# Predicting the ADOS-2 Calibrated Severity Score From Video And Audio Analysis

Andras Sarkany[1 3], Khaleel Kamal[1], Emily Dillon[2], Edith Ocampo[2], Sarely Licona[2], Allie Wainer[2], Latha Soorya[2], Zoltan Toser[1]
(1)Argus Cognitive, Inc., Lebanon, NH, (2)Department of Psychiatry, Rush University Medical Center, Chicago, IL, (3)Department of Software Technology & Methodology, Eötvös Loránd University, Budapest, Hungary

## Background

- Ongoing strides in robust non-intrusive methods allow for the application of computational behavioral assessments in clinical settings, mitigating tedious manual coding processes.
- Using machine learning components, automated behavioral assessment aims to facilitate the detailed coding of naturalistic behavior (including prosody, facial landmarks, gaze, and pose estimations).
- Our previous work has focused on demonstrating the reliability of automated behavior analysis during ADOS-2 diagnostic sessions. Since our previous studies, we have further refined most of the existing components and added new features such as the characterization of eye blinks and the estimation of social initiation/avoidance.

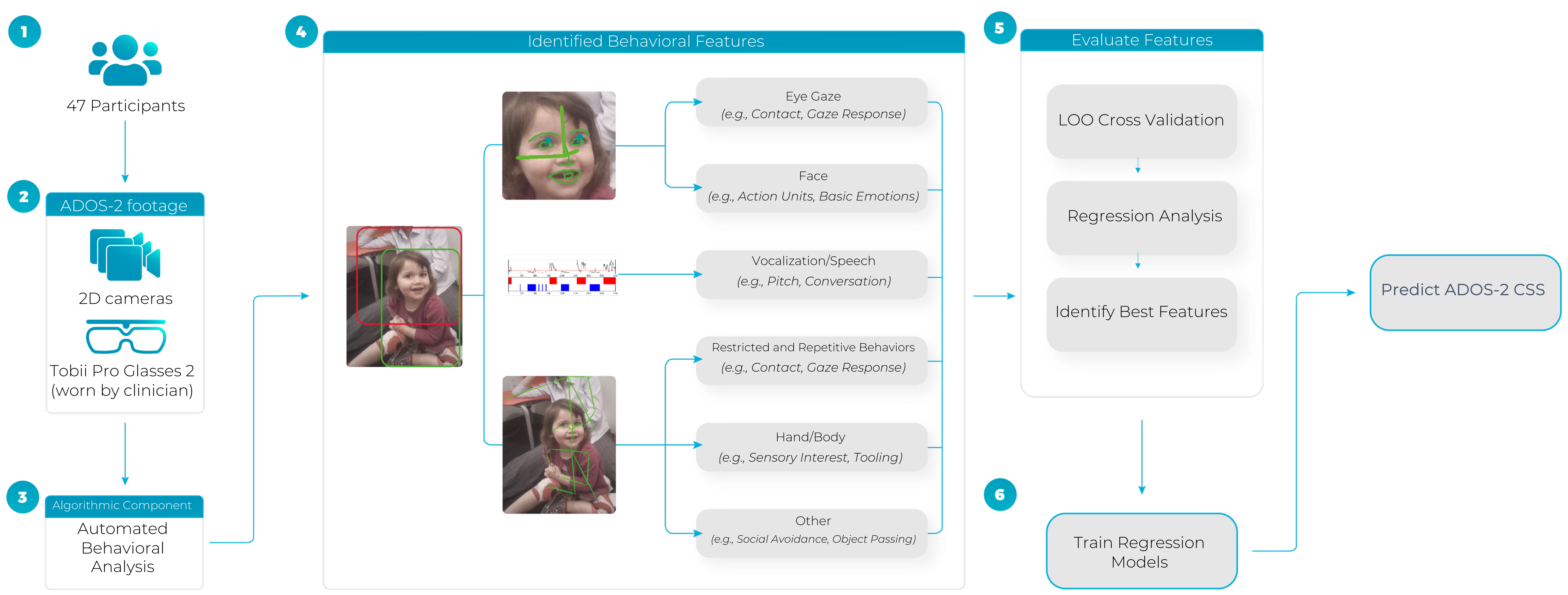
## Procedure and Methodology

**Table 1: Participant Demographics**

| | NDD | TD |
|---|---|---|
| Age in Years (mean, SD) | 8.5 (±2.9) | 6.4 (±2.4) |
| Sex (N, % Female) | 47 (40.4%) | 13 (53.8%) |
| FSIQ (mean, SD) | 76.5 (±31.7) | N/A |
| ADOS-2 CSS (mean, SD) | 6.3 (±2.7) | 1.7 (±2) |

**Table 2: Variable Domains**

| Domain Name | Number of Metrics Generated per Domain |
|---|---|
| Eye Gaze (e.g., Contact, Gaze Response) | 19 |
| Face (e.g., Action Units, Basic Emotions) | 27 |
| Vocalization/Speech (e.g., Pitch, Conversation) | 38 |
| Restricted and Repetitive Behaviors (e.g., Gestures) | 12 |
| Hand/Body (e.g., Sensory Interest, Tooling) | 19 |
| Other (e.g., Social Avoidance, Object Passing) | 28 |




1) 60 participants were recruited, 47 with neurodevelopmental disorders (NDD), including ASD, suspected ASD, and monogenic syndromes associated with ASD, and 13 age-matched typically developing controls (TD).
2) ADOS-2 footage was collected using off-the-shelf 2D cameras and Tobii Pro Glasses 2 worn by the examiner.
3) Argus MDS was used to analyze the footage, automatically identifying patients, analyzing their vocalizations, eye movements, facial expressions, and body pose.
4) Based on discriminative power between the NDD participants and TD controls, 32 features were selected by Research-reliable ADOS-2 raters, from a total of 143 estimated variables (Table 2).
5) The identified features' importance to predict ADOS-2 CSS was evaluated using a hill climbing algorithm combined with a leave-one-out cross-validation and regression analysis.
6) Four types of regression models (random forest, support vector regressor, k-NN, linear regression) were trained to predict ADOS-2 CSS.

## Objective

- Utilize a state-of-the-art automated behavioral assessment pipeline to objectively analyze behavior during autism diagnostic evaluations (within the ADOS-2).
- As subjects interact with the examiner and the environment, automatically detect and estimate detailed metrics across multiple social communication domains, such as speech, emotional valence, and eye contact.
- Evaluate correlations between the variables from the automated behavior analysis system and the relevant clinical variables.
- Utilize the identified variables to assess the predictive power of biometric data for estimating Autism Diagnostic Observation Schedule-2 (ADOS-2) Calibrated Severity Score (CSS).

## Results

The model's predictive performance was measured using root mean squared error (RMSE), mean absolute error (MAE), and R-squared ($R^2$).

To put our results into perspective: Four baseline results were utilized for each regressor type:

1. A model that predicts CSS scores randomly.
2. A model that predicts the median CSS score (5) in all cases.
3. A regression model that uses just age and gender as independent variables.
4. A regression model including age, gender, and FSIQ.

Figure 1 shows the $R^2$ performance of the k-Nearest Neighbors (k-NN) model, including the baseline methods.

**Figure 1:** k-NN Results – $R^2$ (Higher is Better)

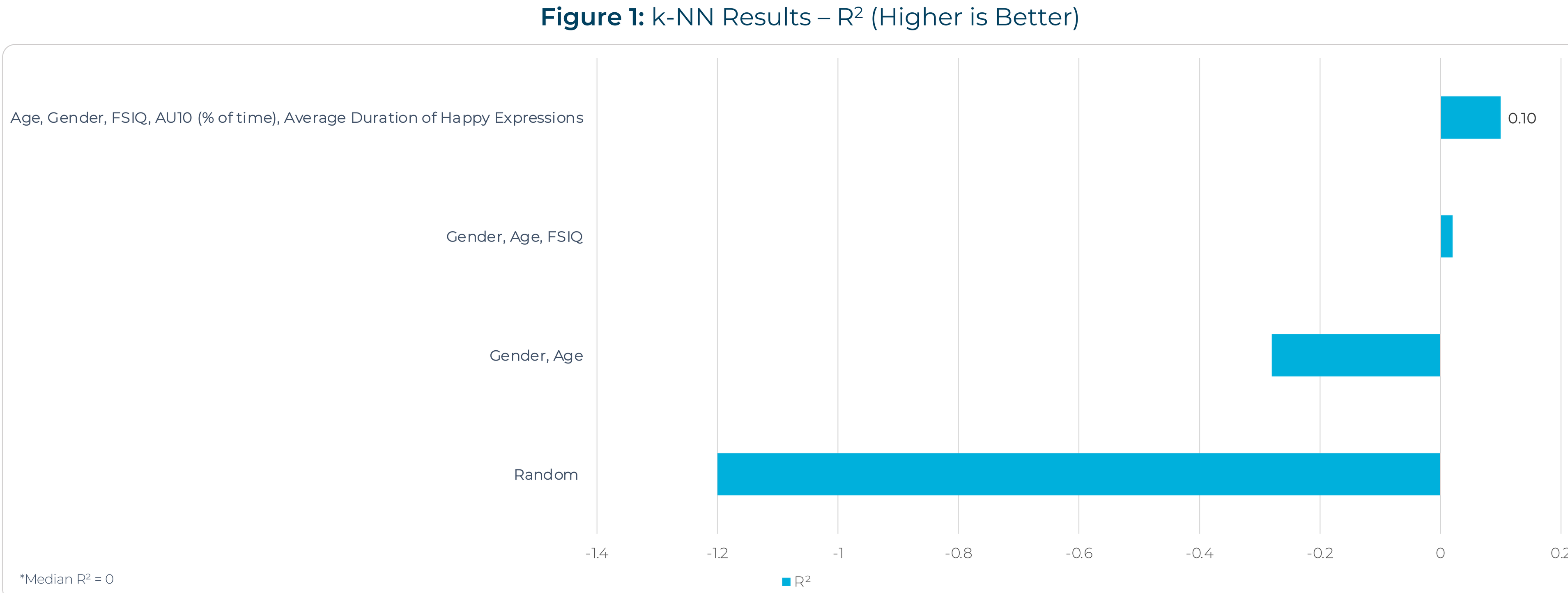



**Table 3: Final Feature Subset**

| Model | Optimal Feature Subset | $R^2$ |
|---|---|---|
| Random Forest | Age, Gender, FSIQ, Eye Contact (% of time), AU10 (% of time), Eye Contact and Sadness Links (% of time) | 0.04 |
| k-Nearest Neighbors (k-NN) | Age, Gender, FSIQ, AU10 (% of time), Average Duration of Happy Expressions | 0.10 |
| Support-Vector Regression (SVR) | Gender, Age, FSIQ, Negative Emotions (% of time), AU1 (% of time), Patient Speech (Average Pause Lengths) | 0.04 |
| Linear Regression | Gender, Age, FSIQ, Vocal Interchanges (% of time), Pitch Variability in Vocalizations | 0.01 |

## Conclusion and Future Works

- The Argus MDS System yields metrics of key social-communication behaviors, including gaze, emotional valence, and vocal communicative abilities.
- Results suggest introducing biometric social markers increases predictive power and reduces the error rate for estimating ADOS-2 CSS. Adding just two variables related to facial expressions increases the explained variability by 10% over demographic variables.
- To further advance understanding of the role of nonintrusive computer-vision–based behavior analysis in clinical settings, we are working on adding speech recognition capabilities to the system and refining gestural analysis.
- Future analyses will include additional components and dependent variables with broader range than the CSS (e.g. Socialization subscale of Vineland Adaptive Behavior Scales-3).
- To determine the generalizability of our approach, future research will include a multicenter, international study that comprises a variety of demographics.

RUSH UNIVERSITY MEDICAL CENTER
ARGUS Cognitive