School Attendance Prediction among Students with Autism: A Deep Learning-based Framework

Mohammed Jarbou  
Binghamton University

Daehan Won (dhwon@binghamton.edu)  
Binghamton University

Jennifer Gillis  
Binghamton University

Raymond Romanczyk  
Binghamton University

Research Article

Keywords: neurodevelopmental, Autism Spectrum Disorder, Long Short-Term Memory, Multilayer Perceptron

DOI: https://doi.org/10.21203/rs.3.rs-691760/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
School Attendance Prediction among Students with Autism: A Deep Learning-based Framework

Mohammed Jarbou\textsuperscript{a}, Daehan Won\textsuperscript{a*}, Jennifer Gillis\textsuperscript{b}, Raymond Romanczyk\textsuperscript{b}

\textsuperscript{a} Department of System Sciences and Industrial Engineering, Binghamton University-State University of New York, Binghamton, NY 13902

\textsuperscript{b} Department of Psychology, Binghamton University-State University of New York, Binghamton, NY 13902

* Corresponding Author: Daehan Won; Tel +1 (607) 777-5024; Fax +1 (607) 777-4094
Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects the areas of social communication and behavior. The term “spectrum” refers to the wide range of symptoms observed across individuals with ASD. Many children with ASD experience difficulty with daily functioning at school and home. ASD prevalence increases in the United States, with the most recent prevalence of 1.9%. Given the wide range of social and learning, difficulties experienced by children with ASD, it is paramount that they are able to attend school to receive the appropriate range of interventions.

School absenteeism (SA) is a significant concern given its association with many negative consequences such as school drop-out. Early prediction of SA would help school district to implement effective interventions to ameliorate this issue.

Due to its heterogeneity, students with ASD show within-group differences concerning their SA. This research introduces a deep learning-based framework for predicting short-and long-term SA of students with ASD. The Long Short-Term Memory (LSTM) algorithm is used to predict short-term SA. Similarly, Multilayer Perceptron (MLP) and Random Forest (RF) algorithms are used to predict long-term SA. The proposed framework achieves a high accuracy of 89% and 90% to predict short-term and long-term SA, respectively.
Introduction

Autism Spectrum Disorder (ASD) is generally defined as a complex neurodevelopmental disorder accompanied by a range of social communication impairments and repetitive patterns of interests and behaviors\(^1\)\(^2\). Although its risk factors are yet not fully defined, different genetic, biological, and environmental factors are addressed in the literature as contributing to the development of ASD\(^2\)\(^3\). Approximately 1.5%–2.0% of the children in the US are diagnosed with ASD.

For some individuals diagnosed with ASD, they will experience a range of significant impairments that adversely impact their quality of life\(^4\). Thus, it is paramount that school-age children with ASD are able to consistently access interventions and specialized educational instruction in school settings\(^5\). For example, attending public school enables students with to interact with their typically developing peers, which may increase their social development, a key area of need for children with ASD. In the same regard, there are many special education schools that provide intensive, specialized interventions and learning opportunities for children with ASD\(^6\)\(^7\).

Recent reports suggest that children with ASD miss school more than other clinical populations, leading to fewer opportunities for these students to benefit from school-based interventions. For example, 5%–28% of typically developing students are reported to have missed school days while this percentage jumps to 40%–53% among students with ASD\(^6\). Similarly, the percentage of chronic absenteeism (CA), defined as missing more than 10% of the annual school days, among typically developing students and students with ASD is 13% and 23%, respectively\(^6\). These statistics clearly illustrate that SA disproportionally affects children with ASD and can serve to negatively impact the effectiveness of ASD school-based interventions\(^6\)\(^7\)\(^8\).

At present, there is limited research on risk factors for SA in children with ASD. Some areas that require further attention include: (1) examination of gender, anxiety, depression, and challenging home settings\(^9\); (2) examination of possible associations between SA and specific child characteristics. For instance, students with particular health conditions (e.g., asthma) miss more school days when they experience severe symptoms\(^9\); and (3) group-level comparison between students with ASD and typically
developing students with respect to SA risk factors. Besides the importance of these studies, the problem
of SA prediction can also be addressed because of the possibility to provide timely interventions to improve
the SA. To the best of our knowledge, this is the first study that introduces Machine Learning- (ML) and
Deep Learning- (DL) based frameworks for predicting SA of students with ASD. The following paragraphs
discuss the challenges of the SA prediction problem and the advantages of using ML and DL techniques
over the conventional statistical analysis tools.

SA prediction aims to establish the probability for each student of the number of missed school
days in the future. If sufficiently accurate, such information could allow school districts and at-home
caregivers to understand SA patterns and perhaps divert attention and resources to specific children with
enough time to intervene adequately. This in turn would enable students to attend school regularly and
benefit from school-based interventions and services.

Students with ASD are heterogeneous as they show different symptoms and severity of symptoms
that might also change with time which can present as a challenge to prediction of SA\(^3\). For instance, in our
case Fig.1 shows that students with equal or similar attendance rates show very different individual
attendance patterns. The figure also shows that ASD students of different risk factors (e.g., food allergy)
are similar with respect to their attendance rate. These findings suggest a hypothesis that the group-level
analysis of SA risk factors does not necessarily explain the SA behavior of the ASD students at the
individual level, which is important in designing customized school-based interventions for SA.

Predicting SA at the individual level requires mining the SA history of each student. The authors
decided to recast the SA prediction problem into a time-series based sequence prediction to accomplish
this. Therefore, we used the students' attendance and maladaptive behaviors, modeled as a time series, as
input data to predict their SA and CA behaviors in the future. Methodology-wise, we used ML and DL
techniques because they outperform the conventional techniques (e.g., ARIMA). More details in this regard
will be given in the following section.

The main hypothesis of this research is twofold: (1) each student with ASD shows different SA
patterns (as shown in Fig. 1); and therefore, (2) SA is better predicted at the individual level. These
hypotheses led the authors to utilize a framework that employs a combination of DL, ML, and time series modeling techniques to model and predict the individual SA and. These techniques are adopted because they outperform the conventional statistical techniques (e.g., ARIMA) in learning the complex patterns and long-range dependencies of the temporal data (e.g., SA behavior).

The results are expected to provide early predictions of when each student will be absent and which students might be at risk of CA in the future. The present research project uses a real dataset for a population of 120 students with ASD. The data was collected at a private special education school in a mid-Atlantic state. More details about the data will be provided in the following sections.

The first objective of this research is to propose a short-term prediction framework to predict the SA at the individual level. This framework efficiently predicts whether a particular student will attend school over a prediction horizon of 10 school days. The second complementary objective is to propose a long-term prediction framework as to whether a particular student will be at risk of CA over the upcoming three months. CA needs to be predicted early enough because it is challenging and demands intensive and systematic interventions to be in place\textsuperscript{10}.

The main contribution of this research is to introduce an ML/DL-based framework for short-term SA and long-term CA prediction. This objective has been set by the authors to address the following gaps in the literature:

1. Provide accurate predictions of SA and CA behavior using DL, ML, and time series modeling.
2. This is the first research that predicts SA and CA of students with ASD at the individual level with consideration of the heterogeneity of ASD.

**Literature Review**

*ASD risk factors and phenotypes*

ASD is a neurodevelopmental disorder with a wide range of symptoms and levels of severity mostly impacting areas of social communication and restricted interests and behavior\textsuperscript{2}. Currently its etiology and risk factors are still not well-defined despite the research efforts dedicated to this purpose\textsuperscript{10,11}. The main common risk factors are genetic, demographic, environmental, and family-related\textsuperscript{6}. The
association between ASD and other factors, such as parental, perinatal, prenatal, and neonatal, are also investigated and discussed in the literature\textsuperscript{11,12,13}.

The advances in genetics research lead to a growing interest in discovering what causes ASD from a genetic perspective. This question is still challenging, and its answer is arguable. While many studies show that autism traits are heritable, the responsible gene factor(s) is(are) not commonly defined\textsuperscript{4}. Some research shows that different gene expressions cause different traits or symptoms of ASD. On the contrary, other studies concluded that different traits could be linked to the same underlying genetic expression\textsuperscript{11}.

As a parallel research stream to ASD diagnosis, a significant amount of research investigates multiple comorbid symptom patterns, psychiatric disorders, and medical conditions. For example, children with ASD are reported to have different sleeping patterns, such as bedtime resistance, night waking, sleep anxiety, and many others\textsuperscript{14}. Food aversion (e.g., eating refusal), social anxiety, and aggressive behavior are all considered as phenotypes of ASD\textsuperscript{15,16}.

The learning capabilities of students with ASD, in relation to the academic goals of school, is another research focus. For example, multiple studies show that reading comprehension skills are negatively affected by ASD\textsuperscript{17}. Many other studies show that students with ASD have different school-related behaviors as compared to typically developing students\textsuperscript{18}. Students with ASD tend to miss more school days than typically developing students\textsuperscript{10}. To the best of our knowledge, the risk factors of SA behavior are not yet well-defined. The following sections will discuss this area in more detail. Fig. 2 summarizes the literature on ASD.

\textit{SA and CA risk factors}

SA is problematic for its long-term impact on students a range of student outcomes. Recent reports show that 13%-16% of US students are chronically absent. This percentage represented eight million students in 2015\textsuperscript{4}. The percentage of CA among autistic students is twice that of typically developing students. Given these alarming figures, the association between SA and ASD has been inadequately studied\textsuperscript{2}. Challenges highlighted in this regard are manifold: (1) students with ASD show different SA patterns, frequency, duration, and expression; (2) SA behavior appears to be idiosyncratic. Therefore, the
population-based investigation does not necessarily represent an individual’s SA, and (3) SA is time variant
due to the vulnerability of the students with autism to the surrounding environment.9

In general, the risk factors of SA are either individual, familial, or environmental.6,9 While anxiety
and poor social relationships are examples of individual risk factors,6,9 familial risk factors include parental
support and home atmosphere.

The school environment has a significant effect in this regard. For example, classroom behavior
management is significantly associated with SA. Also, the transition between classes, grades,
developmental stages, as well as learning demands are other challenging risk factors for ASD.7,18

Awareness has been raised to the schools’ role in managing and controlling SA through early and
well-designed interventions. SA prediction is critical for schools to effectively improve their students'
attendance. To accomplish this, schools need to know, in advance, when and for how long each student
might be absent. This will give the schools enough time to plan for proper and effective interventions.

Statistical models for SA and CA prediction

Many research studies investigate the SA and school refusal factors using different statistical
techniques. For example, the chi-square test has been used to assess the association between SA and other
factors such as gender, presence of siblings, and single parents.19 ANOVA and chi-square tests have been
employed to thoroughly investigate the social, emotional, and behavioral characteristics of autistic students
who have SA.9 Simple hypothesis-testing techniques are also used to prove the significance of anxiety for
school as a predictor factor of SA.20 Simple statistical analysis has been deployed to address the relationship
between CA and some developmental disabilities of students with ASD.10 Regression analysis techniques
are also used to test the significance of different factors as SA predictors.21,22,23

Machine learning in education.

ML is a set of powerful techniques widely used to analyze and obtain useful insights from
multivariate and complex data. Interest is growing to harness ML capabilities in the area of education
research. For example, the association mining algorithm is used to discover the students' behavioral factors
that affect their e-learning courses.24 Clustering algorithms are also used to assign students into
homogeneous groups of similar learning styles. Also, the student's drop-out possibility is predicted using logistic regression and decision tree algorithms. Many other applications have been developed to introduce ML algorithms into other school-related applications such as students' achievement prediction.

SA is another research focus of education research literature. For example, intensive research work has been directed at defining the risk factors of SA. To the best of our knowledge, ML and DL algorithms have not been used to predict the SA behavior of students with ASD or any other child population. This research aims to fill this literature gap by introducing a ML/DL framework for SA and CA prediction among students with ASD.

**Results and discussion**

*Short term SA prediction (univariate and multivariate)*

This research proposes a DL-based framework for predicting the short-term SA of students with ASD. First, a univariate LSTM forecasting model is proposed to provide early predictions of the students' SA behavior dependent upon their attendance history. Expanding upon this, a multivariate LSTM model is then employed by enriching the data source with the students' maladaptive (socially inappropriate) behavior history (e.g., aggressive behavior). As shown in Fig. 3, adding maladaptive behavior improves prediction accuracy and precision while it slightly decreases prediction recall. These results encourage us to dig deeper into investigating the causality relationship between maladaptive and SA of students with ASD. Such an investigation will help design more customized SA interventions that consider these two essential phenotypes of ASD. For example, more customized in-class learning activities could be implemented to improve the students' adaptive behavior, which result in better school attendance.

From a practical perspective, it is of value to know for how far ahead the proposed model can satisfactorily predict SA. Fig. 3 shows that the model is recommended to be used for a time horizon of 10 school days ahead with an accuracy of (90%), and acceptable precision of (80%). As expected, the overall quality of the prediction decreases as the forecasting lead value increase. It implies that the SA of students with ASD might change over time. So, consistent updating mechanisms (e.g., mobile apps) should be in place to record, track, and update attendance.
Long-term SA prediction (scenario I + scenario II)

Fig. 4 shows examples of MLP and RF’s capability to satisfactorily predict long-term CA even when only short attendance history is available. The two algorithms are used to learn the CA history of 120 students with ASD to predict whether each student will be chronically absent over the upcoming three months.

Using different train/test splits, MLP consistently provides better results in terms of all the performance metrics. The prediction performance of the two algorithms is summarized in Table 1 considering the 3 months and 12 months scenarios.

Our results also highlight the causality relationship between maladaptive behavior and SA of students with ASD. More research effort is needed to address this issue quantitatively through different techniques, such as social networks and association mining algorithms. In our opinion, the more the dynamics of ASD phenotypes are investigated, the more the SA interventions will be customized and efficient. Moreover, these research results are expected to encourage school districts to collect, track, and intelligently analyze school-related data, which result in the improvement of overall education quality.

Conclusion

In this research, the ML- and DL-based framework is proposed for the SA and CA prediction of students with ASD. First, the input data is modeled as a time series to represent the students’ attendance and maladaptive behavior history. LSTM algorithm is used for short-term SA prediction. MLP and RF algorithms are then used for long-term CA prediction. Both models show a promising capability to predict SA and CA behavior for 10 school days and three months ahead, respectively. The results are expected to help in designing customized interventions to manage SA effectively. Future research includes (1) improving the adopted algorithms’ performance through hyperparameters optimization, and (2) enriching the proposed framework's data source using other characteristics and behaviors to predict SA and CA.
This research introduces a ML- and DL-based framework to handle short-term SA and long-term CA problems for students with ASD. LSTM algorithm is used for the first problem. In this regard, univariate and multivariate forecasting models are built. Students' attendance history is used as input for the univariate model, while the multivariate model considers the history of students' maladaptive behavior as another data input. The univariate model predicts students' SA based on their attendance behavior. In contrast, the multivariate model depends on students' attendance and maladaptive behaviors to predict their SA. For the CA prediction problem, the individual characteristics are added to the attendance history to enrich the data source. Two different scenarios are also hypothesized for students with long and short attendance history, as detailed later.

Data description

This research's targeted population includes 120 students with ASD with significant impairment and who have an average age of six years, and of which 79% are male, while 21% are female. Also, the population has an attendance rate of 90%, while 23% are reported chronically absent. The data was collected from the Institute for Child Development (ICD) in Binghamton, NY. The research presented in this study was approved by the Binghamton University’s Institutional Review Board (IRB). Also, all methods utilized in this study for data collection were carried out in accordance with relevant regulations. The informed consent was waived off in this study and it was approved by the Binghamton University Human Subjects Research Review Committee (HSRRC), which is the IRB responsible for the review of research.

Table 2 provides more details about the demographic characteristics of the targeted population. We first investigated whether the students' individual characteristics (e.g., communication skills, motor skills, emotional control, and others) are significant predictors of their SA behavior. This investigation is motivated by the lack of research that addresses the relationship between individual characteristics and SA. Statistical hypothesis testing is applied, and the results, depicted in Fig. 1, show no association between these characteristics and the SA of the targeted population. The results also support our hypothesis that (1) SA is heterogeneous and should be predicted at the individual level, (2) SA is better
predicted depending on its history. The association between maladaptive behavior and SA is discussed in
the literature. Therefore, maladaptive behavior will also be used, in this research, to predict SA. This is also
supposed to help design customized interventions to improve SA behavior that considers different ASD
phenotypes.

Short-term SA prediction

Data preprocessing for short-term SA prediction

To predict short-term SA, the history of students' attendance and maladaptive behavior is first
modeled as a time series. Data transformation includes binary encoding of attendance time series (1: attendance, 0: absence) and normalizing the time series of maladaptive behavior. Then, the data is
restructured to take the shape of supervised ML-like data using a rolling forecasting technique such that a
sequence of \((i - 1)\) past events are used to predict the future event \((A_i)\) at time \((T_i)\) where \((I)\) is the value
of the lag parameter. Thus, the entire time series of each student is partitioned into given labels of \((N)\)
sequences each of length \((I)\) as features in addition to \((N)\) events \((A_i)\), to be predicted. For validation
purposes, the data is split using three training-testing thresholds, as will be illustrated later. Other secondary
data cleaning steps are also accomplished.

LSTM algorithm

LSTM is a popular recurrent DL algorithm that is used for to mine the hidden patterns of sequential
data. Many LSTM variations have been introduced to enhance its capability (e.g., diamond LSTM and
bidirectional LSTM). The LSTM areas of application are manifold, which include time series analysis,
natural language processing, and others. In this research, LSTM will be used for the first time to predict
the SA behavior among students with autism.

In this research, the SA of each student is modeled as a time series. Unlike the typical forecasting
techniques (e.g., ARIMA and SARIMA), LSTM is known for its capability to learn the long-term
dependencies of sequential and temporal data. For this reason, LSTM will be used in this research for
short-term SA modeling and prediction. It is worth mentioning that typical forecasting techniques (e.g.,
ARIMA) perform well on the seasonal and linear time series. However, they are less powerful in their ability to capture the long-term dependencies of sequential data than DL (e.g., LSTM)\(^{26}\).

Opposite to the typical DL algorithms, the neurons at each hidden layer are replaced by memory cells that work together with three types of gates: input, forget, and output gates. This characteristic enables the LSTM algorithm to avoid the gradient vanishing problem. In this sense, LSTM is proven in the literature for its superiority of learning and predicting/to learn and predict long sequential data\(^{27}\).

To fulfill the scope of this research, univariate and multivariate LSTM forecasting models are built. The time series of students’ attendance history are used for to train the univariate model as a single input. However, the dataset of the multivariate model is enriched by adding the time series of students’ maladaptive behavior in addition to school attendance. Fig. 5 illustrates how the proposed model works.

LSTM algorithm with a rolling forecasting technique is employed in this research to predict future SA. As with any DL algorithm, LSTM performance is a function of multiple architectural parameters (a.k.a hyperparameters). Tuning these parameters is critical to optimize LSTM accuracy. Multiple optimization algorithms have been introduced in the literature for this purpose\(^{28}\). Parameters optimality is beyond our scope in this research because the main focus will be on the introduction of a new framework for SA prediction for students with ASD.

SA prediction is addressed as a forecasting problem in this research. Therefore, LSTM performance is also a function of two main forecasting parameters: lag and lead. While lag refers to the amount of history needed to predict the next future event, the lead parameter's value represents the number of future events that could be predicted at once using the given lag value. Table 3 summarizes all the LSTM hyperparameters values, which include the forecasting lag/lead values, adopted in this research.

Three training-testing split settings are employed for better model validation. Each of these settings is embedded with a rolling forecasting technique that trains the LSTM model using different data portions. In the same regard, accuracy, precision, and recall are adopted to evaluate the model's performance for each of the validation settings. Accuracy reflects the model's overall prediction quality, while the two other
metrics check the model's capability to predict the attendance events correctly. Fig. 4 shows the model performance over different validation settings.

**Long-term CA prediction**

**Data preprocessing**

In long-term CA prediction, the main objective is to predict whether a particular student will be chronically absent over the upcoming three months. This problem is handled as a pattern recognition problem using MLP and RF algorithms. A combination of a 12-month attendance history and 15 individual characteristics (e.g., medical restrictions, allergy restrictions, and atypicality score) have been used as features. Binary encoding is used to model the monthly attendance history as a binary sequence in addition to the individual binary characteristics (e.g., medication and allergy restrictions). Moreover, the individual numerical features (e.g., age) are normalized. The future CA status is labeled as a binary sequential pattern. For example, (100) means the student will be chronically absent in the second and third months.

Data balancing is necessary to avoid learning bias. Therefore, the input data is also balanced using the standard oversampling technique. Different training-testing splitting thresholds are applied to validate the model. This step will be discussed in detail later in this section. To further validate its robustness, we applied our model to a hypnotized scenario where some students have a short history of school enrollment (three months). The results show our framework's ability to predict CA even for recently enrolled students with a relatively short CA history.

**MLP and RF algorithms**

In this research, long-term CA behavior is also predicted. The problem is formulated as a pattern recognition problem. Each pattern represents the status of students' CA for three months ahead. MLP and RF are two commonly used algorithms for pattern classification problems in the literature. MLP is one of the most common ANN with a broad spectrum of applications. It has a powerful capability to approximate non-linear functions by learning the hidden complex patterns in large, complex, and noisy data. MLP architecture generally consists of one input and one output layer in addition to at least one hidden layer. Inspired by the human brain structure, each layer includes multiple neurons that
work as knowledge processing units. Neurons in each layer are connected to the other layers' neurons through artificial links that hold some value of weights. The backpropagation algorithm is commonly used to train MLP and optimize its weights such that the error function converges to its global or local minima.

RF is a state-of-the-art machine learning algorithm with outstanding prediction and feature selection performance. RF works simply as an ensemble learning algorithm that aggregates \((N)\) independent and deep tree predictors into one powerful final model. In this sense, RF has an outstanding capability to learn complicated and irregular patterns\(^{28}\). In more detail, the FR algorithm trains \((N)\) independent trees \((f_b)\) using different portions of the training data \(\{(X_b, Y_b) \in (X, Y)\}\). Then, the final model \((F)\) is made by averaging the performance of all the individual models \((f_b)\).

MLP and RF have been used to handle the long-term CA prediction as a pattern recognition problem. We applied both algorithms considering two scenarios of twelve- and three-month long histories of school attendance. These scenarios are hypothesized to investigate the robustness of the proposed framework to predict CA for students with different attendance history lengths. The hyperparameters optimization step is not considered as it is beyond the scope of this research. Table 3 summarized the model parameters that are used for each algorithm.

To validate the adopted models' performance, we tested the results using different data splits to train the models using different data portions. In addition, accuracy, recall, and precision metrics are also used to investigate the quality of our predictions.
1. Am Psychiatric Assoc. *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.) (Arlington, 2013).

2. Zablotsky, B. *et al*. A. Prevalence and trends of developmental disabilities among children in the United States: 2009–2017. *Pediatrics, 144(4)*, (2019).

3. Maenner, M.J., Shaw, K.A., and Baio, J. Prevalence of autism spectrum disorder among children aged 337 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 338 2016. *MMWR Surveillance Summaries, 69(4)*, 1 (2020).

4. Kearney, C. A. School absenteeism and school refusal behavior in youth: A contemporary review. *Clinical psychology review, 28(3)*, 451-471 (2008).

5. Cp J., Meyers, SM. Identification and evaluation of children with autism spectrum disorder. *Pediatrics, 120*, 1183-215 (2007).

6. Tonge, B. J., & Silverman, W. K. Reflections on the field of school attendance problems: for the times they are a changing?. *Cognitive and Behavioral Practice, 26(1)*, 119-126 (2019).

7. Munkhaugen, E. K., Torske, T., Gjevik, E., Nærland, T., Pripp, A. H., & Diseth, T. H. Individual characteristics of students with autism spectrum disorders and school refusal behavior. *Autism, 23(2)*, 413-423 (2019).

8. Gottfried, MA. Chronic absenteeism and its effects on students’ academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk (JESPAR), 19(2)*, 53-75 (2014).

9. Black, L. I., & Zablotsky, B. Chronic school absenteeism among children with selected developmental disabilities: national health interview survey, 2014-2016. *National Health Statistics Reports, 118*, (2016)

10. Chaste, P., & Leboyer, M. Autism risk factors: genes, environment, and gene-environment interactions. *Dialogues in clinical neuroscience, 14(3)*, 281 (2012).

11. Dietert, R. R., Dietert, J. M., & DeWitt, J. C. Environmental risk factors for autism. *Emerging health threats journal, 4(1)*, 7111 (2011).
12. Larsson, H. J., Eaton, W. W., Madsen, K. M., Vestergaard, M., Olesen, A. V., Agerbo, E., Mortensen, P. B. Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. *American journal of epidemiology*, **161**(10), 916-925 (2005).

13. Liu, W., Li, M., & Yi, L. Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework. *Autism Research*, **9**(8), 888-898 (2016).

14. Cotton, S. M., & Richdale, A. L. Sleep patterns and behavior in typically developing children and children with autism, Down syndrome, Prader-Willi syndrome and intellectual disability. *Research in Autism Spectrum Disorders*, **4**(3), 490-500 (2010).

15. Peverill, S. *et al.* Developmental trajectories of feeding problems in children with autism spectrum disorder. *Journal of pediatric psychology*, **44**(8), 988-998 (2019).

16. Spain, D., Sin, J., Linder, K. B., McMahon, J., & Happé, F. Social anxiety in autism spectrum disorder: a systematic review. *Research in Autism Spectrum Disorders*, **52**, 51-68 (2018).

17. Chou, Y. C., Wehmeyer, M. L., Palmer, S. B., & Lee, J. Comparisons of self-determination among students with autism, intellectual disability, and learning disabilities: a multivariate analysis. *Focus on Autism and Other Developmental Disabilities*, **32**(2), 124-132 (2017).

18. Roberts, J., & Webster, A. Including students with autism in schools: a whole school approach to improve outcomes for students with autism. *International Journal of Inclusive Education*, 1-18 (2020).

19. Ochi, M., Kawabe, K., Ochi, S., Miyama, T., Horiuchi, F., & Ueno, S. I. School refusal and bullying in children with autism spectrum disorder. *Child and adolescent psychiatry and mental health*, **14**, 1-7 (2020).

20. Adams, D., Simpson, K., & Keen, D. School-related anxiety symptomatology in a community sample of primary-school-aged children on the autism spectrum. *Journal of school psychology*, **70**, 64-73 (2018).

21. Soares, F. R. R., Farias, B. R. F. D., & Monteiro, A. R. M. Consumption of alcohol and drugs and school absenteeism among high school students of public schools. *Revista brasileira de enfermagem*, **72**(6), 1692-1698 (2019).
22. Freeman, N. C., Schneider, D., & McGarvey, P. Household exposure factors, asthma, and school absenteeism in a predominantly hispanic community. *Journal of Exposure Science & Environmental Epidemiology*, **13**(3), 169-176 (2003).

23. Stempel, H., Cox-Martin, M., Bronsert, M., Dickinson, L. M., & Allison, M. A. Chronic school absenteeism and the role of adverse childhood experiences. *Academic pediatrics*, **17**(8), 837-843 (2017).

24. Li, B., Sharma, A., Meng, J., Purushwalkam, S., & Gowen, E. Applying machine learning to identify autistic adults using imitation: An exploratory study. *PloS one*, **12**(8), (2017).

25. Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. Data mining techniques applied in educational environments: literature review. *Digital Education Review*, **33**, 235-266 (2018).

26. Houdt, G.V., Mosquera, C.J., & Nápoles, G. A review on the long short-term memory model. *Artificial Intelligence Review*, 1-27 (2020).

27. Rana, A., Rawat, A. S., Bijalwan, A., & Bahuguna, H. Application of multi-layer (perceptron) artificial neural network in the diagnosis system: a systematic review. In *2018 International conference on research in intelligent and computing in engineering (RICE)*, 1-6 (2018).

28. Chen, W. *et al.* A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, **151**, 147-160 (2017).
Author contribution statement

M.J. performed the technical side of this research, cleaned, and preprocessed the datasets, built, trained, and validated the ML and DL models, analyzed the results, and wrote this paper's main manuscript. D.W. oversaw all steps in this study, designed the ML- and DL-based framework, validated the prediction models' design and the results. Besides providing real datasets for this research, R.R. and J.G. provided the team with all the theoretical background, insights, and directions from the educational and psychological perspectives.

Acknowledgments

This research is supported by Watson Institute for Systems Excellence (WISE) and implemented in cooperation with Institute for Child Development (ICD) at Binghamton University.

Additional Information

Conflict of Interest and Notes

The authors declare that there are no conflicts of interest.
Figure 1 Heterogeneity & time-variant nature of SA behavior. (a) SA pattern for a student with an average attendance rate (85%). (b) Another pattern for a different student with a similar average rate (88%) but less frequency. (c) Attendance rate for students with/without allergy. (d) Attendance rate by the score of living skills. (e) Attendance rate by the score of living emotional control skills. The figures show no association between SA behavior and the individual characteristics of autistic students. The figures also show that SA behavior is better to be predicted/investigated at the individual level.
Figure 2 Schematic diagram of ASD literature review. This research focuses on SA as one of the ASD phenotypes.
Figure 3 SA prediction performance of the proposed models (a) Multivariate model outperforms the univariate model in terms of prediction accuracy (one-way ANOVA, $n = 125, p = 0.02$). (b) Prediction precision of both models is statically indifferent (one-way ANOVA, $n = 125, p = 0.68$). (c) Prediction recall of both models is statically indifferent (one-way ANOVA, $n = 125, p = 0.40$). Adding maladaptive behavior through multivariate model significantly improves the quality of SA prediction in terms of accuracy ($p = 0.02 \leq 0.05$). (d) Prediction accuracy changes by the lead value (days). (e) Prediction precision decrease by lead value (days). (f) Prediction recall change by lead value (days). The model is reliable to be used for predicting ten school days ahead with 90% accuracy and 80% precision.
Figure 4 Long-term CA model performance (a) For a student with 100% prediction accuracy (b) Another student with 67% prediction accuracy. (c-1) Prediction accuracy for a lead value of one day. (c-2) 3 days lead value. (c-3) 5 days lead value. The best prediction performance achieved by (lead value = 1) and (0.70/0.30) train/test validation threshold.
Figure 5 Schematic diagram of the proposed framework (a) univariate and (b) multivariate models
### Table 1 CA prediction performance for 12- & 3-months lag scenarios

| Classifier | Training accuracy | Testing accuracy | Precision | Recall |
|------------|-------------------|------------------|-----------|--------|
| **MLP (0.70/0.30)** | 1.0 | 0.91 | 0.94 | 0.94 |
| **RF (0.80/0.20)** | 0.96 | 0.85 | 0.91 | 0.90 |

| Classifier | Training accuracy | Testing accuracy | Precision | Recall |
|------------|-------------------|------------------|-----------|--------|
| **MLP (0.70/0.30)** | 0.91 | 0.81 | 0.87 | 0.84 |
| **RF (0.70/0.30)** | 0.93 | 0.79 | 0.87 | 0.90 |
Table 2 Demographic characteristics of the targeted population

|                          | Male | Female |
|--------------------------|------|--------|
| Gender                   |      |        |
|                          | 0.79 | 0.21   |
| Age                      |      |        |
| ≤ 5                      | 0.65 |        |
| > 5                      | 0.35 |        |
| Medical restrictions     |      |        |
| With                     | 0.45 | 0.55   |
| Without                  |      |        |
| Diet restrictions        |      |        |
| With                     | 0.58 |        |
| Without                  | 0.42 |        |
| Allergy restrictions     |      |        |
| With                     | 0.55 |        |
| Without                  | 0.45 |        |
Table 3 LSTM & MLP hyper-parameters and forecasting values.

| Parameters          | LSTM algorithm | MLP algorithm |
|---------------------|----------------|---------------|
| Number of hidden layers | 3              | 3             |
| Activation function      | "linear"       | "tanh"        |
| Batch size           | 32             | 20            |
| Number of epochs     | 100            | 100           |
| Optimizer            | "Adam"         | "lbfgs"       |
| Lag (days)           | 5              | --            |
| Lead (days)          | 1              | --            |