Behavior-based Risk Detection of Autism Spectrum Disorder Through Child-Robot Interaction

Hifza Javed
Department of Biomedical Engineering
The George Washington University
Washington, DC, USA
hifzajaved@gwu.edu

Chung Hyuk Park
Department of Biomedical Engineering
The George Washington University
Washington, DC, USA
chpark@gwu.edu

ABSTRACT

This work presents a method to identify children at risk for Autism Spectrum Disorder using behavioral data extracted from video analysis of child-robot interactions. Robots were used as a tool to elicit social engagement from the children in order to capture their social behaviors. A Convolutional Neural Network was used to classify the behavioral data as either at-risk ASD or Typical Development. The network performance was compared to two machine learning classifiers and the utility of the proposed method as a way to streamline existing diagnostic procedures was discussed.

CCS CONCEPTS

• **Computer systems organization** → Robotics • **Computing methodologies** → Machine learning • **Human-centered computing** → User studies

KEYWORDS

Autism spectrum disorder, child-robot interaction, multi-modal behaviors, deep learning, machine learning, diagnosis

ACM Reference format:

Hifza Javed and Chung Hyuk Park. 2020. Behavior-based Risk Detection of Autism Spectrum Disorder through Child-Robot Interaction. In *Proceedings of ACM/IEEE Human-Robot Interaction conference (HRI'20). ACM/IEEE, Cambridge, UK, 2 pages.* https://doi.org/10.1145/1234567890

1 Introduction

Children with Autism Spectrum Disorder (ASD) typically experience difficulties in social communication and interaction. As a result, they display a number of distinctive behaviors including atypical facial expressions and repetitive behaviors such as hand flapping and rocking. Given the subjective, cumbersome and time intensive nature of the current methods of diagnosis, this paper presents a behavior-based approach to identify children at risk for ASD. This can serve as a means to streamline the standard diagnostic procedures by facilitating rapid detection and clinical prioritization of at-risk children.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3378382

Previous approaches to use machine learning for ASD diagnosis include classifications based on data from standardized ASD assessments [1][2] and brain image datasets [3]. [4] used kinematic analysis of upper body movements during a specialized task to detect ASD-specific behaviors.

The current work uses facial expressions and upper body movement patterns to detect ASD. We designed a robot-assisted interaction where robots were used as a tool to elicit social engagement from the children in order to capture their social behaviors. We collected multimodal behavioral data from the child-robot interactions to train a Convolutional Neural Network (CNN) in order to evaluate the utility of proposed approach for ASD risk detection in children.

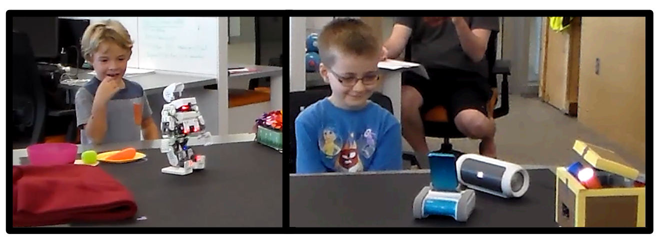

Figure 1: Children interacting with the robots within a sensory maze setup.

2 Methods

2.1 Child-robot interaction design

The interaction scenario for this work used two robots (a small humanoid and an iPod-based mobile robot) in a sensory maze setup where they walked around the maze to find different stations and interacted with the sensory stimuli presented at each station [5]. The activity was designed to include various opportunities for conversations initiated by the robot, encouraging active participation from the child to facilitate a joint sensory experience (Figure 1).

2.2 User study

A user study was conducted to collect a video dataset from two groups of children, typically developing (TD) and with ASD, between the ages of 5 and 10 years. The study was approved by the Institutional Review Board (GW IRB#111540). The TD group consisted of 7 boys with a mean age of 7.43±2.30 years. The ASD group consisted of 5 boys with a mean age of 8.2±1.10 years. The total interaction time was 3931 seconds (mean=561.57) for the TD group and 3442 seconds (mean=688.4) for the ASD group making the two datasets comparable in size.

2.3 Multimodal data and feature selection

Since the robots were set up on a table with the children either standing or seated around it, we extracted only upper body tracking data and facial keypoints from video recordings of the interactions using OpenPose [6]. We then used Laban movement analysis [7] to derive features that can be used to analyze the intent behind human movement, which included the *weight*, *space* and *time* features, as defined in [8][9]. This was done using a moving time window of 1 second to capture the temporal nature of the data. The 3 derived movement features were combined with 68 facial key-points (originating from the nose and eyes) to form a dataset comprising a total of 71 features.

2.4 Network Architecture

A Convolutional Neural Network (CNN) was designed to process the dataset given the usefulness of the CNN to process ordered data (Figure 2). The network had 2 1D convolution layers to extract high-level features from the temporal data. Given the non-linear data structure, the first 2 dense layers were used to spread the feature dimensions and the last was used to generate the output. To avoid overfitting, the four dropout layers were used at a dropout rate of 20%.

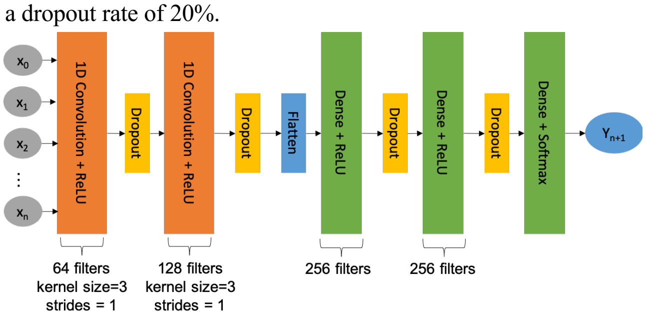

Figure 2. Our CNN used for the ASD risk detection.

3 Results

The goal of ASD risk detection was modeled as a binary classification problem. The CNN was trained on 80% of the interaction data and the remaining 20% were used to validate its performance. The CNN achieved a training accuracy of 0.883 and a training loss of 0.232. Two additional machine learning classifiers (Random Forests [10] and K-Nearest Neighbor [11]) were used to situate the performance of the CNN by comparing their accuracy, precision, and recall values (Table 1). These

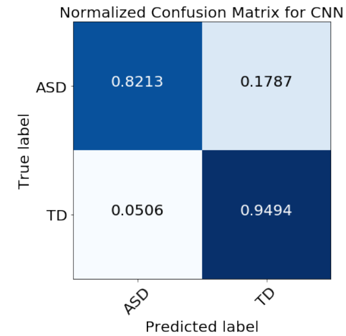

Figure 3. Confusion matrix for classifications from the CNN.

machine learning classifiers have simpler structures than a deep CNN with fewer hyperparameters tat require fine-tuning. Figure 3 shows the confusion matrix resulting from the classifications generated by the CNN.

4 Discussion and conclusion

All 3 classifiers attained high accuracy, precision, and recall values that were close to 0.90 (Table 1). These results are encouraging, given the small size of the training dataset (on average ~18597 data points per subject). Given the false negative value of 0.1787 (Figure 3), this method cannot be claimed to be ready to be used singularly as a diagnostic tool for ASD. However, this proposed behavioral approach can be useful as a more accessible layer of screening to identify at-risk children and streamline the diagnosis process.

The three classifiers attained similar overall performance, indicating that the simpler machine learning methods suffice for the given dataset. However, with larger number of participants and repeated interaction sessions, the size of the dataset is expected to grow rapidly. With the incorporation of additional modalities to use a more comprehensive representation of ASD behaviors, complexity of the dataset is bound to increase, in which case a CNN may stand out as the better choice.

Of course, this work is limited by the number of participants, the inclusion of a single gender and the exclusive use of behaviors available from video data. Inclusion of other complex behaviors observable through physiological, eye gaze and vocal data can help to improve the reliability of the networks.

Unlike previous works, however, this paper presents a method to identify at-risk children with ASD based only on behavioral data captured through video recordings of a naturalistic interaction with social robots. The movement of the child was not restricted and no obtrusive sensors were used. This means that this method can easily be generalized to other interactions from, for example, play time at home, where much larger datasets can be obtained.

Table 1. Performance metrics for the three classifiers.

Classifier	Accuracy	Precision	Recall
CNN	0.8846	0.8912	0.8853
RF	0.8852	0.8876	0.8856
KNN	0.8874	0.8920	0.8878

REFERENCES

[1] Wall, Dennis Paul, J. Kosmicki, T. F. Deluca, E. Harstad, and Vincent Alfred Fusaro. "Use of machine learning to shorten observation-based screening and diagnosis of autism." Translational psychiatry 2, no. 4 (2012): e100.

[2] Kosmicki, J. A., V. Sochat, M. Duda, and D. P. Wall. "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning." Translational psychiatry 5, no. 2 (2015): e514.

[3] Heinsfeld, Anibal Sólon, Alexandre Rosa Franco, R. Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi. "Identification of autism spectrum disorder using deep learning and the ABIDE dataset." NeuroImage: Clinical 17 (2018): 16-23.

[4] Crippa, Alessandro, Christian Salvatore, Paolo Perego, Sara Forti, Maria Nobile, Massimo Molteni, and Isabella Castiglioni. "Use of machine learning to identify children with autism and their motor abnormalities." Journal of autism and developmental disorders 45, no. 7 (2015): 2146-2156.

[5] Javed, Hifza, Rachael Burns, Myounghoon Jeon, Ayanna M. Howard, and Chung Hyuk Park. "An Interactive Robotic Framework to Facilitate Sensory Experiences for Children with ASD." arXiv preprint arXiv:1901.00885 (2019).

[6] Cao, Zhe, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." arXiv preprint arXiv:1812.08008 (2018).

[7] Groff, Ed. "Laban movement analysis: Charting the ineffable domain of human movement." Journal of Physical Education, Recreation & Dance 66, no. 2 (1995): 27-30.

[8] Masuda, Megumi, Shohei Kato, and Hidenori Itoh. "Emotion detection from body motion of human form robot based on laban movement analysis." In International Conference on Principles and Practice of Multi-Agent Systems, pp. 322-334. Springer, Berlin, Heidelberg, 2009.

[9] Wakayama, Yuki, Seiji Okajima, Shigeru Takano, and Yoshihiro Okada. "IEC-based motion retrieval system using Laban movement analysis." In International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, pp. 251-260. Springer, Berlin, Heidelberg, 2010.

[10] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[11] Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). K-nearest neighbor classification. In Data mining in agriculture (pp. 83-106). Springer, New York, NY.