Towards Forecasting Engagement in Children with Autism Spectrum Disorder using Social Robots and Deep Learning

1st Ruchik Mishra
Department of Electrical and Computer Engineering
University of Louisville
Louisville, U.S.A
ruchik.mishra@louisville.edu

2nd Karla Conn Welch
Department of Electrical and Computer Engineering
University of Louisville
Louisville, U.S.A
karla.welch@louisville.edu

SoutheastCon 2023 | 978-1-6654-7611-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/SoutheastCon51012.2023.10115150

***Abstract*—The personalization of therapy for children with Autism Spectrum Disorder (ASD) has been found to be crucial in comparison to a universal approach. This personalization in therapy demands the ability to adapt to the individual's needs and engagement levels to avoid disinterest or meltdowns. This paper proposes the first step towards forecasting engagement of children with ASD during therapy sessions using Blood Volume Pulse (BVP). The BVP data is collected from an interactive session between two children with ASD in the presence of a NAO robot, and the forecast is made using a Deep Learning architecture combining Convolutional Neural Networks (CNNs) and Long-short term Memory (LSTM). Out of the three networks tested: LSTM, CNN and CNN+LSTM, the latter was found to outperform the others and gave a coefficient of determination of 0.955. The forecast was done using less than 3 minutes of prior BVP data to forecast 3 minutes into the future time steps.**

***Index Terms*—autism spectrum disorder, robotics, engagement forecast, Deep Learning, CNN, LSTM, affective computing**

I. Introduction

Social robots have substantiated the evidence of more positive outcomes of intervention for children with ASD as compared to a human therapist [5], [7], [16]. This hypothesis has further been extended to having a personalized approach to therapy with children with ASD which has been highlighted by the authors in [2], [17], [21]. This personalization has been spread across various activities including perspective taking, puzzle solving, etc. [20].

Personalized robotic intervention gives an opportunity to adapt to the affective states of the child with ASD in order to reciprocate appropriately [13], [20]. This is important because meltdowns, challenging behaviors, non-compliance, etc. are common behaviors observed in ASD which might effect the growth and development of the individual [10], [12], [14]. This personalized adaptation in a therapy session becomes even more challenging if the intervention is done with the help of a robot in any of the possible modes of operation: 1) teleoperated or Wizard of Oz approach where the robot is controlled completely by the therapist [15], [27], 2) robotic therapy where a robot is semi-autonomous in the sense that decisions can be overridden by the human therapist [8], [25], and 3) fully autonomous solution with the robot acting on its own [13], [18], [20].

This research was supported by the National Science Foundation (NSF) under Smart and Connected Health (SCH) Grant 1838808.

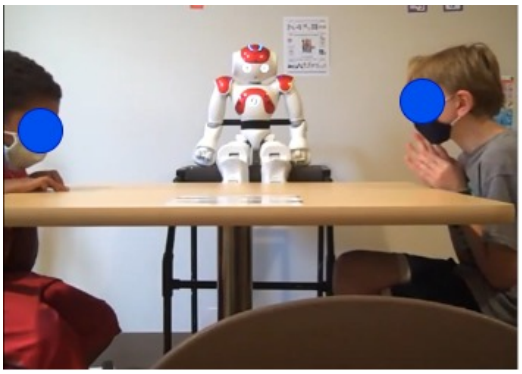


Fig. 1: Interactive session of children with ASD with robot as prompter.

In this paper, we have proposed a method that takes the first step towards forecasting engagement during the robotic intervention of children with ASD. This is done using physiological signals collected during intervention of children with ASD. This approach leverages physiological signals as implicit indicators of affective states, which children with ASD may have difficulty expressing outwardly in ways similar to developmentally typical children [28]. The signals are collected during an interactive session between two children with ASD in the presence of a NAO robot as prompter as can be seen from Figure 1. The approach is motivated by the idea of time-series forecasting of physiological signals using the CNN+LSTM literature.

This paper has been arranged in the following manner: Section II describes the related works in the literature, followed by Section III on data acquisition. Further, the problem formulation is described in Section IV followed by the methodology to achieve it. The results have been presented in Section VI followed by the limitation of this paper in Section VII. Finally

Section VIII presents the concluding remarks of this paper.

II. Related Works

The authors in [17] have used a deep learning network called 'PPA-Net' for finding the engagement of participants using their affective states. In addition, the authors used multiple modalities: audio, facial expressions, body movement, and physiological signals for making this estimation of engagement in individuals from different cultural backgrounds. The major limitation of this work is that all of the predictions were done offline.

Further, modeling engagement has been approached as a binary classification problem by the authors in [11]. They have applied this classification problem for a long-term and in-house intervention using the Kiwi robot for the human-robot interaction, which provides feedback to the child. The scope of this work spans across both individualized and group models for engagement classification and aims at a more real-time classification.

Another example where physiological signals have been used to model different emotions based on the valence and arousal dimensions can be found in [19]. Based on these emotions, the authors presented promising evidence towards engagement perception in autism therapy to overcome the limitations of using self-reports of emotional experiences from children with ASD.

Other works that involve the use of physiological signals for emotion recognition include approaches from the Autoregressive literature. This approach has been followed by the authors in [26] where they have used a Non-linear Autoregressive Integrative model based on the point process model. This paper was an attempt to model the heartbeat in terms of non-linear dynamics as compared to the approach in [4] where the heart rate variability was modelled as a linear model.

A more recent work for detecting challenging behaviors of children with ASD have been proposed by the authors in [1] where they have focused on real-time detection of challenging behaviors using physiological signals using wearable sensors (i.e., Empatica E4 wrist band). These signals were used to classify the behavior of the child as challenging or not challenging.

Unlike the works mentioned in the literature, the authors in [9] have proposed a method using a variation of Logistic Regression to predict the challenging behaviors by extracting the features from the physiological signals collected. These extracted features then are used to predict the onset of aggressive behaviors 1 minute prior to it based on 3 minutes of data collected prior to it.

This work is different from [9] in forecasting using physiological signals in the way that instead of extracting features from time series to predict the onset of challenging behaviors, we propose to forecast the time series itself. This method of time series-forecast is more robust as compared to the approach in [9] in the sense that it is not limited to the domain of challenging/non-challenging behavior classification but can be used for forecasting engagement by attaching a classification model as in [11] or by forecasting different emotions by attaching the classification model used in [19] at the end of our forecast model respectively. In addition, our use of Deep Learning networks allows to forecast much longer sequences ahead in time.

III. Data Acquisition

A. Participants

Six subjects completed this study. All were male and aged between 10.4-11.9 years (M = 11.4 years, SD = 0.86). Subjects were recruited from the population of a university-affiliated autism center. All participants and their caregivers completed consent forms approved by the University's Internal Review Board. All subjects had a diagnosis of ASD, based on Diagnostic Statistical Manual 5^{th} edition [3] completed by a clinical provider at the center or a referring physician/psychologist.

B. Physiological data

The data was acquired using the Empatica E4 wristband for collecting physiological signals from children with ASD. The motivation behind using the E4 device is that it is portable and hence does not restrict any major physical constraints for the movement of children during therapy. Data was collected during an interactive session in which participants were grouped as a pair of two children with ASD and instructed to get to know each other (refer to Figure 1). During this session, the robot acts as the prompter to facilitate the conversation in case of silence for more than 30 seconds or if one participant has dominated the conversation for more than 1 min. The session considered for this paper lasts for approximately 7 minutes; hence, the BVP data used represents 7 minutes of univariate time series data collected at a frequency of 64 Hz. Data from one participant was fully analyzed for this paper.

IV. Problem formulation

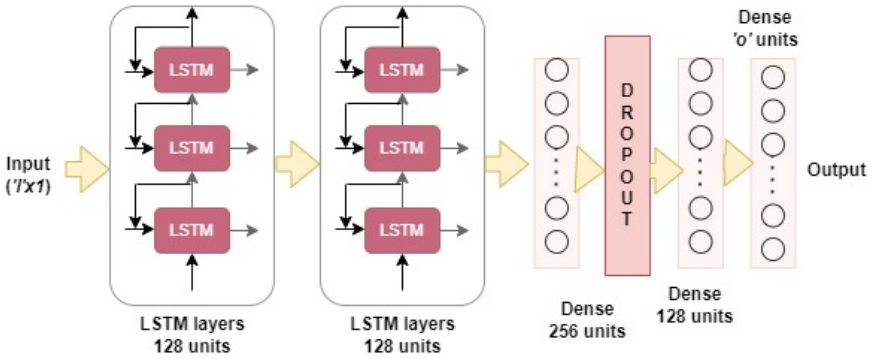
The forecast of the univariate Blood Volume Pulse (BVP) data is given by equation 1 which is similar to the approach used in [6].

$$f([s_{j-q}, s_{j-q+1} \ldots, s_j]) = [\widehat{s}_{j+1}, \ldots, \widehat{s}_{j+r}] \tag{1}$$

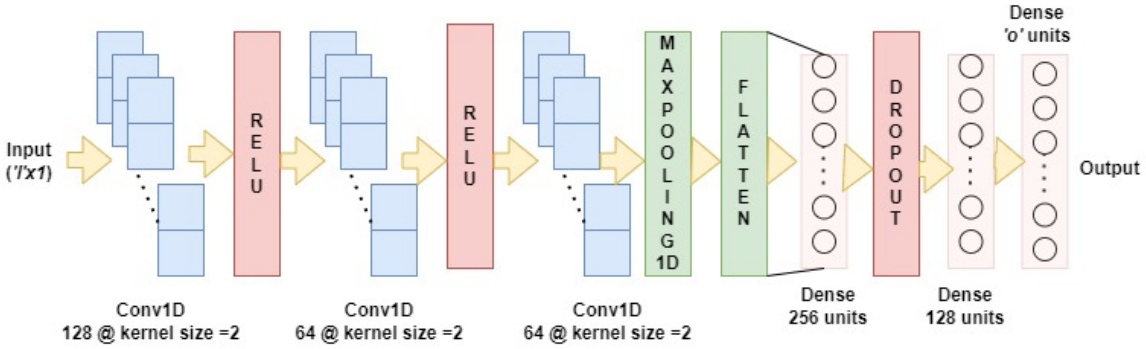
where $f(.)$ is the forecast function, j is the number of data point, q is the number of previous data points used to forecast the BVP signals s for future time steps. A Deep Neural Network is used to estimate this function $f(.)$ using Mean Squared Error (MSE) loss and Coefficient of Determination (R2 score) as the metric as given by equation 2:

$$R^2 = 1 - \frac{\sum_{i=0}^{N}(y_i - \widehat{y}_i)^2}{\sum_{i=0}^{N}(y_i - \widehat{\overline{y_i}})^2} \tag{2}$$

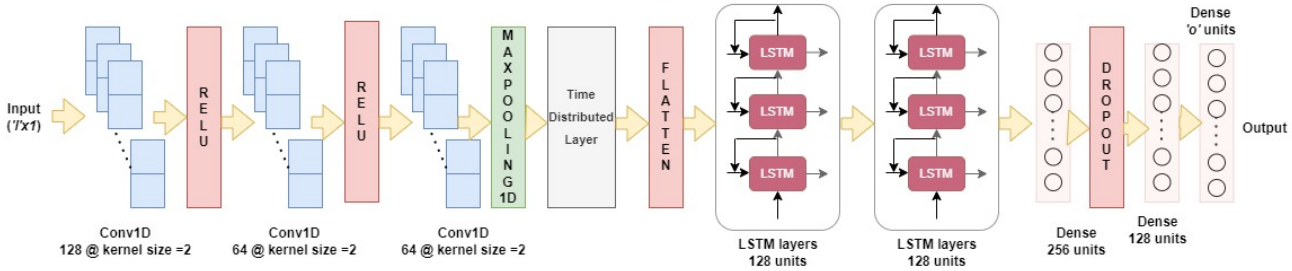
where y_i is the actual value of the signal ($[s_{j+1}, \ldots, s_{j+r}]$), $\widehat{y}_i$ is the predicted value by the function $f(.)$ i.e. ($[\widehat{s}_{j+1}, \ldots, \widehat{s}_{j+r}]$), and N is the total number of data points.



(a) Model 1: LSTM only.



(b) Model 2: CNN only.



(c) Model 3: CNN+LSTM.

Fig. 2: Network architectures used in this paper.

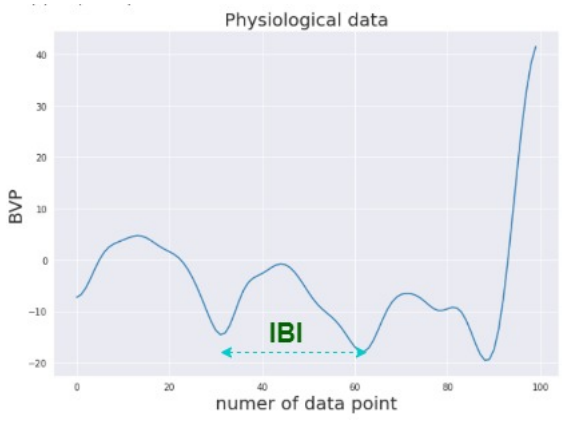
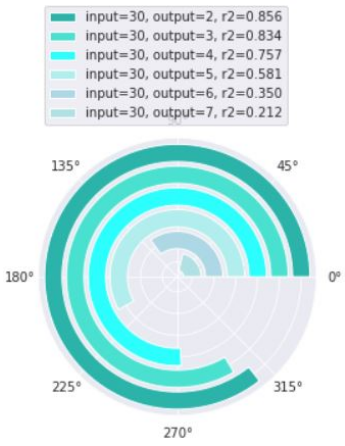


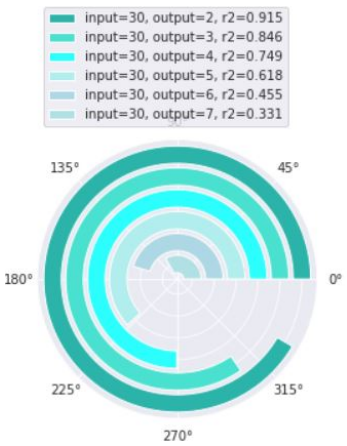
Fig. 3: Number of data points for an Interbeat Interval (IBI).

V. Methodology

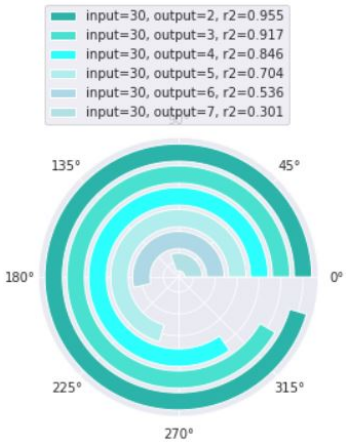
Once the BVP data is collected, we convert into the training and test data in the form that can be used by our Deep Learning model to be used in training. The input dimension of the data is equal to the number of prior data points that are used to make the prediction for the future time steps. Similarly, the output dimension is equal to the number of data points into the future that the model is predicting. More details have been mentioned in Algorithm 1. The training process for this work has been done offline from the data collected during the sessions. The data collected from the Emaptica E4 wrist band is first stored in Empatica's web portal, and then later the data is fetched for offline analysis. The computations are done using Google Colab's premium version that provides decent computational power for the work described in this paper. In addition, the framework used was Tensorflow and Keras for all parts of this



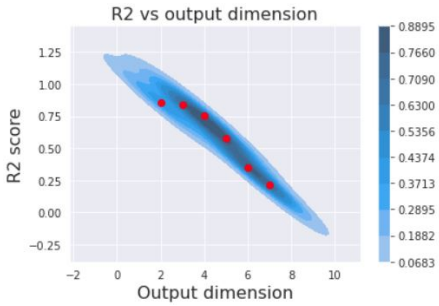
(a) Test data R2 scores for different output dimension options for LSTM model.



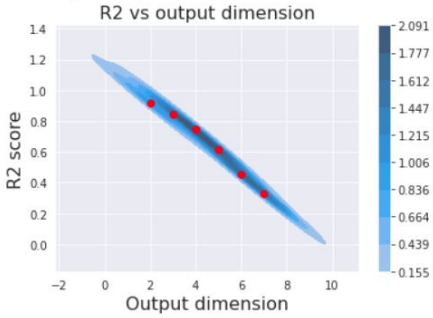
(b) Test data R2 scores for different output dimension options for CNN model.



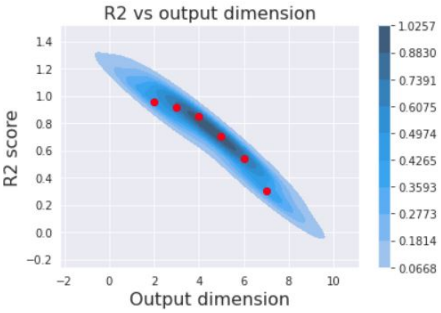
(c) Test data R2 scores for different output dimension options for CNN+LSTM model.



(d) R2 score vs output dimension for LSTM model.



(e) R2 score vs output dimension for CNN model.



(f) R2 score vs output dimension for CNN+LSTM model.

Fig. 4: Results on test data.

project that involved elements of Deep Learning.

Algorithm 1 Forecasting algorithm pseudocode

Require: $\mathcal{X}_{\text{train}}, \mathcal{Y}_{\text{train}}, \mathcal{X}_{\text{test}}, \mathcal{Y}_{\text{test}}$
Collect BVP data: $\mathcal{X}_{\text{BVP}}$
n_{in} = input dimension
n_{out} = output dimension
for i in range(number of data points) **do**
$\mathcal{X}_{\text{data}} = \mathcal{X}_{\text{BVP}}[i : i + n_{in}]$
$\mathcal{Y}_{\text{data}} = \mathcal{X}_{\text{BVP}}[i + n_{\text{in}} : i + n_{in} + n_{\text{out}}]$
end for
$\mathcal{X}_{\text{train}}, \mathcal{Y}_{\text{train}}$=split($\mathcal{X}_{\text{data}}, \mathcal{Y}_{\text{data}}$, train split $= 80\%$)
$\mathcal{X}_{\text{test}}, \mathcal{Y}_{\text{test}}$=split($\mathcal{X}_{\text{data}}, \mathcal{Y}_{\text{data}}$, test split $= 20\%$)
Model($\mathcal{X}_{\text{train}}, \mathcal{Y}_{\text{train}}$) (training step)
Evaluate($\mathcal{X}_{\text{test}}, \mathcal{Y}_{\text{test}}$) (evaluate using test data)

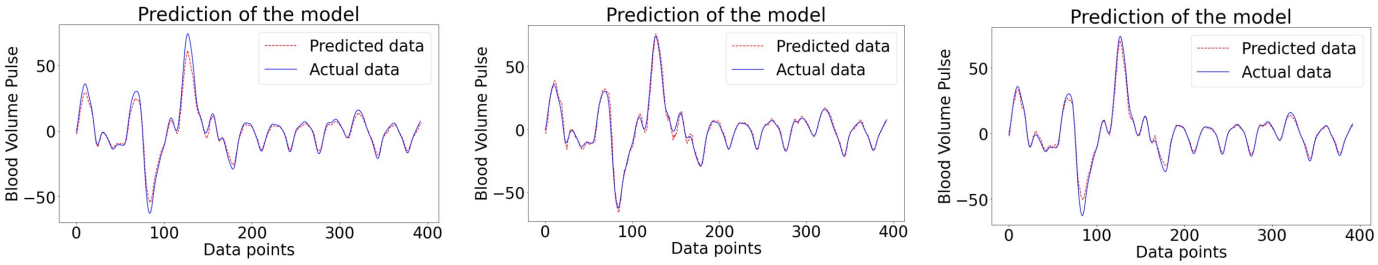
VI. Results and Discussions

A. Model 1: Using LSTM only

Since BVP data is a time-series data, a sequential model (here LSTM) was initially employed for the forecasting of the time series (see Figure 2a for model architecture). The duration 't' of the training data used to make the forecast was $2 < t < 3$ minutes. This training data was then used to forecast the time for the next t where $t \in [2, 3]$ minutes. In Figure 2a, the dimension of the input is $l \times 1$ where 'l' is the dimension of the input and 'o' is the dimension of the output which is the number of steps to predict in the future. We have the value of 'l' as 30 in this paper since that is approximately the Interbeat Interval (IBI) as is shown in Figure 3. The reason for using an input dimension comparable to the value of the IBI is based on previous research in the literature that has shown the merit of using IBI for emotion recognition [26] and heart-rate variability [4]. We evaluated the performance of the model keeping the input dimension the same and varying the output dimension to see our model performance. Figure 4a shows the R2 scores (coefficient of determination) on the test data of different output dimensions being used for the LSTM model. Further, Figure 4d shows how the R2 score decreases with increasing output dimensions.

B. Model 2: Using CNN only

From the previous section (Section VI-A), the max R2 score obtained was found to be 0.856. So, we tried a CNN approach to extract features from the univariate BVP data to make forecasts. The model architecture has been shown in Figure 2b.



(a) Forecast visualization on subset of test data for LSTM model.

(b) Forecast visualization on subset of test data for CNN model.

(c) Forecast visualization on subset of test data for CNN+LSTM model.

Fig. 5: Forecast visualization on subset of test data.

Further, the model's R2 score on test data for different output parameters can be seen from Figure 4c, and the variation of the output dimension on the R2 score can be seen in Figure 4e. Since the data used is a univariate time series data, we have used Conv1D layers from Keras to extract features from the BVP data [24]. The motivation behind using CNN for time series data can be attributed to their ability to extract spatial features well [29]. Hence, the formulation of forecasting used in this paper allows us to leverage this property of the CNNs. The training data used for our CNN model is the same as for the LSTM case too. The training constitutes data from the first 't' minutes where $t \in [2, 3]$ and can forecast the series for the next three minutes. However, in this case, the use of CNNs boosts up the R2 score for the forecasts to 0.915 as compared to just 0.856 for the best choice of output dimension for the LSTM network.

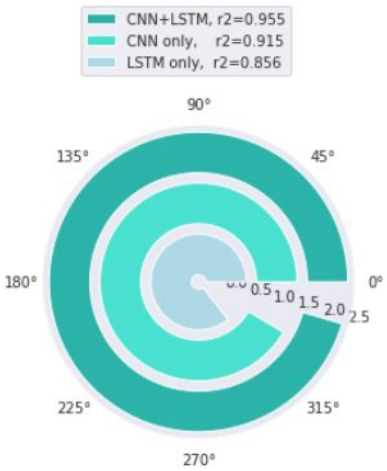


Fig. 6: Comparing the best R2 scores of each model.

C. Model 3: Using CNN+LSTM network

The architecture used for this network has been shown in Figure 2c. As can be seen from the figure, it combines the architecture of both Model 1 and Model 2. This has an advantage as the CNN captures the spatial features of the univariate BVP data and the LSTM layers capture the temporal sequence of the data. This advantage can be seen from the increased R2 score (0.955) for the output dimension two as compared to the other models used. This R2 score is calculated over the forecasted data for the next 3 minutes given 't' minutes of training data where $t \in [2, 3]$.

To further compare the predictions of the forecast models discussed above on the subset of the test data, the outputs of each of the models has been shown in Figure 5 and their best R2 scores have been shown in Figure 6.

It can be clearly seen from Figure 6 that the CNN+LSTM model performs the best as a forecast model for the data set considered in this paper since its R2 score is higher than the other models used.

VII. Limitations and Future work

This paper has focused only on a univariate time series data, which in this case is the Blood Volume Pulse signal. In future work, we would like to use multivariate time series data for the forecast model. This would include the use of other signals from the E4 (Electrodermal Activity (EDA) for the Skin Conductance Level (SCL)), the temperature of the body, and accelerations.

In addition to the use of physiological signals, we would like to extend our current work to video frame predictions and concatenate it with the physiological signal forecast. This is because the visual modality has been incorporated to general engagement perception as in [11], [17] and also has been linked to heart rate measurement as has been shown in [22] [23]. Lastly, we would like to incorporate the Transformers network architecture for time series forecast and compare it to our current approach [30].

VIII. Conclusion

In this paper, we presented three network architectures that provided the first step towards engagement forecast in the context of therapy using social robots for children with ASD. The motivation for this work is based on the need for a real-time engagement feedback as has been mentioned by the authors in [17], [20], [28] and from the need to predict challenging or non-compliant behaviors during therapy [9], [18] so as to make interventions more personalized and adaptable [11], [13], [17], [18], [20].

Our current approach can take in less than three minutes of Blood Volume Pulse data amd can forecast the values of BVP for the next three minutes. Three deep learning model architectures were used to achieve this: LSTM, CNN and a combination of both. Among these models, it was found that the CNN+LSTM model outperformed the other approaches where just LSTM or just CNN was used. This work forms the basis of our future work on forecasting engagement in real time during a live therapy session in the presence of socially-assistive, adaptive robot.

IX. Acknowledgements

The authors wish to acknowledge student researchers Janet Pulgares Soriano, Nathaniel Dugan, Jacob Adair, Aamira Shah, and Fareed Haidar for their assistance with collecting the data. We also want to thank the staff at the Norton Children's Autism Center, the subjects, and their families.

References

[1] Ahmad Qadeib Alban, Malek Ayesh, Ahmad Yaser Alhaddad, Abdulaziz Khalid Al-Ali, Wing Chee So, Olcay Connor, and John-John Cabibihan. Detection of challenging behaviours of children with autism using wearable sensors during interactions with social robots. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 852–857. IEEE, 2021.

[2] Fady Alnajjar, Massimiliano Cappuccio, Abdulrahman Renawi, Omar Mubin, and Chu Kiong Loo. Personalized robot interventions for autistic children: an automated methodology for attention assessment. *International Journal of Social Robotics*, 13(1):67–82, 2021.

[3] A American Psychiatric Association, American Psychiatric Association, et al. *Diagnostic and statistical manual of mental disorders: DSM-5*, volume 10. Washington, DC: American psychiatric association, 2013.

[4] Riccardo Barbieri, Eric C Matten, AbdulRasheed A Alabi, and Emery N Brown. A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability. *American Journal of Physiology-Heart and Circulatory Physiology*, 288(1):H424–H435, 2005.

[5] Momotaz Begum, Richard W Serna, David Kontak, Jordan Allspaw, James Kuczynski, Holly A Yanco, and Jacob Suarez. Measuring the efficacy of robots in autism therapy: How informative are standard hri metrics'. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 335–342, 2015.

[6] Jason Brownlee. *Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end*. Machine Learning Mastery, 2016.

[7] John-John Cabibihan, Hifza Javed, Marcelo Ang, and Sharifah Mariam Aljunied. Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *International journal of social robotics*, 5(4):593–618, 2013.

[8] Hoang-Long Cao, Pablo G Esteban, Madeleine Bartlett, Paul Baxter, Tony Belpaeme, Erik Billing, Haibin Cai, Mark Coeckelbergh, Cristina Costescu, Daniel David, et al. Robot-enhanced therapy: Development and validation of supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE robotics & automation magazine*, 26(2):49–58, 2019.

[9] Matthew S Goodwin, Ozan Özdenizci, Catalina Cumpanasoiu, Peng Tian, Yuan Guo, Amy Stedman, Christine Peura, Carla Mazefsky, Matthew Siegel, Deniz Erdoğmuş, 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*, pages 201–207, 2018.

[10] Tiffany L Hutchins and Patricia A Prelock. Using communication to reduce challenging behaviors in individuals with autism spectrum disorders and intellectual disability. *Child and Adolescent Psychiatric Clinics*, 23(1):41–55, 2014.

[11] Shomik Jain, Balasubramanian Thiagarajan, Zhonghao Shi, Caitlyn Clabaugh, and Maja J Matarić. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Science Robotics*, 5(39):eaaz3791, 2020.

[12] Amanda N Kelly, Judah B Axe, Ronald F Allen, and Russell W Maguire. Effects of presession pairing on the challenging behavior and academic responding of children with autism. *Behavioral Interventions*, 30(2):135–156, 2015.

[13] Changchun Liu, Karla Conn, Nilanjan Sarkar, and Wendy Stone. Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE transactions on robotics*, 24(4):883–896, 2008.

[14] Johnny L Matson and Marie Nebel-Schwalm. Assessing challenging behaviors in children with autism spectrum disorders: A review. *Research in Developmental Disabilities*, 28(6):567–579, 2007.

[15] Victor Monroy. *A Wizard-of-Oz Study to Determine the Efficacy of an Automated Prompting System for Children with Autism*. University of Toronto (Canada), 2010.

[16] Paola Pennisi, Alessandro Tonacci, Gennaro Tartarisco, Lucia Billeci, Liliana Ruta, Sebastiano Gangemi, and Giovanni Pioggia. Autism and social robotics: A systematic review. *Autism Research*, 9(2):165–183, 2016.

[17] Ognjen Rudovic, Jaeryoung Lee, Miles Dai, Björn Schuller, and Rosalind W Picard. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, 3(19):eaao6760, 2018.

[18] Mohammad Nasser Saadatzi, Robert C Pennington, Karla C Welch, and James H Graham. Small-group technology-assisted instruction: Virtual teacher and robot peer for individuals with autism spectrum disorder. *Journal of autism and developmental disorders*, 48(11):3816–3830, 2018.

[19] Sarah Sarabadani, Larissa C Schudlo, Ali Akbar Samadani, and Azadeh Kushski. Physiological detection of affective states in children with autism spectrum disorder. *IEEE Transactions on Affective Computing*, 11(4):588–600, 2018.

[20] Brian Scassellati, Laura Boccanfuso, Chien-Ming Huang, Marilena Mademtzi, Meiying Qin, Nicole Salomons, Pamela Ventola, and Frederick Shic. Improving social skills in children with asd using a long-term, in-home social robot. *Science Robotics*, 3(21):eaat7544, 2018.

[21] Michał Stolarz, Alex Mitrevski, Mohammad Wasil, and Paul G Plöger. Personalized behaviour models: A survey focusing on autism therapy applications. *arXiv preprint arXiv:2205.08975*, 2022.

[22] Arvind Subramaniam and K Rajitha. Estimation of the cardiac pulse from facial video in realistic conditions. 2019.

[23] Arvind Subramaniam and K Rajitha. Spectral reflectance based heart rate measurement from facial video. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3362–3366. IEEE, 2019.

[24] Keras Team. Keras documentation: Conv1d layer.

[25] Serge Thill, Cristina A Pop, Tony Belpaeme, Tom Ziemke, and Bram Vanderborght. Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook. *Paladyn*, 3(4):209–217, 2012.

[26] Gaetano Valenza, Luca Citi, Antonio Lanatà, Enzo Pasquale Scilingo, and Riccardo Barbieri. A nonlinear heartbeat dynamics model approach for personalized emotion recognition. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2579–2582. IEEE, 2013.

[27] Michael Villano, Charles R Crowell, Kristin Wier, Karen Tang, Brynn Thomas, Nicole Shea, Lauren M Schmitt, and Joshua J Diehl. Domer: A wizard of oz interface for using interactive robots to scaffold social skills for children with autism spectrum disorders. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 279–280. IEEE, 2011.

[28] Karla Conn Welch. Physiological signals of autistic children can be useful. *IEEE Instrumentation & Measurement Magazine*, 15(1):28–32, 2012.

[29] Chao Yang, Wenxiang Jiang, and Zhongwen Guo. Time series data classification based on dual path cnn-rnn cascade network. *IEEE Access*, 7:155304–155312, 2019.

[30] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11106–11115, 2021.