2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) | 979-8-3315-4090-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/UEMCON62879.2024.10754700

Enhancing Early Diagnosis of Autism With Machine Learning Algorithms Using Postural Control Features

Maya Nachega
Dept. of Biomedical Engineering
Vanderbilt University
Nashville, USA
maya.s.nachega@vanderbilt.edu

Sugru A. Yahaya
Weldon School of Biomedical Engineering
Purdue University
West Lafayette, USA
syahaya@purdue.edu

Ila Wallace
Dept. of Computer Science
Texas State University
San Marcos, USA
iew2@txstate.edu

Ting Liu
College of Graduate Studies
Texas A&M University-San Antonio
San Antonio, USA
tliu@tamusa.edu

Damian Valles
Ingram School of Engineering
Texas State University
San Marcos, USA
dvalles@txstate.edu

Rachel M. Koldenhoven
Dept. of Health & Human Performance
Texas State University
San Marcos, USA
rmr214@txstate.edu

Yumen Li
Dept. of Health & Human Performance
Texas State University
San Marcos, USA
yumeng.li@txstate.edu

***Abstract*—Autism Spectrum Disorder (ASD) is a common neurodevelopmental disorder whose biological cause is still not well understood. Due to this, as well as the gradual onset patterns of the disorder's symptoms, early identification and diagnosis of ASD proves to be challenging. This study aimed to provide a potential early diagnostic tool for ASD by training machine learning classifiers with linear and nonlinear postural control data features. For the research methodology, center of pressure (COP) data for 38 children ages 5-16 was collected using a force plate during both eyes opened and eyes closed conditions. Each trial was 20 seconds, and three trials were completed for each condition. The raw data was preprocessed, and six COP variables were calculated for each trial: anteroposterior (AP) displacement, mediolateral (ML) displacement, elliptical sway area, COP travel distance, AP multiscale entropy complexity, and finally, ML multiscale entropy complexity. After preprocessing, these data features were trained to seven machine learning models to classify participants as having ASD or typical development (TD). Our experiment showed that all seven machine learning models could appropriately classify participants with an accuracy of 79% and above. The best classification performance had an accuracy of almost 97% and was done by the model that utilized the random forest algorithm for classification. Additionally, the data feature that was the most important across machine learning algorithms was AP complexity. These results have strengthened the validity of a machine learning approach for early ASD diagnosis. In future research, more data should be collected beyond the 38 participants obtained in this study, or data augmentation should be done to generate more data to train the models. Future research should also put efforts into obtaining more robust and representative datasets, exploring best multiscale entropy parameter values, and exploring why certain data features influence classification more than others.**

Keywords—*Autism Spectrum Disorder, machine learning, postural control, data features, early diagnosis*

This work was funded by the National Science Foundation – Research Experiences for Undergraduates (NSF-REU) under Grant #2150135.

I. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects 1.85% of children in the US, and its prevalence is increasing globally [1]. However, there is no direct clinical test for diagnosing ASD. Instead, ASD is diagnosed by observing behavior characterized as deficits in social communication, excessively repetitive behaviors, restricted interests, and an insistence on sameness [2]. Children with ASD also commonly experience motor skill deficits manifesting as atypical gait and poor postural control [3].

Postural control is maintaining the balance of the body's center of pressure (COP). An individual's COP is the projection of their center of mass onto the ground, and stability is achieved when the COP is within their base of support, which during standing includes both feet and the space between them [4]. It is vital to the development of children because children with better stability are more likely to lead more active and social lifestyles. However, literature has established the fact that children with ASD experience reduced postural stability [5]. Postural control and stability are mainly regulated by the vestibular, visual, and somatosensory systems [6]. A standard measure of these systems' effectiveness is using a kinetic device such as a force platform [6]. Using this device allows for ease in tracking the COP of a subject. COP tracking is essential to the early diagnosis of ASD in young children, as many studies have found that children with ASD have more significant sway displacement, sway areas, and sway velocities compared to their neurotypical counterparts [5].

Children with ASD have been found to have low postural complexity compared to age-matched children without ASD [7]. Complexity is a nonlinear measure of a system's irregularity. To find a system's complexity, the entropy must first be calculated. There are many methods for calculating entropy, but multiscale entropy is preferable to other methods because it quantifies the overall complexity of a system over multiple time scales, which reflects the dynamics of the human body [7].

To create a tool to assist clinicians with ASD diagnoses, we trained machine learning algorithms to detect if a child has ASD based on their postural stability. Machine learning was used because of its ability to handle large and irregular datasets [8]. In addition, several previous studies have used simple supervised machine learning algorithms to detect postural characteristics [9][10]. Each of the many different types of machine learning algorithms has its benefits and detriments, so we selected several to evaluate the effectiveness of the models. This study aimed to create an early diagnostic tool that uses both linear and nonlinear measures of postural stability to train a machine learning algorithm to categorize children by whether they had ASD or were neurotypical.

II. Methodology

A. Participants

The participants in this study were children who attended an Autism Summer Camp and typical development (TD) children from the surrounding community between 2021 and 2023. The ages of the children ranged from 5 to 16. There were 38 total participants analyzed, and the two types of participants in this study were categorized as follows:

1) Children with a formal ASD diagnosis from a licensed professional represented the ASD group.

2) Children with typical neurodevelopment represented the TD group and served as the control.

In the ASD group, the sex distribution was two girls and twenty-four boys. For the TD group, the sex distribution was zero girls and twelve boys. The participants' parents and guardians were informed of the study and the risks associated with participating. The participants and their respective parents and guardians consented to their child's participation in the study voluntarily. Participants were also screened for prior health complications or conditions negatively impacting postural control. These participants were not included in the dataset.

B. Equipment Utilized

The participants' COP measurements were collected using a force plate (Bertec Portable Force plate, Columbus, OH) placed on a flat, level surface on top of a blue rubber mat, as shown in Fig. 1. The rubber mat was used to keep the force plate in place and prevent potential scuffs when in contact with the floor. The force plate was connected to a laptop with Bertec Digital Acquire 4.1.20 software installed. This software was used to collect the raw data of each participant, which was the coordinate position of the COP over each 20-second trial. These trials were then exported as Excel files for later data processing.

C. Procedures

Both groups of children were instructed to stand on the force plate without moving to the best of their ability, as shown in Fig. 2. They were to stand on the force plate three times with their eyes open and closed. Each trial lasted 20 seconds. If the child moved during a trial, the data collected was discarded, and the trial was attempted again. If the child no longer wished to participate in the study or was too agitated to continue, they could stop participating.

Fig. 1. Portable Bertec Force Plate resting on the blue rubber mat.

D. Data Analysis

After data collection, the data from each trial was preprocessed to calculate the following values: anteroposterior and mediolateral displacement of center of pressure (COP), the elliptical sway area, the total distance of COP, and the anteroposterior and mediolateral multiscale entropy complexities. The machine learning algorithms used were logistic regression, k-nearest neighbor, decision trees, random forest, gaussian Naïve-Bayes, support vector machine (SVM), and discriminant analysis. The machine learning algorithms were trained using 80% of the collected data. The remaining 20% of the data was used for model performance evaluation. Each algorithm's effectiveness was evaluated by its accuracy score (1) and F1 score (4) to assess the different outcomes of classification when given the test data.

$$Accuracy = \frac{\#\ of\ Correct\ Predictions}{Total\ \#\ of\ Predictions} \tag{1}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{2}$$

$$Recall\ = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{3}$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision +\ Recall} \tag{4}$$

Accuracy and F1 scores were used. F1 scores are calculated when classes do not have equal samples. The four groups for analyzing the algorithm's performance were true positive, false positive, true negative, and false negative. The number of false positives and false negatives and a quick evaluation of model performance were visualized using confusion matrices. The features calculated from the raw data were also analyzed using the importance of permutation features to determine which features were the most influential in classification. Each feature was shuffled 50 times to obtain the mean absolute error value. Permutation feature importance was run on each model.

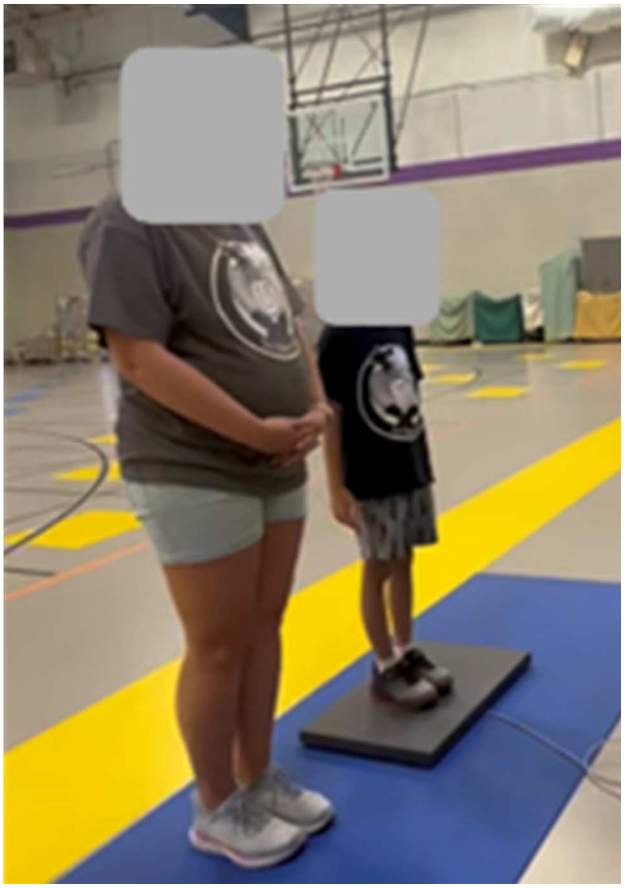

Fig. 2. A participant is standing on the force plate.

III. Results

After the raw data from each participant was preprocessed and the COP values were calculated, they were organized and exported to one master spreadsheet. Table I compares various movement and stability metrics between autism spectrum disorder (ASD) and Typically Developing (TD) groups, detailing means and standard deviations for each. Metrics such as anteroposterior and mediolateral displacement show higher averages and variability in ASD, suggesting more pronounced movement. Elliptical sway area and distance traveled are also noted, with TD individuals traveling further, indicating possibly more controlled movement. Entropy measures, reflecting randomness in movement, are more significant in TD across both anteroposterior and mediolateral directions, as are the complexities associated with these entropies, implying that TD movements exhibit higher variability and complexity than ASD.

TABLE I. Means of Data Features

Data Features	**Mean (Standard Deviation)**	
	ASD	**TD**
Anteroposterior Displacement (m)	0.060 (0.037)	0.033 (0.016)
Mediolateral Displacement (m)	0.065 (0.058)	0.037 (0.015)
Elliptical Sway Area (m^2)	0.004 (0.007)	0.00 (0.001)
Distance Travelled (m)	0.819 (0.638)	2.164 (1.876)
Anteroposterior Entropy	0.289 (0.105)	0.463 (0.224)
Mediolateral Entropy	0.270 (0.111)	0.412 (0.222)
Anteroposterior Entropy Complexity	1.123 (0.407)	2.488 (1.224)
Mediolateral Entropy Complexity	1.050 (0.442)	2.207 (1.536)

After training via stratified k-fold cross-validation, all models could classify participants appropriately as either ASD or TD with accuracy scores above 79% and F1 scores above 84%. Table II lists each of the classifiers, their accuracy scores, and their F1-scores. Table II presents the performance evaluation of various classifiers distinguishing ASD and TD individuals. The evaluation metrics include accuracy and F1 scores, essential indicators of model effectiveness, particularly in datasets where the balance between classes might be a concern. The best-performing model was the random forest model, with an accuracy score of about 97% and an F1-score of 97%. The model demonstrated superior capability in classifying participants, highlighting its effectiveness in handling complex patterns and interactions within the data.

TABLE II. Model Performance Evaluation Via Accuary and F1-Scores

Classifier	**Accuracy Scores**	
	Test Set Accuracy Score	**F1 Score**
Logistic Regression	0.83	0.88
K-Nearest Neighbor	0.86	0.9
Decision Tree	0.93	0.94
Random Forest	0.97	0.97
Naïve-Bayes	0.79	0.84
SVM	0.79	0.86
Discriminant Analysis	0.93	0.95

Fig. 3 contains confusion matrices for several classifiers distinguishing ASD and TD individuals. The Random Forest model shows the most effective performance, accurately identifying all true negatives and most true positives with no false positives and only two false negatives. On the other hand, the Naïve-Bayes and Support Vector Machine classifiers demonstrate more modest results, with higher instances of false positives and false negatives. Logistic Regression and Discriminant Analysis classifiers perform relatively well, with fewer false positives and a balanced detection of true positives and negatives. Decision Tree and K-Nearest Neighbor models also show good efficacy, with a balanced classification performance characterized by a slightly higher number of true positives. Overall, these confusion matrices reveal substantial variations in the effectiveness of each classifier in correctly identifying ASD and TD participants, highlighting the strength of ensemble methods like Random Forest in handling such classification tasks.

Each model was evaluated using the importance of permutation features to determine which feature was the most influential to classification. A ranking system determines which feature was the most influential across all models. Since there were eight total features, features were ranked 0-8 per model, with eight indicating that one feature was the most influential and one being the least influential. Some models did not have results for all features, so those were ranked zero. Table III depicts the rankings across models. Fig. 4 depicts that AP complexity was the most influential data feature for classification across all models tested. Path distance and ML complexity were the second and third most influential, respectively.

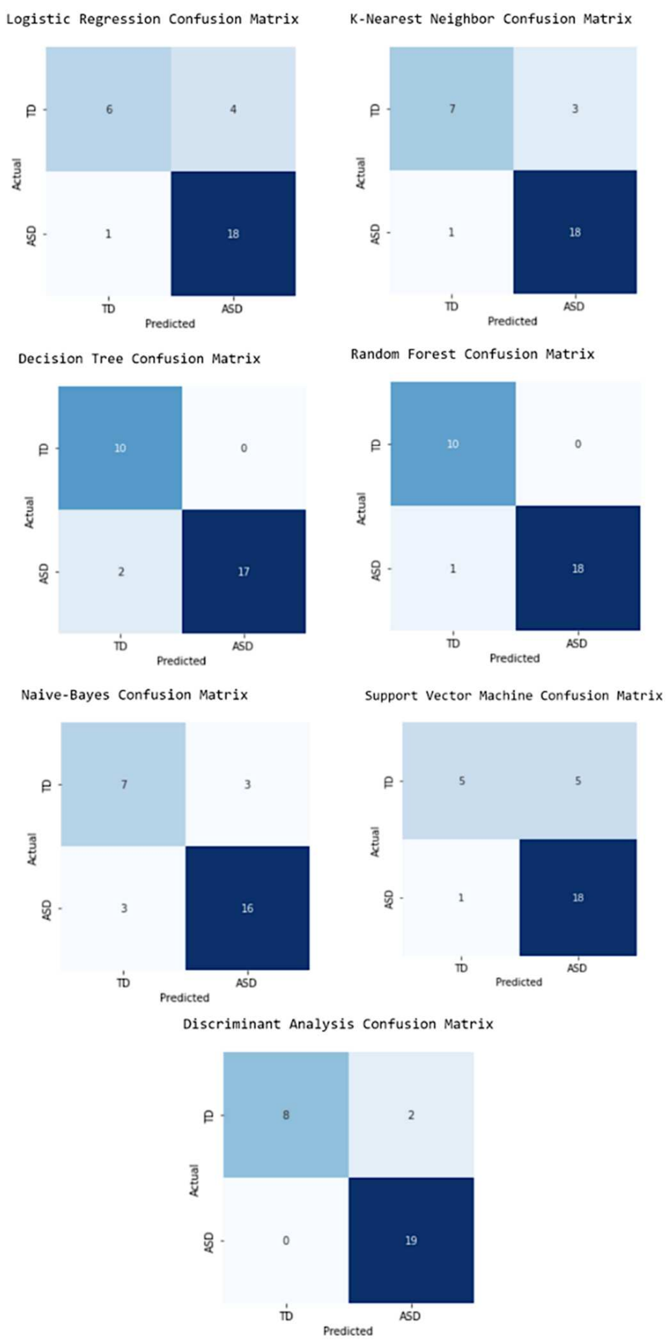

Fig. 3. Confusion Matrices for each of the seven machine learning classifiers analyzed.

TABLE III. RANKED FEATURES IMPORTANCE

Machine Learning Models	Data Features							
	AP Complexity	Path Distance	ML Complexity	ML Displacement	AP Entropy	AP Displacement	Sway Area	ML Entropy
Logistic Regression	8	7	6	0	0	0	0	0
K-Nearest Neighbor	8	7	6	0	4	0	0	5
Decision Tree	6	8	0	7	0	0	0	0
Random Forest	7	8	0	0	0	6	5	0
Naïve-Bayes	8	6	7	4	5	3	1	2
Support Vector Machine	8	7	6	0	0	0	0	0
Discriminant Analysis	8	6	2	5	7	4	3	1
Mean Score	7.57	7	3.86	2.29	2.29	1.86	1.29	1.14

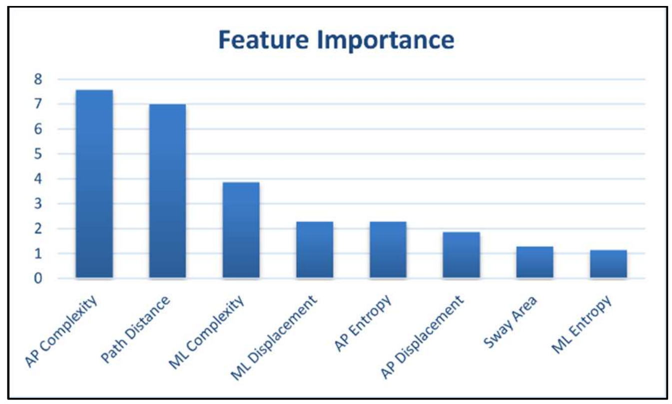

Fig. 4. Data features are ranked by influence using the importance of permutation features.

IV. DISCUSSION

This study used a machine learning approach to examine the differences between children with ASD and TD children using postural control. Our results showed that all the machine learning algorithms selected could accurately differentiate between the groups with an accuracy of at least 79%. All the models were able to distinguish ASD versus TD participants with generally good accuracy, but the best classification performance was seen in the random forest model. Random forest is an easy-to-use and flexible algorithm that tends to yield high accuracy in classification tasks even without hyperparameter fine-tuning. Its strength lies in using many decision trees by splitting various nodes and randomly choosing features to create the decisions. This method is robust and leaves out little to no outliers, which may be why it worked well with our dataset. Even though random forest performed the best in this study, future related research should still employ other algorithms. Every dataset is different; thus, different algorithms might work better or worse for various experiments, even if the datasets present similar data. For instance, in a previous study that assessed postural control data for training machine learning classifiers for ASD diagnosis, the best classification was done by the Naïve-Bayes model [11]. However, in our study, the Naïve-Bayes model was one of the lower-performing classifiers. Therefore, trying several models and comparing classification performance across the chosen algorithms would be good practice in follow-up research.

The implicitly defined parameters were used in our machine learning models since they did not negatively affect the accuracy of any models. However, in future studies, experimentation with hyperparameter fine-tuning may be beneficial due to parameter selection having the ability to impact the performance of a model greatly. Parameter modification may also be helpful if a new data set is employed since a change in data may alter the performance of the models tested. Parameter selection also affects the data preprocessing stage regarding entropy and complexity calculations. The embedding dimension (m) of our sample entropy-based multiscale entropy function was six, and the tolerance (r) was 0.25. These values were based on the recommendations for the COP time series of previous studies, but the preferred m- and r-values are determined case-by-case [12]. Preliminary work to ensure appropriate parameter selection based on each unique dataset can ensure that optimal results are attained and that they are representative.

The results of the permutation feature showed that across all machine learning models, the three most influential data features were AP complexity, path distance, and ML complexity. Compared to the raw data collected, this is a logical conclusion. The difference between the average AP complexity of children with ASD and children with typical development was 1.365, the most significant difference from all features. Path distance had the second most significant difference of 1.345 m, and ML complexity had the third with a difference of 1.157. None of the other six features had more substantial differences than one, reflected in Fig. 4 by the drop between the ML complexity and displacement bars. Although the results of the raw data and the permutation analysis agree, they conflict with previous studies in this field. Multiple prior studies have found that there was no significant difference in AP complexity between children with ASD and children with typical development and found that ML complexity was the feature that had an essential difference between the two groups [7][11]. Various factors, such as different participants, equipment, and testing environments, may explain conflicting study results. Future studies must be done to investigate this discrepancy.

Although the study results were generally positive, some limitations and challenges should be addressed in future research to ensure more progress is made in machine learning for early ASD diagnosis. Firstly, our study had a relatively small dataset, which was the root of many initial challenges in creating well-performing models. In machine learning, the more data a model must learn from, the better it performs. This is because it allows the model to gain a robust understanding of the variability that can be available. Therefore, it can better classify participants that may not fit a particular pattern. A common issue faced with smaller datasets is overfitting, meaning the model becomes too accustomed to the patterns in a particular small set of data and poorly evaluates new/unseen data that it wasn't trained with. Stratified *k*-fold cross-validation training is a common technique to alleviate the issue of overfitting, and it effectively solved our overfitting problem. In the future, it would be beneficial to acquire more participants beyond the 38 participants we had or to perform data augmentation to generate more data from the small amount available.

Another limitation lies in selecting parameters for entropy (as briefly mentioned before). There is no clear set of rules for how parameters are chosen for entropy because the physiological connection between ASD, entropy, and postural control data is not yet clearly understood. Therefore, the certainty of our parameter selections in our dataset could be higher. A great avenue for future research could be exploring parameter selection for multiscale entropy to evaluate postural complexity.

V. Conclusion

In conclusion, this study has strengthened the validity of a machine learning approach to early ASD diagnosis using postural control data features. At the very least, it presents a valid preliminary step to clinical evaluation for ASD diagnosis via the American Psychiatric Association's *Diagnostic and Statistical Manual*, Fifth Edition (DSM-5). In future related research studies, more investigation should be done on the best multiscale entropy parameter values for postural complexity analysis since limited research is available. Furthermore, future studies should also try to obtain more participants as well as a more representative distribution of ASD children. Additionally, the dataset lacked female participants, so having more females could lead to a more robust dataset and better classification results. If obtaining more data and access to children with ASD poses a problem, another suggestion would be to perform data augmentation via an encoder so that the machine learning models can be trained on data generated from a small amount of existing data. This would allow for a more robust data training and validation process for the models and avoid common issues with smaller datasets, such as overfitting.

Overall, the results of our study show that there is strong potential that machine learning classifiers could aid in the streamlined and efficient early diagnosis of ASD in younger children. Many improvements could be made in future research studies that can further validate this approach to ASD diagnosis as well as begin to strengthen knowledge on specific aspects of postural control that best aid in the early diagnosis of ASD. Gaining a clearer understanding of what features most aid in diagnosis can also have many implications for better understanding the biological basis and causes behind ASD.

References

[1] C. Bougeard, F. Picarel-Blanchot, R. Schmid, R. Campbell, and J. Buitelaar, "Prevalence of Autism Spectrum Disorder and Co-morbidities in Children and Adolescents: A Systematic Literature Review," *Front Psychiatry*, vol. 12, 2021, [Online]. Available: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.744709

[2] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, 2013. doi: doi:10.1176/appi.books.9780890425596.

[3] I C. J. Zampella, L. A. L. Wang, M. Haley, A. G. Hutchinson, and A. de Marchena, "Motor Skill Differences in Autism Spectrum Disorder: a Clinically Focused Review," Oct. 01, 2021, *Springer*. doi: 10.1007/s11920-021-01280-6.

[4] S. J. Hall, "Chapter 13. Equilibrium and Human Movement," in *Basic Biomechanics, 5e*, New York, NY: The McGraw-Hill Companies, 2007. [Online]. Available: accessphysiotherapy.mhmedical.com/content.aspx?aid=6063819.

[5] Y. H. Lim, K. Partridge, S. Girdler, and S. L. Morris, "Standing Postural Control in Individuals with Autism Spectrum Disorder: Systematic

Review and Meta-analysis," *J Autism Dev Disord*, vol. 47, no. 7, pp. 2238–2253, Jul. 2017, doi: 10.1007/s10803-017-3144-y.

[6] T. Paillard and F. Noé, "Techniques and Methods for Testing the Postural Function in Healthy and Pathological Subjects," 2015, *Hindawi Publishing Corporation*. doi: 10.1155/2015/891390.

[7] Y. Li, M. A. Mache, and T. A. Todd, "Complexity of Center of Pressure in Postural Control for Children With Autism Spectrum Disorders Was Partially Compromised," *J Appl Biomech*, vol. 35, no. 3, pp. 190–195, doi: 10.1123/jab.2018-0042.

[8] S. Ng, S. Masarone, D. Watson, and M. R. Barnes, "The benefits and pitfalls of machine learning for biomarker discovery," Oct. 01, 2023, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s00441-023-03816-z.

[9] R. Sun, K. L. Hsieh, and J. J. Sosnoff, "Fall Risk Prediction in Multiple Sclerosis Using Postural Sway Measures: A Machine Learning Approach," *Sci Rep*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-52697-2.

[10] J. Kamruzzaman and R. K. Begg, "Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait," *IEEE Trans Biomed Eng*, vol. 53, no. 12, pp. 2479–2490, 2006, doi: 10.1109/TBME.2006.883697.

[11] Y. Li, M. A. Mache, and T. A. Todd, "Automated identification of postural control for children with autism spectrum disorder using a machine learning approach," *J Biomech*, vol. 113, p. 110073, 2020, doi: https://doi.org/10.1016/j.jbiomech.2020.110073.

[12] L. Montesinos, R. Castaldo, and L. Pecchia, "On the use of approximate entropy and sample entropy with centre of pressure time-series," *J Neuroeng Rehabil*, vol. 15, no. 1, p. 116, 2018, doi: 10.1186/s12984-018-0465-9.