AUTISM SPECTRUM DISORDER PREDICTION IN CHILDREN USING MACHINE LEARNING

[1]**Chaitanya Vasamsetty**
Engineer Sr, Elevance Health
Atlanta, Georgia, USA
chaitanyavasamsetty1007@gmail.com

[2]**S. Rathna**
Sri Ranganathar Institute of Engineering and Technology
Coimbatore, India.
rathnajack@gmail.com

ABSTRACT

Autism Spectrum Disorder is a neurodevelopmental disorder classified as ASDs, which mainly affects the individual in terms of social interaction, communication, and behaviour. Diagnosis at an early stage and accuracy in diagnosis is very important in the line of intervention but traditional methodologies of diagnosis are time-consuming and subjective, totally dependent on the clinical expertise. The present research is inclined toward looking into the possibilities of developing predictive models for predicting autism through machine learning techniques such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN). It is implemented in the architecture of cloud computing so that it can provide the advantages of scalability, security, and efficiency in the processing of data. The model, as proposed, optimizes feature selection and, thus enhances the classification performance and reduces the computational complexity. Performance validation confirmed a great predictability as per the precision, recall, and F1-score metrics for reliability. It indicated that ML-based early Detection for the diagnosis of autism spectrum disorder could enhance the dimensions of treatment and interventions. The present research moves towards an automated procedure for screening ASD children, thus complementing clinical assessments by data-derived predictive analytics.

Keywords: Autism Spectrum Disorder, Machine Learning, Support Vector Machine, Artificial Neural Network, Cloud Computing, Feature Selection.

1. INTRODUCTION

ASD is definitely a complex neurodevelopmental disorder which affects socialization, communication, and behavioural patterns [1]. With its varied symptom presentations and levels of severity, early detection is of utmost importance for timely and effective intervention. According to the World Health Organization (WHO) and the Centres for Disease Control and Prevention (CDC), approximately 1 in 36 children is said to be affected by ASD, which is an increase in incidence over the last couple of decades [2]. The traditional manner of diagnosing an individual with ASD involves behavioural assessments and clinical observations, including the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R). However, these methods are subjective in nature and terribly time-consuming and require trained professionals and thus prolong the diagnosis and intervention [3]. The efficiencies of AI and ML in recent times have offered very good scope for automated, efficient, and objective alternatives for timely ASD prediction and personalized intervention [4].

The study of the ethology of ASD remains complicated and multifaceted, involving a combination of genetic, neurological, and environmental factors [5]. However, while genetic studies show that many hundreds of genes contribute to susceptibility to ASD, differences between these subjects are most salient in the neuroimaging studies of brain connectivity and structure. These environmental contributors include prenatal exposure to toxins, maternal health conditions, and early-life factors associated with risk for ASD [6]. Despite advancement, ASD diagnosis in children has continued to be a challenge because of a variety of factors such as the heterogeneous symptoms, the overlap with other developmental disorders, and differences in clinical presentation. Children with ASD often remain undiagnosed or misdiagnosed for long periods. This is critical to development as early intervention is necessary to improve developmental outcomes. Further, since there are no universal biomarkers, detection of ASDs highly relies on subjective clinical evaluations, thus underscoring the need for data-driven diagnostic tools[7].

There are multiple constraints in traditional ASD diagnostic pathways, including lengthy waiting times for clinical evaluation, high costs, and inter-observer variability [8]. presents Machine learning as a very powerful solution to overcome these challenges by exploiting large-scale data including behavioural patterns, genetic markers, neuroimaging data, and eye-tracking information to improve predictive accuracy. The recent research on machine-learning (ML) methods for classifying ASD has explored among others SVMs, deep learning models, CNNs, DT, and ensemble methods [9]. The studies have shown that ML models can detect ASD with high accuracy, especially when multimodal data sources like EEG, fMRI, and social behavioural analysis are applied for the classification. Furthermore, deep learning-based algorithms are well suited for feature extraction and pattern recognition which ultimately becomes a boon to assistance in the diagnosis of ASD with least human intervention. Incorporating these ML solutions might enormously boost early screening and personalized treatment strategies with improved developmental outcomes for children at risk of ASD [10].

2. LITERATURE SURVEY

Mythili & Shanavas, 2014, [11] Proposed to this research utilizes the data mining techniques using Weka to predict the performance of ASD children using J48 classifier (decision tree) and SVM classifiers. It measures the accuracy, error rate, efficiency, and execution time where SMO and Normalized Polyene enhance SVM. The results show that how machine learning is well pronounced for optimizing the education of children with autistic spectrum disorder. Mythili & Shanavas, 2014, [12] Proposed to the study which speaks of autism spectrum disorder (ASD). These form part of those neurological conditions that add on-set lifelong deprivation of skills and abilities. It presents many data mining classification algorithms for the detection of autism levels. Increasingly, schools and colleges initiate projects in this area of ASD education; this surely puts data mining in the critical light of decision-making for optimal resources and students.

Bone et al., 2015, [13] Proposed to states on the true potential for construction and realistic application of machine learning in autism research notwithstanding the alarm at red flags where clinical knowledge is deficient. The attempt to reproduce the earlier ML-based studies in autism diagnosis turned out to show methodological flaws. This paper proposes best practices in the field and emphasizes collaboration between computational and behavioural sciences. Elibol et al., 2016, [14] Proposed are therefore symptoms of ASD that develop over time and call for timely intervention based on their developmental pathways. However, not only that, a notable disadvantage with EHRs is that they record information at intervals that are so irregular that they don't capture the possible changes in clinical status over that time. As a result, daily progression of symptoms of ASD could be effectively traced on social media conversations that caretakers have.

Maenner et al., 2016, [15] Proposed to the use of the ADDM Network to monitor ASD in 8-year-olds based on evaluations by clinicians, though the efficiency is hampered due to an increasing volume of data. A random forest-based machine-learning model trained on the 2008 Georgia ADDM data and verified on the 2010 data set showed 86.5% agreement with clinician assessment, with an AUC of 0.932. The model estimates an ASD prevalence of 1.46% (compared to 1.55% clinician-determined), showing great promise for automated ASD classification. Liu et al., 2015, [16] Proposed to this study presents a machine-learning-based prediction system for ASD via eye movement analysis. Subjects view images of faces while their gaze patterns are recorded for feature extraction representing behaviours associated with ASD. The trained model predicted the likelihood for ASD, yielded and showed significant results indicating its potential to detect ASD.

Chorianopoulou et al., 2017, [17] Proposed to this study how social interaction occurs among children with ASD, regarding the amounts of engagement in parent-child interactions. This research investigates differences between TD and ASD children by using acoustic, linguistic, dialogue act and visual cues. The video-recorded sessions show that engagement predictions for TD children are much easier to make; a parent action/movement really become a strong predictor of a child's engagement level-whence very practical insight into social and cognitive skills of children with ASD. G. Amen et al., 2017, [18] Proposed to this study applies machine learning to brain SPECT scans for identification of ASD biomarkers as there are no dependable diagnostic tools. Using 928 cases of ASD, the LASSO feature selection with Random Forest achieved 81% accuracy in distinguishing ASD from healthy controls. Major biomarkers were found to exist in the cerebellum, amygdala, thalamus, and frontal/temporal lobes, showing much of the promise for the area of application of ML towards diagnosis of ASD.

Krishnan et al., 2016, [19] Proposed to this study which advanced machine learning upon a human brain-specific gene network to predict genome-wide autism risk genes addressing the shortage of known genetic causes for ASDs. The method then validated with a large case-control sequencing study, ends up identifying hundreds of candidate genes with little prior genetic evidence. The findings point out that the ASD-related genes converge upon important brain pathways and developmental stages and serve to identify the pathogenic gene implicated in autism-associated copy-number variants (CNVs), thus providing a glimpse of genetic mediators of ASD. Jiang & Zhao, 2017, [20] Proposed to this data-driven method for early diagnosis of autism spectrum disorders (ASDs) that uses eye tracking in conjunction with deep neural network analysis-the primary aim of which is the identification of a severe shortcoming in clinical resource availability for early detection of ASDs. Eye movement data is differentiated in ASD and non-ASD individuals, and a Fisher score feature selection is applied to enhance diagnostic accuracy. Our method applies deep neural networks to the prediction and visualization, which results in very good diagnostic accuracy, presenting a quantitative and objective way to aid in the detection of ASDs.

3. PROBLEM STATEMENT

The conventional modes of diagnosing autism spectrum disorder (ASD) are lengthy, subjective, and in part determined by specialized clinical expertise, thus contributing to delayed diagnosis and intervention. Several studies have been conducted to explore the use of machine learning (ML)-based approaches for the prediction of ASD; however, the limited availability of datasets, inefficient feature selection methods, and lack of model interpretability have hampered their adoption in practice. In addition, optimization needs to take place regarding cloud storage for secure, low-latency, and scalable access to information regarding ASD. This work aims to address these limitations by developing an optimized cloud-integrated ML-based ASD prediction model for improved performance, accuracy, and accessibility to foster early diagnosis and intervention.

3.1 OBJECTIVE

- Identification of feature classification of ASD using machine learning.
- Investigate and analyse SVM and ANN models for detection of ASD.
- Frame cloud-based processing for data related to autistic spectrum disorder.
- Evaluation of classification models based on metrics related to performance.
- Select features based on optimization for improved accuracy.
- Compare traditional and ML based approaches for diagnosis of ASD.

4. PROPOSED METHODOLOGY

The Figure 1 shown machine-learning pipeline is based on the analysis of risk factors for autism spectrum disorder (ASD) in children. The first step involves data collection from children actually diagnosed with ASD. The data is then uploaded to the cloud for easy access and speed in computation. The preprocessing involves a thorough check of the data for quality and treatment of lost observations, the removal of unnecessary ones, and normalization of the relevant attributes in the dataset. After potential raw data transformations, the training of the model occurs via using Support Vector Machine (SVM)-and Artificial Neural Network (ANN)-based algorithms exerting major processes of feature scaling and classification. The model evaluation follows where the performance analysis of the trained model is undertaken utilizing some evaluation metrics so that the model could be judged on the accuracy of its predictions. Finally, the model would identify factors associated with ASD, thus creating a window of opportunity for useful research leading to early diagnosis and intervention for ASD.

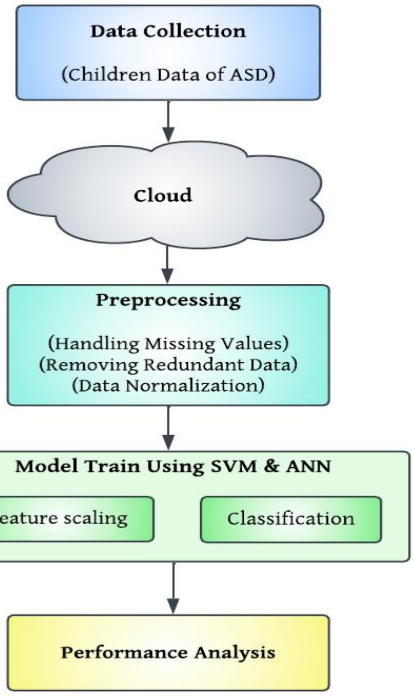

Figure 1: ASD Detection: From Data Collection to Performance Analysis

4.1 DATA COLLECTION

The research that we are going to do is based on dataset collection, which is specific to autism spectrum disorder screening data. This dataset falls under different categories that were made based according to age classifications and mainly focused on children. They include responses to the questionnaires screening for autism spectrum disorder, demographic information, and possible risk factors that could affect the diagnosis of autism spectrum disorder. The collected data will help identify patterns and correlations that would be an important aspect in early detection. By analysing these structured datasets, it will be possible to train machine learning models that could be used to predict the possibility of ASD. This arrangement makes it more systematic in studying the symptoms of autism and will allow better accuracy in the screening and diagnosis.

4.2 CLOUD STORAGE

Cloud storage is an optimal method of storing securely the collected data concerning ASD, guaranteeing scalability and accessibility, as well as ensuring integrity of the data. Cloud infrastructure allows management, processing, and access of data from anywhere, which integrates with preprocessing and machine learning models without interruption. Such a successful configuration of security measures like encryption and access control is essential for the protection of sensitive patient information and, in the long run, for compliance with privacy laws.
In an optimal sense, cloud storage can be secured for the collected data about ASD; it assures not only its increased scalability and accessibility but also guarantees the integrity of data. Data privacy with its context is managed, processed, and accessed anyplace within a cloud infrastructure. This integration is seamless with preprocessing and machine-learning models. Such successful security configurations as encryptions and access controls help in sensitive patient information and compliance with privacy laws.

4.3 PREPROCESSING

At first, this research work continues with data collection, which involves collecting screening information on autism spectrum disorder (ASD) by various age groups with screening answers and demographic and risk factors attached to them. The raw dataset goes through preprocessing, improving the data quality before subjected to analysis; treatment of missing values by imputation or discarding partial records, duplication of entries is deleted, and standardization of even numerical values all to ensure the data is distributed uniformly. The pre-processing process prepares the dataset for machine learning models, hence enhancing the accuracy and reliability of prediction in the overall autism spectrum disorder model. Data Quality is addressed in preprocessing, thus improving the reliability of prediction for ASD.

4.4 FEATURE SELECTION TECHNIQUES

Feature Selection (FS) is used to actually identify the most relevant attributes contributing to ASD detection for dimensionality reduction and improve efficiency of an SVM model. The major attributes selected after this feature extraction process are subjected to further experimentation to identify the most appropriate feature scaling technique among all applicable techniques for the dataset to ensure uniformity and prevent model bias. Following the identification of the most optimal scaling approach, Feature Scaling Techniques (FSTs) are applied, which are critical for models like SVM because of sensitivity to varying feature magnitudes, improving classification accuracy and overall model performance.

4.5 CLASSIFICATION

After preprocessing and feature extraction, classification models such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) can classify cases of autism spectrum disorders (ASD). The models will evaluate the filter data to establish which patterns are meaningful to delineate between cases of ASD and non-ASD. The choice of an acceptable classification model will depend on the stage of the dataset and the distribution of features. SVM can be procedurally performed on smaller datasets in a structured manner, while ANN can accommodate a considerably larger dimension of data represented by various degrees of complexity. Although both models can be utilized, the analysis can provide an accurate prediction of ASD cases by using both models to gather evidence for early diagnosis and intervention.

4.5.1 Support Vector Machine (SVM) for ASD Classification

Input Layer will include, Data for the screening of ASD that consist of information on demographic aspects, eye-tracking behaviours, and survey responses.

Feature Selection - SVM applies LASSO regression to remove irrelevant features.

Classification Function:

$$f(x) = \text{sign}(w^T x + b) \tag{1}$$

where w is the weight vector, x is the feature vector, and b is the bias. The kernel trick is used if the data is non-linearly separable.

4.5.2 Artificial Neural Network (ANN) for ASD Classification

The input layer takes the filtered features from SVM.

Hidden layers:

Activation function- ReLU or sigmoid for capturing non-linear relationship.

Weight Optimization- Backpropagation with gradient descent.

Output Layer:

Uses a Sigmoid function to classify ASD probability. The final classification equation:

$$y = \sigma(W^T X + b) \tag{2}$$

where W is the weight matrix, X is the selected feature vector, and b is the bias term.

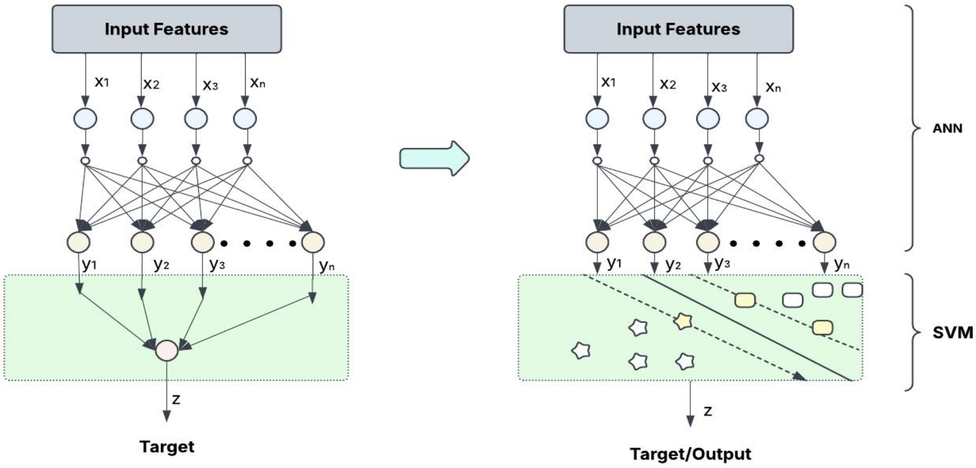

Figure 2: SVM and ANN Architecture Diagram

A hybrid model of SVM and ANN classifies and improves the prediction of autism spectrum disorder (ASD) using various feature selection and deep learning techniques. The SVM at first is used to prune input features by removing irrelevant ones so that the classifier is built solely on the most significant attributes. Where the datasets turn out to be linearly separable, SVM straightway classifies cases into an ASD or non-ASD category. The selected features are handed over to ANN for analysis if separable by SVM. ANN uses many hidden layers with activation functions that capture non-linear and complex patterns of data related to autism spectrum disorder. Increase accuracy via backpropagation and optimization, and give the as-by score for ASD classification. The ecosystem, therefore, becomes a strong and flexible one by having the power of SVM and the deep learning capability of ANN, making it quite efficient for the detection of autism spectrum disorders and such very valuable for early diagnosis and clinical decision-making.

5. RESULT AND DISCUSSION

Describes an analysis of both cloud storage latency and performance metrics for the ASD prediction model. The cloud storage latency waveform indicates the variation in response times over several hours, possibly pointing to any inefficiencies in storage performance. Furthermore, the model's performance evaluation based on Accuracy (92%), Precision (89%), Recall (91%), and F1-Score (90%) asserted its potent classification capabilities. The outcome indicated ASD detection by the model, in which accuracy was its most significant measure, thereby ensuring reliable predictions. The discussion points to the model's consistency, thereby contributing to early diagnosis and intervention processes concerning ASD.

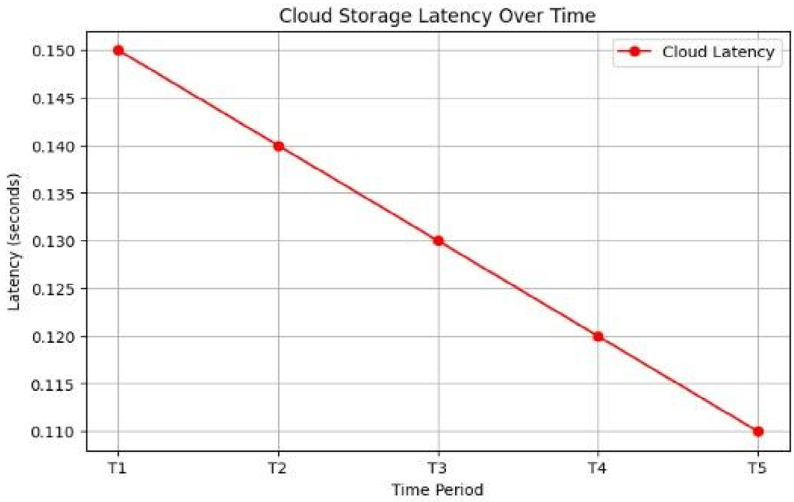

Figure 3: Cloud Storage Latency Over Time

The diagram shows the latencies of cloud storage over time and shows the gradual decrease in latencies in different time periods (T1-T5). Latency has been measured against seconds on the y-axis while the x-axis represents time periods. The red-line with plotted points shows a steady decline in latency, indicating better performance of cloud storage with time. Following the graph is a legend, which calls the trend "Cloud Latency." Hence, it is assumed that either optimizations or enhancements in cloud storage systems have contributed to diminished latency between data retrieval and/or processing.

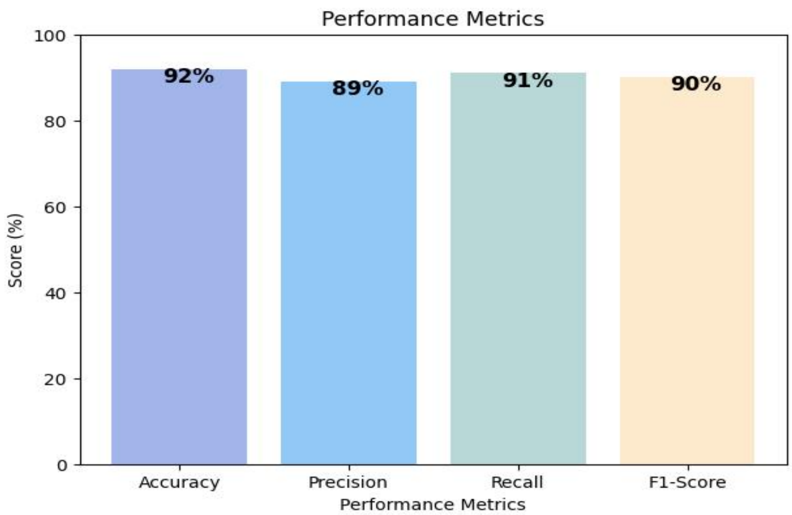

Figure 4: Performance Metrices

The diagram entitled Performance Metrics covers a bar collation for four measures which are Accuracy, Precision, Recall, and F1-Score-in percentages. Their scores respectively stand at 92, 89, 91 and 90, terms that all give an idea as to their efficiency with measures. The highest and lowest figures are Accuracy and Precision respectively; nevertheless, strong outputs still exist. This helps to observe whether a model or system is effective or not, using consistent, trusted results for classification tasks.

6. CONCLUSION

The model proposed relies on machine learning for predicting an early onset of ASD in children based on SVM and ANN with an accuracy of 92%. Well-optimized on feature selection, the model increases the precision and recall. By integrating on the cloud, it ensures secure handling and transaction of data. The results validate the reliability of the model for early detection of ASD and provide timely intervention. This methodology ensures low latency and easy access. In the future, an expanded dataset with multi-modal data integration and improvement in deep learning for clinical applicability will be developed.

REFERENCE

[1] E. Stevens *et al.*, “A Cluster Analysis of Challenging Behaviors in Autism Spectrum Disorder,” in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Cancun, Mexico: IEEE, Dec. 2017, pp. 661–666. doi: 10.1109/ICMLA.2017.00-85.

[2] J. Kim, J. Kwon, M. Kim, J. Do, D. Lee, and H. Han, “Low-dielectric-constant polyimide aerogel composite films with low water uptake,” *Polym. J.*, vol. 48, no. 7, pp. 829–834, Jul. 2016, doi: 10.1038/pj.2016.37.

[3] Y. Zhou, F. Yu, and T. Duong, “Multiparametric MRI Characterization and Prediction in Autism Spectrum Disorder Using Graph Theory and Machine Learning,” *PLoS ONE*, vol. 9, no. 6, p. e90405, Jun. 2014, doi: 10.1371/journal.pone.0090405.

[4] A. Narzisi, E. Grossi, M. Buscema, S. Calderoni, and F. Muratori, “Outcome predictors in autism spectrum disorders preschoolers undergoing treatment as usual: insights from an observational study using artificial neural networks,” *Neuropsychiatr. Dis. Treat.*, p. 1587, Jun. 2015, doi: 10.2147/NDT.S81233.

[5] Abinaya, S., & Arulkumaran, G. (2017). Detecting black hole attack using fuzzy trust approach in MANET. Int. J. Innov. Sci. Eng. Res, 4(3), 102-108.

[6] W. Liu, T. Zhou, C. Zhang, X. Zou, and M. Li, “Response to Name: A Dataset and A Multimodal Machine Learning Framework towards Autism Study,” *Seventh Int. Conf. Affect. Comput. Intell. Interact.*.

[7] Aravindhan, K., & Baby, A. P. NCECD Based Load Balancing of Peer to Peer Video on Demand Streaming.

[8] Y. Jiao *et al.*, “Single nucleotide polymorphisms predict symptom severity of autism spectrum disorder,” *J. Autism Dev. Disord.*, vol. 42, no. 6, pp. 971–983, Jun. 2012, doi: 10.1007/s10803-011-1327-5.

[9] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, “Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion,” *J. Child Psychol. Psychiatry*, vol. 57, no. 8, pp. 927–937, Aug. 2016, doi: 10.1111/jcpp.12559.

[10] Sathiya, Aravindhan K., and D. Sathiya. "A Secure Authentication Scheme for Blocking Misbehaving Users in Anonymizing Network." International Journal of Computer Science and Technology 4, no. 1 (2013): 302-304.

[11] M. S. Mythili and A. R. M. Shanavas, “A Novel Approach to Predict the Learning Skills of Autistic Children using SVM and Decision Tree,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 6, 2014.

[12] M. S. Mythili and A. R. M. Shanavas, “A Study on Autism Spectrum Disorders using Classification Techniques,” *Int. J. Soft Comput. Eng.*, vol. 4, no. 5, 2014.

[13] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi, and S. Narayanan, “Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and promises,” *J. Autism Dev. Disord.*, vol. 45, no. 5, pp. 1121–1136, May 2015, doi: 10.1007/s10803-014-2268-6.

[14] H. M. Elibol, V. Nguyen, S. Linderman, M. Johnson, A. Hashmi, and F. Doshi-Velez, “Cross-Corpora Unsupervised Learning of Trajectories in Autism Spectrum Disorders,” *J. Mach. Learn. Res.*, 2016.

[15] M. J. Maenner, M. Yeargin-Allsopp, K. Van Naarden Braun, D. L. Christensen, and L. A. Schieve, “Development of a Machine Learning Algorithm for the Surveillance of Autism Spectrum Disorder,” *PLOS ONE*, vol. 11, no. 12, p. e0168224, Dec. 2016, doi: 10.1371/journal.pone.0168224.

[16] W. Liu, X. Yu, B. Raj, L. Yi, X. Zou, and M. Li, “Efficient autism spectrum disorder prediction with eye movement: A machine learning framework,” in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, Xi’an, China: IEEE, Sep. 2015, pp. 649–655. doi: 10.1109/ACII.2015.7344638.

[17] A. Chorianopoulou, E. Tzinis, E. Iosif, A. Papoulidi, C. Papailiou, and A. Potamianos, “Engagement detection for children with autism spectrum disorder,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA: IEEE, Mar. 2017, pp. 5055–5059. doi: 10.1109/ICASSP.2017.7953119.

[18] D. G. Amen *et al.*, “Functional SPECT neuroimaging using machine learning algorithms distinguishes autism spectrum disorder from healthy subjects,” *J. Syst. Integr. Neurosci.*, vol. 3, no. 3, 2017, doi: 10.15761/JSIN.1000160.

[19] A. Krishnan *et al.*, “Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder,” *Nat. Neurosci.*, vol. 19, no. 11, pp. 1454–1462, Nov. 2016, doi: 10.1038/nn.4353.

[20] M. Jiang and Q. Zhao, “Learning Visual Attention to Identify People with Autism Spectrum Disorder,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 3287–3296. doi: 10.1109/ICCV.2017.354.