AWARD NUMBER: W81XWH-18-1-0456

TITLE: Predicting Situational Onset of Aggression in Minimally Verbal Youth with Autism Using Biosensor Data and Machine Learning Algorithms

PRINCIPAL INVESTIGATOR: Matthew Goodwin, Ph.D.

CONTRACTING ORGANIZATION: Northeastern University, Boston, MA

REPORT DATE: December 2022

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE December 2022	**2. REPORT TYPE** Final	**3. DATES COVERED** 01Sep2018-31Aug2022
4. TITLE Predicting Situational Onset of Aggression in Minimally Verbal Youth with Autism Using Biosensor Data and Machine Learning Algorithms		**5a. CONTRACT NUMBER** W81XWH-18-1-0456 **5b. GRANT NUMBER** AR170209 **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Matthew Goodwin, PhD E-Mail:m.goodwin@northeastern.edu		**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Northeastern University 360 Huntington Ave Boston MA 02115		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Development Command Fort Detrick, Maryland 21702-5012		**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Unpredictable aggressive behavior by youth with autism isolates them from educational, social, and family activities. Approximately 2/3 of youth with autism display aggression, a common reason for treatment referral. However, evidence-based pharmacological and behavioral interventions for aggression in ASD are frequently ineffective. Aggression is particularly impairing in the 30-40% of youth with autism who are minimally verbal and cannot verbalize their distress. Aggression may represent a maladaptive attempt to express or modulate distress-related physiological arousal. We hypothesized that physiological arousal precedes aggressive behavior. We aimed to predict aggression in minimally verbal autism participants before it occurs using data collected from a wrist-worn physiological sensor and time-synchronized behavior observation. Using sophisticated machine learning algorithms linking observable aggression to preceding physiological signals (heart rate, skin conductance), we demonstrate that aggression can be predicted three minutes before it occurs with 80-90% accuracy. These findings enable new opportunities for pre-emptive intervention.

15. SUBJECT TERMS

Autism Spectrum Disorder, ASD, Minimally Verbal, Aggression, Prediction, Physiological Arousal, Arousal Modulation, Machine Learning.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRDC
Unclassified	Unclassified	Unclassified	Unclassified	29	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

Page

1. INTRODUCTION:

Unpredictable and potentially dangerous aggressive behavior by youth with autism spectrum disorder (ASD) isolates them from important educational, social, and family activities, thereby increasing difficulties and costs associated with the condition. As many as 2/3 of youth with ASD display aggression, one of the primary reasons they get referred for treatment. Aggression presents serious safety risks for the individual and others in the environment and frequently occurs with agitation, meltdowns, and other concerning behaviors that are difficult to manage. Families report that aggression increases their stress, isolation, and financial burden and decreases available support options. Aggression toward others is significantly impairing and challenging to manage in the 30-40% of youth with ASD who are minimally verbal (MV-ASD). Their difficulty verbalizing distress can lead to behaviors that seem to occur without warning, sometimes long after any observable trigger. This unpredictability makes aggression toward others in MV-ASD dangerous and presents a barrier to accessing the community. Evidence-based pharmacological and behavioral interventions for ASD aggression are frequently ineffective due to significant medication side effects or insufficient time to provide de-escalation strategies. Aggression toward others may represent a maladaptive attempt to express or modulate distress-related physiological arousal. Thus, we hypothesized that physiological arousal precedes aggressive behavior.

This project aimed to predict aggression toward others in MV-ASD before it occurs using machine learning on data collected from a commercially available wrist-worn wireless physiological sensor. The project provides predictive information (i.e., the onset of aggressive behavior in the proximal future using physiological data from the recent past) that may enable new opportunities for intervention. This innovative approach has the potential to improve our ability to identify escalating distress in youth with MV-ASD, overcoming their inherent difficulty conveying feelings and emotions. By linking observable aggressive behavior to the detection of preceding physiological signals (e.g., heart rate, sweating), we hope to move the field of behavioral assessment and treatment in autism towards a more biologically-based, data-informed approach that is focused on prospective monitoring, prevention, and real-time intervention.

2. KEYWORDS:

Autism Spectrum Disorder, ASD, Minimally Verbal, Aggression, Prediction, Physiological Arousal, Arousal Modulation, Machine Learning.

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Goal 1: Establish physiological biomarkers of imminent aggression. We will observe and record aggression toward others in 40 MV-ASD inpatient youth during repeated naturalistic observations in an inpatient psychiatric unit while they wear a validated wireless autonomic biosensor that measures physiological arousal (i.e., cardiovascular and electrodermal) and motor activity. These data, in combination with time-synchronized coding of aggression and non-aggression by research staff using a mobile application, will be analyzed by machine learning algorithms to create a set of properties (a "classifier") that predict imminent aggression (i.e., the onset of aggressive behavior in the proximal future using physiological data from the recent past).

Goal 2. Evaluate the positive predictive value and reliability of imminent aggression prediction. We will apply the highest-performing classifiers from Aim 1 to validate aggression prediction prospectively in an independent MV-ASD inpatient youth sample (n=20) and examine classifier performance and stability over time.

What was accomplished under these goals?

The proposal funded data collection from 60 MV-ASD participants. However, no data was received from the Partnering PI during the funding period or no-cost extension. While COVID-19 may explain a delay, a reason for the absence of any shared data has yet to be provided. Considering this absence of data transmitted from the Partnering PI, the team at Northeastern generated the following results using data from 20 MV-ASD inpatient participants collected before the funding period.

Overview of the research rationale

Due to the limited amount of available data, we explored models with different levels of complexity. Specifically, we explored *logistic regression* (LR), *deep neural networks* (DNNs), and *support vector machines* (SVMs). LR is the least complex model and, thus, less prone to overfitting. DNNs and SVMs are more flexible but can easily overfit. In our subsequently reported results, we impose different regularization terms when training the DNNs to reduce complexity. Our results indicate that calibration data from new study participants and repeated data from the same study participant lead to higher predictive power (see Appendix: Tables & Figures). Thus, we also explored a semi-supervised domain adaptation strategy to personalize population models.

Limitations of the PETRA paper

In our most recent publication [1], analyzing 3 minutes of biosensor data in the past and predicting aggression onsets 3 minutes in the future, we obtained *area under the curve* (AUC) values ranging from 0.86 to 0.98 with LR and SVMs, respectively. In these analyses, we assumed that all feature vectors extracted from the biosensor time series were independent. However, upon further examination, we learned that past time windows had a 75% overlap which may have led SVM to overfit the data. Generalizing these results to unseen data was impossible due to the scarcity of data. This became clear when extending the scenarios for leaving individuals or sessions out of cross-validation, resulting in much lower accuracy for all models.

To address this problem, we performed a series of experiments under different data split paradigms to avoid overlap between training and testing samples. Specifically, we considered *session splits* (SS), where the first 80% of the session is used for model training while the last 20% is used for testing; *leave individuals out* (LIO), where specific individuals are used only for testing; and *leave sessions out* (LSO), where sessions are left out for testing. Consistent with our previously published work [1-4], we investigated population and person-dependent models, PM and PDM, respectively. LIO was used only with PMs, while LSO was only used with PDMs.

Extensive exploration using LR

As in [1], we investigated the prediction performance for different observational windows (time in the past Tp) and prediction windows (time in the future Tf) of 60, 120, and 180 seconds. We also investigated the influence of augmenting feature vectors with behavior-dependent features. Specifically, for *augmented feature vectors,* we considered a boolean feature to determine if an aggression episode happened and a real feature counting the time since an onset of aggression occurred. We also investigated prediction performance for different aggression intensities by introducing an intensity label obtained by clustering the norm of acceleration data from the biosensor during aggression episodes. We classified aggression episodes into low, medium, and high intensities. When investigating PDMs performance, we only selected participants with a minimum number of sessions (five and seven) to enable more reliable test performance estimation. This is important since most data in our domain is unbalanced (i.e., fewer aggression episodes than non-aggression periods). Selecting participants with more sessions leads to more aggression episodes in the test data.

Analyzing the results, we observe the following:

- Observing past windows of 60 or 120 seconds and predicting up to 120s in the future, PDM-SS models obtained the best results in participants with at least seven sessions (Exp.3), reaching AUCs of 0.73. Filtering only for participants with at least five sessions led to a considerable decrease in AUCs (0.58 was the best result; Exp.4).

- PM-SS obtained the second-best LR results with AUCs of 0.66, observing 120s in the past and predicting 120s in the future.

- Overall performance decreases comparing PMs and PDMs with SS and LIO or LSO indicates the need for more calibration data within each participant and each session.

- Comparing classifier performance with different aggression intensities (Exp. 7), we observe an apparent performance increase in predicting high-intensity aggression episodes. These results demonstrate that more intense aggression episodes are more easily predicted. And contrary to other results wherein augmenting feature vectors did not seem to improve the predictive power of the models, feature vector augmentation appeared to enhance the prediction capability of high-intensity aggression behaviors.

The above results indicate that more data is needed to: (i) estimate more accurate test performances; (ii) train more sophisticated classifiers; and (iii) improve the performance of population and person-dependent models. Furthermore, behavior-dependent features improved the prediction accuracy of high-intensity aggression episodes. However, the lack of available data presents challenges confirming this hypothesis. Finally, the superior performance of SS and PDM-SS models indicates that further predictive gains may be obtained if more person-dependent calibration data were available.

SVM for aggression prediction

We analyzed the performance of SVMs following the same experimental setup used for LR above. The results are summarized in tables Ex8 - Ex12.

Analyzing the results, we observe the following:

- For PM-SS (Ex8), some combinations of observation window and prediction window lengths results were comparable to LR, achieving maximum AUCs of 0.66 using both feature and augmented feature vectors. However, LR generated better performance when making predictions further into the future.

- PM-LIO (Ex9) SVMs presented poor performance, slightly worse than LR in the same scenario (Ex2). The limited amount of data plays a central role in the generalization capability of the models. This appears especially true for more flexible models.

- For similar reasons, SVMs performed poorly in PDM with SS and LSO. While the performance of PDM-SS (Ex10) is considerably worse than the LR counterpart, results were not so in the LSO configuration, indicating that both methods could benefit from additional across-session calibration data.

- When analyzing SVM performance for different aggression intensities (Ex12), the gain in performance when classifying high-intensity episodes is not as evident as for LR.

Overall, SVMs generated prediction performances for PM with SS comparable to the LR results for the same scenario. In all other scenarios, SVMs perform worse than LR, indicating poorer generalization of SVMs for unseen sessions or individuals. These results are expected, given the limited data available and intrinsic variability across sessions and participants.

Neural network for aggression prediction

We analyzed the performance of DNNs following the same experimental setup used for LR above. The results are summarized in tables Ex13 - Ex17.

The model consisted of a multilayer perceptron with three hidden layers, a total of 35,000 parameters, and rectified linear unit (ReLU) as activation functions of the hidden layer. We trained the model using backpropagation and the Adam optimizer. To avoid overfitting and improve generalization to unseen data, we used dropout layers between each layer [5] and L1 and L2 regularizations [6].

Analyzing the results, we observe the following:

- Regularized DNNs provide the best results for PM-SS (Ex13), achieving AUCs consistently close to 0.70 and up to 0.73 for some combinations of Tp and Tf.

- When considering PM-LIO, DNNs had slightly worse generalization than LR but better than SVMs.

- Results for PDM models were worse than DNNs compared to LR for SS (Exp15) and comparable to LSO since all models struggled to generalize across sessions for PDMs. This is expected since the data available for individual models was extremely limited.

- Like SVMs, DNNs perform best at predicting high-intensity aggression episodes (Ex17). Additionally, DNNs benefitted from augmented feature vectors, suggesting that information on aggression frequency, duration, and periodicity may further improve prediction.

Overall, DNNs performed well for population models and benefited from behavior-related features. This is expected since DNNs present flexible models while regularizations reduce the risk of overfitting. Furthermore, DNN-based PDMs with SS performed considerably better than SVMs despite the lack of data in the PDM scenarios. We attribute this gain to the L1-L2 regularizations and dropout layers. Continuing to explore well-regularized DNNs is a promising direction to explore when more data becomes available.

Supervised and unsupervised domain adaptation across sessions

One open question regards the possibility of adapting PMs to specific individuals. An active topic of research in the machine learning community, domain adaptation employs models trained in a source domain using data from different domains. In our research, we explored labeled and unlabeled adaptation of PMs to specific individuals.

While the procedure is straightforward when labels are assumed known, we resort to a pseudo-labeling approach for the unlabeled scenario [7]. The procedure consists of classifying all features of unlabeled sessions and attributing labels to data points if the classifier is sufficiently confident about its label (i.e., the class posterior is greater than a threshold). New pseudo-labeled data points are then included in the training data, and the classifier is retrained. This cycle continues in an iterative procedure until a stop criterion is achieved or the classifier’s prediction is sufficiently confident for all unlabeled data points. We applied these strategies to individuals unseen by the PMs, adding sessions individually.

We computed AUCs for the new individual with the remaining unseen sessions to assess prediction accuracy.

Table Ex18 summarizes the results of our unsupervised domain adaptation. We present the percentage of median AUC change between the initial PM AUC and the final AUC obtained after all adaptation cycles across all individuals.

Analyzing the results, we observe the following:

- In almost all combinations of Tp and Tf, the procedure led to positive improvements, corroborating our conjecture that session-wise calibration data can be used to improve model performance.

- For this experiment, we also plot AUC versus the number of sessions used for adaptation for each participant. Plots are provided for the best median improvement results and are bolded in Table Ex18 for both augmented feature vectors (Figure 1) and feature vectors (Figure 2). We observe in both cases that five out of the eight participants demonstrated performance improvements. We also observed that adaptation was successful for all participants with six or more sessions (three out of eight) in both cases.

To further explore median improvements concerning features of the data set, we plot the number of aggressions, average aggression duration (in seconds), and observation duration (duration of all sessions combined in hours) versus the median of the maximum AUC improvement, where the median was taken across all Tp and Tf results. Figures 3 and 4 illustrate augmented feature vectors and feature vectors, respectively. In these plots, we observe that domain adaptation did not improve model accuracy in participants with the least amount of data available (participants 1121 and 1168). Participant 1017's prediction performance was also not improved using domain adaptation; however, we cannot identify why. More data from more participants are needed to determine.

Overall, it appears that domain adaptation techniques can be beneficial if enough data is available. However, the lack of theoretical convergence guarantees of unsupervised domain adaptation models makes it challenging to make categorical affirmations about its efficiency, especially with such limited data. Nevertheless, we are excited by these initial results since unlabeled data will be abundant in real-life applications of challenging behavior prediction systems.

Conclusions

Across the project period, we analyzed the performance of various machine learning models with different approximation capabilities and data requirements. Our results indicate that model performance varies depending on different operating scenarios. For instance, DNNs provided the best results under PM-SS, while LR yielded higher accuracy for PDM-SS. This is understandable since PDMs contain less information, leading to better performance of less flexible and simplistic models such as LR. When working with LIO or LSO, we notice that all models generalized poorly, leading to AUCs close to 0.5 (random chance).

For this reason, we explored an unsupervised domain adaptation technique based on pseudo-labeling. This approach led to promising results wherein AUCs consistently improved for a subset of participants. Finally, our experiments categorically demonstrate the need for more data. Estimating test performance and training PDMs is extremely challenging when working with such a limited and unbalanced data set.

Future work

Our planned next steps using biosensor data and machine learning algorithms to predict aggression in ASD before it occurs include: (i) further exploring domain adaptation techniques to improve model performances for both LIO and LSO scenarios; (ii) exploring the rate of aggression from a point-process perspective, including Markov-modulated point process models, since we hypothesize that different arousal states may lead to different rates of aggression arrival; and (iii) expanding prediction of aggression to other prevalent challenging behaviors in ASD, including self-injurious behavior, tantrums, property destruction, and elopement.

References

1. mbiriba, T., umpanasoiu, D. ., Heathers, ., oannidis, S., rdo mu , D., & Goodwin, M. S. (2020, June). Biosensor prediction of aggression in youth with autism using kernel-based methods. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (pp. 1-6).

2. Goodwin, MS, Mazefsky, C, Ioannidis, S, Erdogmus, D, Siegel, M (2019). Predicting aggression to others in youth with autism using a wearable biosensor. *Autism Research*, 12(8), 1286-1296. NIH Matilda White Riley Early-Stage Investigator Paper Award.

3. Ozdenizci, O Cumpanasoiu, C, Mazefsky, C, Siegel, M, Erdogmus, D, Ioannidis, I, Goodwin, MS (2018). Time-series prediction of proximal aggression onset in minimally-verbal youth with autism spectrum disorder using physiological

biosignals. In *Proceedings of 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5745-5748.

4. Goodwin, MS, Ozdenizci, O, Tian, P, Cumpanasoiu, C, Guo, A, Stedman, A, Peura, C, Mazefsky, C, Siegel, M, Erdogmus, D, Ioannidis, I (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*, 201-207.

5. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., ... & Li, C. L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, *33*, 596-608.

6. Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929-1958.

7. Ying, Xue. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series. 1168.* 022022. 10.1088/1742-6596/1168/2/022022.

What opportunities for training and professional development has the project provided?

This project enhanced interdisciplinary research experience, including machine learning involving human subjects' data, for eight students (three at the post-doctoral level, three doctoral, and two masters).

How were the results disseminated to communities of interest?

Nothing to Report.

What do you plan to do during the next reporting period to accomplish the goals?

Nothing to Report.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

Individuals with autism who are severely impacted and engage in challenging behavior are understudied and underserved. The predictive analytic methods produced in this project lay the foundation for expanding data collection on precursor behaviors associated with aggression and the future development of just-in-time adaptive intervention mobile health systems that enable new opportunities for pre-emptive intervention.

What was the impact on other disciplines?

The work completed in this project advances signal modeling, analysis, and machine learning theories, leading to methods that allow rapid, seamless individualization and online adaptation of models that make many conceptual technologies in the lab more practical for widespread societal use. We anticipate the applicability of our work to fields such as brain/body interfaces to computers, robots, or other cyber-physical systems (CPS), multimodal sensor-based user/human state estimation, and intent inference for assistive technology, teleoperation, and collaborative heterogeneous human-CPS teams. Our techniques are also likely to have general applicability to other health application contexts that involve event prediction (e.g., seizure detection) and human action/decision-making (e.g., intensive care unit triage).

What was the impact on technology transfer?

Nothing to Report.

What was the impact on society beyond science and technology?

Nothing to Report.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

The direction and scope of this work remain the same. However, we have ended our reliance on data from the Partnering PI due to performance issues.

Actual or anticipated problems or delays and actions or plans to resolve them

Over a four-year period, the absence of data provided by the Partnering PI created a lost opportunity to more rapidly advance this machine learning/behavior prediction research.

Changes that had a significant impact on expenditures

Nothing to Report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

Nothing to Report.

Significant changes in use or care of vertebrate animals

Nothing to Report.

Significant changes in use of biohazards and/or select agents

Nothing to Report.

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Published:

1. Imbiriba, T, Cumpanasoiu, D, Heathers, J, Ioannidis, S, Erdogmus, D, Goodwin, MS (2020). Biosensor prediction of aggression in youth with autism using kernel-based methods. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA)*,1-6.

2. Goodwin, MS, Mazefsky, C, Ioannidis, S, Erdogmus, D, Siegel, M (2019). Predicting aggression to others in youth with autism using a wearable biosensor. *Autism Research*, 12(8), 1286-1296. NIH Matilda White Riley Early-Stage Investigator Paper Award.

3. Ozdenizci, O Cumpanasoiu, C, Mazefsky, C, Siegel, M, Erdogmus, D, Ioannidis, I, Goodwin, MS (2018). Time-series prediction of proximal aggression onset in minimally-verbal youth with autism spectrum disorder using physiological biosignals. In *Proceedings of 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5745-5748.

4. Goodwin, MS, Ozdenizci, O, Tian, P, Cumpanasoiu, C, Guo, A, Stedman, A, Peura, C, Mazefsky, C, Siegel, M, Erdogmus, D, Ioannidis, I (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*, 201-207.

Books or other non-periodical, one-time publications.

Nothing to Report.

Other publications, conference papers and presentations.

1. Goodwin, MS (2022). The Promises of Wearable Bio-Sensing Technology for Youth with Autism. <u>Keynote address</u> at the Symposium on Innovative Solutions for Improving Quality of Life for Individuals with Autism, Calgary University, Virtual, May 5-6.

2. Goodwin, MS (2022). Predicting Challenging Behavior in Individuals with Autism using Wearable Biosensors and Machine Learning Classifiers. <u>Invited address</u> at the Munroe-Meyer Institute Grand Rounds, Omaha, NE, Mar 31.

3. Goodwin, MS (2022). Predicting Challenging Behavior in Individuals with Autism using Wearable Biosensors and Machine Learning Classifiers. <u>Keynote address</u> at the 15th Annual Autism Update Conference, Stanford University, Virtual, Mar 19.

4. Goodwin, MS (2021). Predicting Challenging Behavior in Individuals with Autism using Wearable Biosensors and Machine Learning Classifiers. <u>Keynote</u> at the Riverview School, Cape Cod, July 29.

5. Goodwin, MS (2021). Predicting Challenging Behavior in Individuals with Autism using Wearable Biosensors and Machine Learning Classifiers. Invited address at the 14th NIH Matilda White Riley Behavioral and Social Sciences Honors, Virtual, May 5.

6. Goodwin, MS (2021). Predicting Challenging Behavior in Minimally-Verbal Youth with Autism Using Wearable Biosensor Data and Machine Learning Classifiers. Invited address at Translational Research Day 2021, Tufts Clinical and Translational Science Institute, Tufts University, April 27.

- **Website(s) or other Internet site(s)**

Nothing to Report.

- **Technologies or techniques**

Nothing to Report.

- **Inventions, patent applications, and/or licenses**

Nothing to Report.

- **Other Products**

Nothing to Report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	*Matthew Goodwin*
Project Role:	*PI*
Researcher Identifier (e.g. ORCID ID):	0000-0002-4237-601X
Nearest person month worked:	*1*
Contribution to Project:	*No change.*
Funding Support:	*The Simons Foundation and Nancy Lurie Marks Foundation provided additional, complementary support toward this project, as explained in materials submitted to the DoD; these foundations provide additional funding to open more sites and enroll more patients.*

Name: *Tales Imbiriba*
Project Role: *Post Doc*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *3*

Contribution to Project: *Performed machine learning analyses on N=20 data set, the precursor to this funding.*

Funding Support:

Name: *Ahmet Demirkaya*
Project Role: *Graduate Student*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *8*

Contribution to Project: *Performed machine learning analyses on N=20 data set, the precursor to this funding.*

Funding Support:

Name: *Natasha Yamane*
Project Role: *Graduate Student*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *4*

Contribution to Project: *Assisted with machine learning analyses preformed on machine learning analyses on N=20 data set, the precursor to this funding.*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Title: SCH: Enhancing Automated Prediction of Challenging Behavior in Individuals with Autism Using Biosensor Data and Machine Learning.

Role: Goodwin PI

Project Number: R01LM014191 (06/2022 – 05/2026, 1.25 Sum mo Yr 1-4)

Source: National Library of Medicine (NLM)/National Cancer Institute (NCI)

Goals: This project aims to advance fundamental machine learning theory and techniques that facilitate rapid model individualization and continuous online model adaptation with little or no labeled data. To this end, we will contribute to areas including domain adaptation, transfer learning, lifelong learning, and robust modeling and inference. Three Specific Aims (SA) guide the project: (SA-1) Rapid physiological and behavioral data model individualization; (SA-2) Continuous lifelong physiological and behavioral data model adaptation; and (SA-3) Validation of model individualization and adaptation techniques with prospective data collected in a clinical setting from our partners at the Marcus Autism Center at Emory University who specialize in functional analysis of challenging behavior in individuals with autism. Across these Aims, we will explore applications of semi-supervised learning theory, data importance weighting, Support Vector Machines, neural network models, Hierarchical Markov-Modulated Point Process Models, and Bayesian evidence fusion.

What other organizations were involved as partners?

Nothing to Report.

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS:

QUAD CHARTS:

Not applicable. The work reported involved an NCE without participation from the Partnering PI.

9. APPENDICES: .

See Appendix: Tables & Figures.

APPENDIX: Tables & Figures

Table 1: Classification Results (using N=20 pre-award corpus). Experiments 1-19 were performed considering two scenarios. In scenario 1, augmented features (human-in-the-loop annotations) were constructed using onsets of aggression, and feature vectors ending within a positive episode were kept. In scenario 2, augmented features were constructed based on offsets of aggression and feature vectors ending within positive episodes were removed. Bolded AUCs represent the best performing.

Logistic Regression Results:

Ex1: LR, Population Models with Session Splits (PM-SS), 80% training, 20% testing						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.65	0.62	0.59	0.65	0.63	0.60
$\tau_p = 120$	**0.66**	0.62	0.60	**0.66**	0.63	0.60
$\tau_p = 180$	0.65	0.61	0.60	0.65	0.62	0.60

Ex2: LR, Population Models with Leave-Individuals-Out (PM-LIO)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.60	0.58	0.56	0.50	0.49	0.48
$\tau_p = 120$	0.59	0.56	0.55	0.49	0.48	0.47
$\tau_p = 180$	0.57	0.55	0.54	0.48	0.47	0.47

Ex3: LR, Person-Dependent Models with Session Splits (PDM-SS) for individuals with more than 7 sessions (3 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	**0.72**	0.66	0.60	**0.73**	0.66	0.61
$\tau_p = 120$	**0.72**	0.64	0.60	**0.73**	0.64	0.60
$\tau_p = 180$	0.69	0.60	0.57	0.69	0.61	0.57

Ex4: LR, Person-Dependent Models with Session Splits (PDM-SS) for individuals with more than 5 sessions (7 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.56	0.57	0.51	0.58	0.58	0.56
$\tau_p = 120$	0.55	0.54	0.49	0.58	0.57	0.54
$\tau_p = 180$	0.53	0.51	0.48	0.56	0.55	0.54

Logistic Regression Results (cont.)

Ex5: LR, Person-Dependent Models with Leave-Sessions-Out (PDM-LSO) for individuals with more than 7 sessions (3 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.54	0.52	0.53	0.51	0.52	0.53
$\tau_p = 120$	0.51	0.53	0.52	0.49	0.51	0.51
$\tau_p = 180$	0.52	0.54	0.55	0.50	0.50	0.53

Ex6: LR, Person-Dependent Models with Leave-Sessions-Out (PDM-LSO) for individuals with more than 5 sessions (7 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.55	0.55	0.52	0.47	0.48	0.48
$\tau_p = 120$	0.54	0.54	0.51	0.47	0.48	0.47
$\tau_p = 180$	0.54	0.53	0.53	0.48	0.48	0.48

Ex7: LR, Population models with leave subjects out and aggression intensity determined by clustering the norm of acceleration data (PM-CLUST)							
		Augmented Feature Vectors			Feature Vectors		
		$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
Low	$\tau_p = 60$	0.62	0.59	0.56	0.53	0.53	0.51
	$\tau_p = 120$	0.60	0.57	0.54	0.52	0.50	0.50
	$\tau_p = 180$	0.58	0.56	0.55	0.51	0.50	0.50
Mid	$\tau_p = 60$	0.63	0.60	0.58	0.60	0.57	0.56
	$\tau_p = 120$	0.62	0.59	0.58	0.59	0.57	0.57
	$\tau_p = 180$	0.61	0.60	0.58	0.58	0.58	0.57
High	$\tau_p = 60$	**0.72**	0.70	0.68	0.67	0.65	0.63
	$\tau_p = 120$	**0.72**	0.68	0.65	0.67	0.63	0.63
	$\tau_p = 180$	0.67	0.64	0.61	0.63	0.59	0.57

SVM Results:

Ex8: SVM, Population Models with Session Splits (PM-SS)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	**0.66**	0.66	0.64	**0.66**	0.65	0.64
$\tau_p = 120$	0.60	0.61	0.53	0.60	0.62	0.56
$\tau_p = 180$	0.56	0.60	0.50	0.57	0.57	0.53

Ex9: SVM, Population Models with leave-individuals-out (PM-LIO)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.56	0.55	0.54	0.54	0.53	0.52
$\tau_p = 120$	0.52	0.53	0.53	0.52	0.53	0.52
$\tau_p = 180$	0.50	0.50	0.50	0.51	0.50	0.50

Ex10: SVM, Person-dependent models with session splits (PDM-SS) for individuals with more than 7 sessions (3 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.59	0.64	0.62	0.60	0.64	0.62
$\tau_p = 120$	0.50	0.51	0.52	0.51	0.52	0.53
$\tau_p = 180$	0.50	0.50	0.50	0.51	0.50	0.50

Ex11: SVM, Individual models with leave sessions out (PDM-LSO) for individuals with more than 7 sessions (3 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.50	0.50	0.50	0.50	0.50	0.50
$\tau_p = 120$	0.50	0.50	0.50	0.50	0.50	0.50
$\tau_p = 180$	0.50	0.50	0.50	0.50	0.50	0.50

SVM Results (cont.)

Ex12: SVM, Population models with leave subjects out and aggression intensity determined by clustering the norm of acceleration data (PM-CLUST)							
		Augmented Feature Vectors			Feature Vectors		
		$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
Low	$\tau_p = 60$	0.50	0.51	0.52	0.49	0.49	0.51
	$\tau_p = 120$	0.50	0.51	0.52	0.49	0.51	0.52
	$\tau_p = 180$	0.51	0.50	0.50	0.50	0.50	0.50
Mid	$\tau_p = 60$	0.59	0.58	0.56	0.57	0.55	0.54
	$\tau_p = 120$	0.54	0.56	0.55	0.55	0.54	0.54
	$\tau_p = 180$	0.50	0.50	0.50	0.50	0.50	0.50
High	$\tau_p = 60$	0.61	0.60	0.62	0.59	0.58	0.59
	$\tau_p = 120$	0.54	0.54	0.55	0.54	0.54	0.55
	$\tau_p = 180$	0.50	0.50	0.50	0.50	0.50	0.50

DNN Results:

Ex13: DNN, Population Models with Session Splits (PM-SS)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	**0.72**	0.70	0.67	**0.73**	0.70	0.68
$\tau_p = 120$	**0.72**	0.70	0.66	**0.73**	0.70	0.66
$\tau_p = 180$	0.70	0.66	0.64	**0.73**	0.68	0.66

Ex14: DNN, Population Models with leave-individuals-out (PM-LIO)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.63	0.58	0.57	0.56	0.54	0.52
$\tau_p = 120$	0.61	0.58	0.56	0.56	0.53	0.52
$\tau_p = 180$	0.60	0.57	0.56	0.55	0.53	0.52

DNN Results (cont.)

Ex15: DNN, Person-Dependent Models with Session Splits (PDM-SS) for individuals with more than 7 sessions (3 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.62	0.57	0.53	0.62	0.55	0.50
$\tau_p = 120$	0.67	0.60	0.54	0.62	0.62	0.55
$\tau_p = 180$	0.61	0.61	0.58	0.59	0.60	0.54

Ex16: DNN, Person-Dependent Models with leave sessions out (PDM-LSO) for individuals with more than 7 sessions (3 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	0.51	0.51	0.51	0.51	0.51	0.52
$\tau_p = 120$	0.51	0.50	0.51	0.51	0.51	0.51
$\tau_p = 180$	0.51	0.51	0.49	0.51	0.51	0.51

Ex17: DNN, Population models with leave individuals out and aggression intensity determined by clustering the norm of acceleration data (PM-CLUST)							
		Augmented Feature Vectors			Feature Vectors		
		$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
Low	$\tau_p = 60$	0.60	0.56	0.52	0.51	0.50	0.50
	$\tau_p = 120$	0.58	0.54	0.51	0.50	0.51	0.49
	$\tau_p = 180$	0.51	0.55	0.52	0.50	0.48	0.49
Mid	$\tau_p = 60$	0.63	0.60	0.58	0.61	0.58	0.56
	$\tau_p = 120$	0.63	0.61	0.60	0.60	0.57	0.55
	$\tau_p = 180$	0.60	0.65	0.61	0.60	0.59	0.56
High	$\tau_p = 60$	0.66	0.67	0.70	0.61	0.63	0.64
	$\tau_p = 120$	0.67	0.69	0.68	0.62	0.65	0.61
	$\tau_p = 180$	0.68	0.67	0.66	0.62	0.61	0.61

LR Domain Adaptation Results:

Ex18: Domain Adaptation (Unsupervised) for individuals with more than 4 sessions (8 individuals)						
	Augmented Feature Vectors			Feature Vectors		
	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$	$\tau_f = 60$	$\tau_f = 120$	$\tau_f = 180$
$\tau_p = 60$	-0.05	7.67	**8.78**	9.81	10.22	6.61
$\tau_p = 120$	7.63	1.06	1.01	9.74	0.02	-4.03
$\tau_p = 180$	6.75	5.05	4.72	**12.18**	-2.26	-10.61

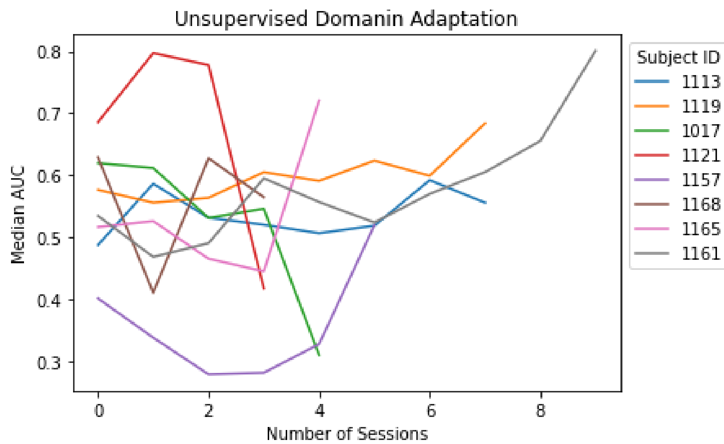

Figure 1: AUC vs. the number of sessions used for adaptation for all individuals with four or more sessions. Exp19, $\tau_p = 60$, $\tau_f = 180$. *Augmented Feature Vector. DA using Pseudo Labels.*

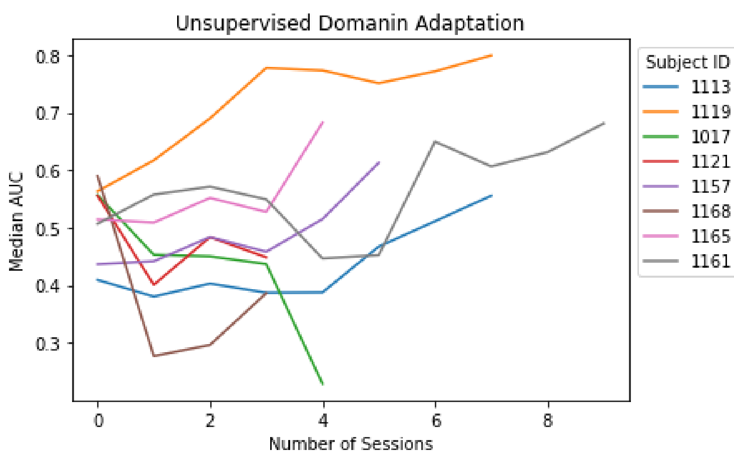

Figure 2: AUC vs. the number of sessions used for adaptation for all individuals with four or more sessions. Exp19, $\tau_p = 180$, $\tau_f = 60$, *with feature vectors and DA using pseudo labels.*

LR Domain Adaptation Results (cont.)

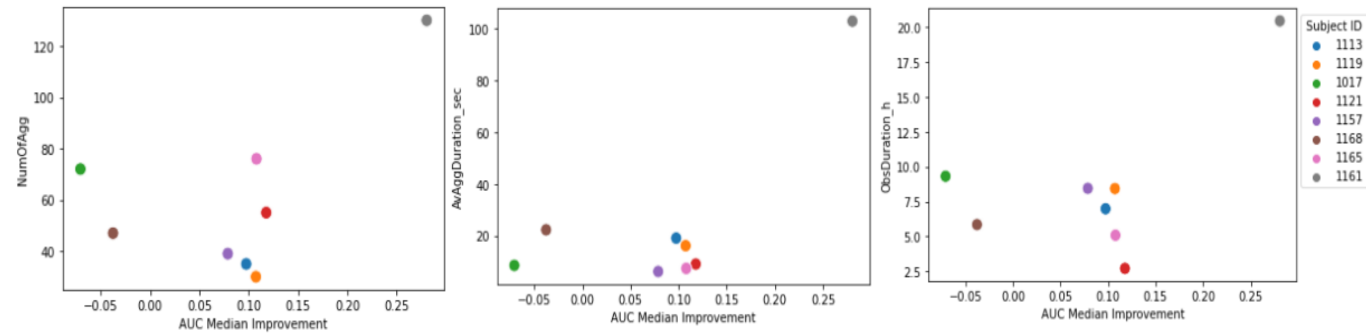

Figure 3: Domain adaptation with augmented feature vectors.

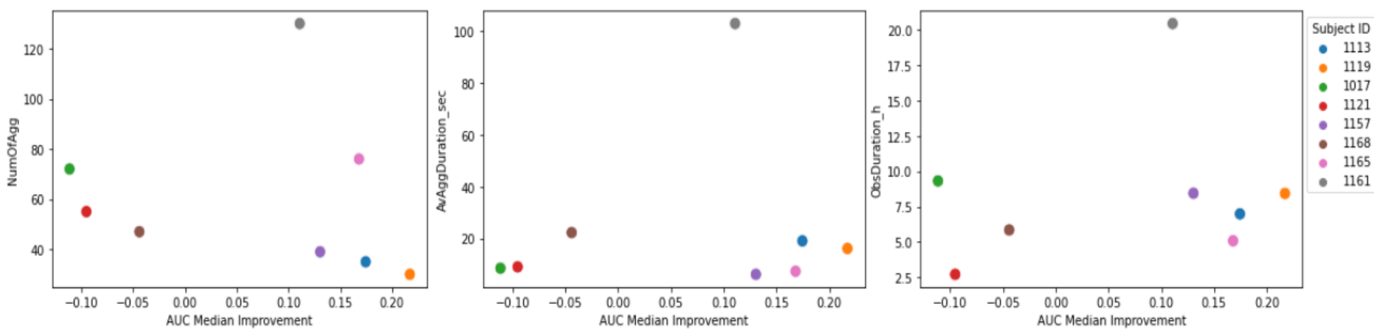

Figure 4: Domain adaptation with feature vectors.