

Screening of Autism Spectrum Disorder using Machine Learning Approach in Accordance with DSM-5

1st Shreea Bose

Department of Post Graduate and Research Computer Science
St Xavier's College
Kolkata, India
boseshreea.official@gmail.com

2nd Prakriti Seth

Department of Biotechnology and Medical Engineering
National Institute of Technology
Rourkela, India
prakritiseth.bio@gmail.com

Abstract—Autism Spectrum Disorder (ASD) is a specific category of neurodevelopmental disorder that can be associated with several behavioral conditions and has no known cure to date. ASD can be detected from a very early stage in childhood and upon successful detection can be ameliorated. There have been several clinical diagnosis procedures and they can be error-prone and time-consuming. Thus, machine learning-based prediction models for early-stage ASD as well as in adolescents and adults are being developed over the years. In our study, several parameters of ASD detection were implemented with open-source ASD datasets and analysed using several machine learning models like Logistic regression, XGboost, SVC and Naive Bayes. Among these XGboost showed the best performance. The outcomes of such analytical approaches demonstrate that, when suitably optimized, machine-learning techniques can offer robust predictions of Autism Spectrum Disorder (ASD) status. These findings imply that it may be feasible to employ these models for the early ASD detection, thereby enhancing the prospects of timely and effective intervention. XGBoost has given best results throughout all datasets, including cross validation. An accuracy of 100% has been achieved, making the model best for prediction.

Index Terms—Autism Spectrum, Logistic regression, XGboost, SVC, Naive Bayes, DSM-5, Q-CHAT

I. INTRODUCTION

Autism Spectrum-Disorder (ASD) is a type of neurodevelopmental condition that includes social interaction, behavioral, and severe communication difficulties in individuals, which exhibits irritability, repetitive behavior, and concentration issues [1]. Since the publication of the 5th edition of the DSM (Diagnostic and Statistical Manual of Mental-Disorders, DSM-5), ASD has been recognised as a diagnostic entity that previously reflected several discrete disorders, such as Asperger's Syndrome, Autistic Disorder, and Pervasive Developmental Disorders [2], [3]. Although ASD is incurable, early discovery allows for more effective mitigation treatment. However, standard behavioral investigations make detecting and diagnosing ASD extremely challenging. It is usually discovered around two years age, however it might be discovered at later stages based on the severity of the symptoms [4]. While there are a variety of clinical methods available to diagnose ASD as early

as feasible, they include time consuming diagnostic methods that are rarely employed unless there is a high probability of ASD development. Clinical diagnostic procedures include ADOS-R (Autism Diagnostic Observation Schedule- Revised), ADI (Autism Diagnostic Interview), and many more [5], [6]. Clinical diagnostic approaches have demonstrated competitive performance in screening ASD patients. However, as the verbal parts cannot be answered appropriately for the patient, the kid who is too young and exhibits delayed speech impairments often scores just 25% of the total ADI-R questions. Furthermore, completing an interview with a carer by a professional takes 90-150 minutes, which is time-consuming and frequently results in data gaps. The identification of ASD with the help of ADOS-R, on the other hand, is based on scoring measurements based on the responses supplied. One of the key drawbacks of this technique is the propensity to overclassify adolescents with other clinical problems [7]. As a result, healthcare practitioners are in desperate need of a time efficient and simple ASD screening approach that can effectively predict whether any patient with particular measurable features has ASD and advise persons on whether they should seek a formal clinical diagnosis. Currently, there are few datasets available that are connected with clinical diagnosis, which is largely genetic, such as NDAR, Boston Autism Consortium (AC) and AGRE [8]–[10]. Advanced computational and engineering approaches have been used in recent years to address the demands of cross-disciplinary applications in psychology and psychiatry. Machine learning approaches have shown promise in bioinformatics, affective computing, behavioural informatics, and medical diagnostics [11]. Researchers are particularly interested in machine learning, which is based on advanced mathematical learning, statistical estimation, and information theory, as a broadly applicable computational-framework for uncovering meaningful patterns in vast volumes of data automatically. A learnt data representation, for example, can give insights into the processes that created the data, support people in clinical decision making by visualising data, and predicting a target-variable from a collection of input attributes (i.e., classification) [12]–[14]. Machine learning can be effectively

utilized in research related to autism to develop an accurate and sturdy diagnostic algorithm by analyzing human-coded behaviors from established diagnostic tools such as the ADI and the Autism Diagnostic Observation Schedule. As these instruments are carefully crafted by experts and backed by robust statistical analysis, it is plausible that objective machine learning techniques can enhance the performance and efficiency of the instruments by eliminating redundancy. Improving the speed of diagnostic procedures could enhance the ability to gather the extensive cohorts necessary to effectively uncover the intricate neurogenic causes of ASD. While attempting to obtain rapid categorical assessments of ASD, researchers have sometimes compromised on sensitivity/specificity or unintentionally skewed population sampling towards individuals with more severe impacts. [15], [16]. Machine learning appears to be a potential solution for speeding up these diagnostic efforts by detecting crucial nosological components, minimising duplication while retaining accuracy [17]. In this article, we apply various techniques for classification in order to obtain results with increased accuracy in the detection of ASD. The main findings of this work can be summarised in the following:

- We have analyzed different parameters of Toddler, Child, Adolescent and Adult datasets of ASD, and identify associations between the demographic information and ASD cases. The research aims to reveal fresh patterns and connections that help increase screening accuracy by evaluating a varied set of traits and their links with ASD.
- The research highlights the possibility for automating and improving the accuracy of ASD screening by using machine learning approaches to the screening process. Through our prediction models we have reached to the **accuracy of 100%** using the XGBoost Classifier.
- Machine learning-based screening approaches have the potential to be scaled and deployed in healthcare systems, bringing them to a larger population and meeting the growing demand for autism screening and diagnosis.

Overall, the work contributes to current efforts to improve accuracy, efficiency, and accessibility of ASD screening, and it is consistent with the larger goals of advancing autism research and healthcare.

II. RELATED WORKS

There are several research works that describe several machine learning-based approaches to determine ASD. Some of them are mentioned henceforth. According to the findings of Kosmicki and colleagues, a new method has been proposed to identify autism using the least number of features. The study used a machine-learning approach for the evaluation of clinical assessment of ASD, where a subset of autism-related behaviors in children was assessed using ADOS which is a tool consisting of four modules. The researchers employed eight different machine learning algorithms to identify stepwise backward features from the score sheets of 4540 participants. The results showed that by using 9 out of the 28 behaviors from Module 2 and 12 out of the 28 behaviors from Module 3, an ASD risk can be identified with an overall accuracy of

98.27% and 97.66%, respectively. [18]. In their study, Thabtah et al. suggested a screening method for ASD that integrates Machine-Learning Adaptation and the DSM-5. Screening instruments are often used to achieve some goals in screening of ASD. This research evaluated the benefits and limitations of the ASD Machine Learning classification. The author sought to highlight the challenges with current ASD screening methods and their reliability by comparing the DSM-IV manual to the DSM-5 text. [19]. Abbas et al. integrated the ADOS and ADI-R machine learning approaches into one evaluation technique and used that feature for encoding techniques [20]. Research conducted by Mythili et al. focused on utilizing Classification Techniques to better understand ASD. Their primary objective was to identify the various degrees of autism and the associated challenges. To achieve this, they employed SVM and Fuzzy approaches with WEKA tools within a Neural Network to analyze students' behavior and social interactions. [4]. Heinsfeld et al. employed a deep learning-based strategy to categorise a total of 1,035 individuals comprising of 505 ASD and 530 controls studies and obtained 70% accuracy [21]. Duda et al. used multiple classifiers to analyse ASD data and discovered that 5 out of 65 variables were adequate to differentiate ASD from ADHD [22][23]. Baron-Cohen et colleagues. created the Autistic-spectrum Quotient (AQ) as a screening technique [23][24]. In 2012, Allison and colleagues introduced DI (a discriminant index) technique to reduce the number of elements in the Q-CHAT and AQ tools from 50 to 10. The DI-based tool comprised five distinct sections, namely detail attention, attention switching, communication, social and imagination skills. This questionnaire-based technique was designed to optimize the ASD screening process and improve its accuracy, making it a promising avenue for future research in this field. [1].

III. PROPOSED MODEL

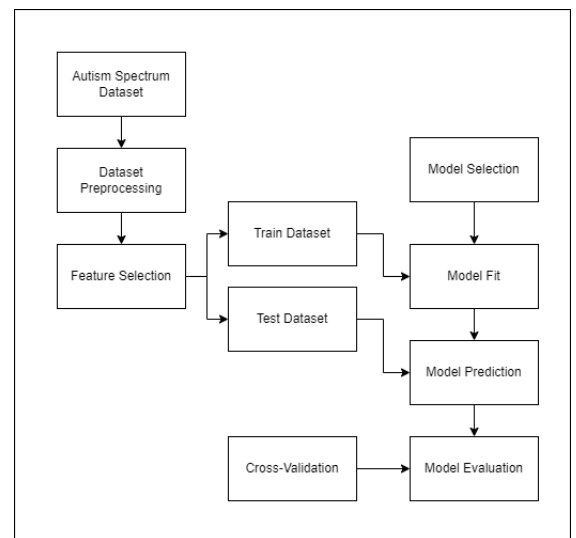


Fig. 1. Flowchart of ASD prediction using Machine Learning

A. Data Collection

A wide range of factors linked to autism spectrum disorder (ASD) make up the dataset used to predict ASD. It includes the Autism Spectrum Quotient, which ranges from A1 to A10 and is likely to correspond to specific evaluation or diagnostic criteria, Class/ASD, which indicates the presence or absence of ASD, and Q chat scores, which may provide further insights. Other columns provide demographic and background information, such as Sex, Ethnicity, Jaundice, and Family Members with ASD. Furthermore, the Class/ASD features column may contain ASD-related features. This dataset contains a diverse mix of features that can be used to create predictive models to identify prospective cases of ASD.

B. Data Preprocessing

Using machine learning models to predict autism is a challenging endeavor that calls for applying the right algorithms and examining a variety of data sources. We look for any missing data in these columns and take appropriate action. We may decide to eliminate rows with missing data or impute missing values, depending on the data. Preprocessing ensures that the machine learning model can handle categorical variables like "Sex" and "Ethnicity" by converting them into numerical representation using one-hot encoding or label encoding.

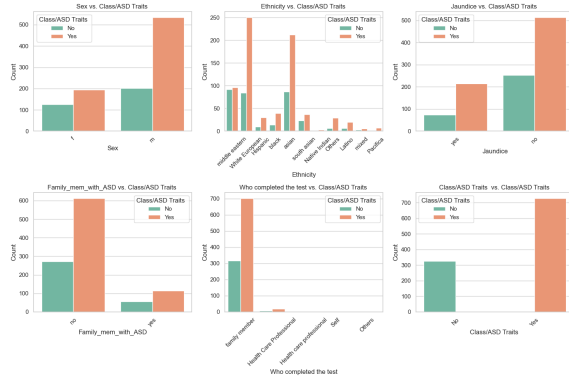


Fig. 2. Comparison Chart and Exploratory Data Analysis

Gender vs ASD: We examine the gender distribution of ASD patients. This comparison can reveal whether or not there is a gender difference in ASD prevalence. Here we find that the male dominates the female. This can be due to several reasons. **Ethnicity vs. ASD:** We try to investigate how the prevalence of ASD differs among ethnic groups. This comparison can aid in identifying any potential differences in ASD diagnoses between ethnic groups. **ASD vs. Jaundice:** We try to look into the association between a history of jaundice and the existence of ASD. It can play an important role in determining potential early risk factors for ASD development.

C. Feature Selection

A survey that participants complete themselves called the Autism Spectrum Quotient (AQ) is used to assess the prevalence of autistic traits in adults. There are fifty questions in

all, and the answers on a Likert scale go between "agree" to "disagree." The Autism Questionnaire (AQ) is a quantitative tool used in research to determine whether individuals without autism spectrum disorder (ASD) exhibit autistic symptoms. The A1 to A10 AQ questions are used for screening and cover a broader range of features and behaviors associated with autism. We are essentially investigating whether the presence of particular autistic traits, as indicated by the answers to these questions, is associated with an increased likelihood of having an autism diagnosis (represented by the values 0 or 1 in A1 to A10) when we compare AQ scores (A1 to A10) to the likelihood of having autism. It may be possible to ascertain whether particular AQ traits are indicative of ASD through this comparison.

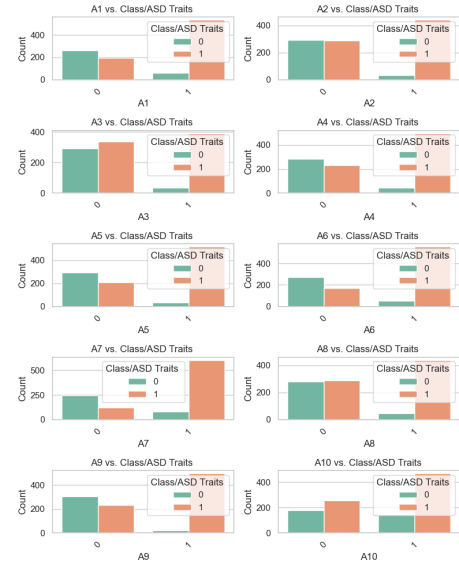


Fig. 3. Comparison Chart of Class vs AQ scores

D. Model Selection

1) Logistic Regression: When estimating the probability that a data item belongs to one of two classes—binary classification tasks—this statistical technique and well-liked machine learning algorithm are employed. There are typically two classes when it comes to forecasting autism spectrum disorder (ASD): ASD present (1) and ASD absence (0). In logistic regression, the relationship between a set of independent variables, or features, and the probability of a binary outcome is represented by the logistic function, also called the sigmoid function.

$$P(Y = 1) = 1 / (1 + e^{(p_0 + p_1 * X_1 + p_2 * X_2 + \dots + p_n * X_n)}) \quad (1)$$

Where $P(Y=1)$ is the probability of the outcome being class 1 (ASD present), e is the base of the natural logarithm, b_0 is the intercept and $(p_0 + p_1 * X_1 + p_2 * X_2 + \dots + p_n * X_n)$ are the coefficients for the independent variables X_1, X_2, \dots, X_n which are features in our dataset. The logistic function converts

the linear combination of characteristics and coefficients to a number between 0 and 1, which represents the likelihood of class 1.

2) *XGBoost Classifier*: Classifying Autism Spectrum Disorder (ASD) can be done with the use of the XGBoost (Extreme Gradient Boosting) classifier, a sophisticated machine learning algorithm that performs well in binary classification tasks. By combining the predictions of several decision trees mathematically, XGBoost improves prediction accuracy. It modifies the weights of samples that are misclassified while minimizing a loss function. The average of the outcomes from each individual tree yields the final forecast. Hyperparameters can be used to adjust XGBoost in order to maximize performance. The method works by training it on a dataset including pertinent variables, like demographic and behavioral data, and producing a likelihood score for the existence of ASD that may be thresholded for binary classification.

3) *Support Vector Machine*: Using Support Vector Machines (SVM), one can categorize people with autism spectrum disorder (ASD) by determining which hyperplane best divides data points that represent those with and without ASD. In mathematical terms, this means minimizing misclassifications and optimizing the margin between two classes. In this high-dimensional space, SVM seeks to identify a hyperplane that optimizes the margin between the two classes (false and authentic news). The optimal hyperplane to divide the two classes is found through SVM.

$$w * x + b = 0 \quad (2)$$

Here w is the weight vector and b is the bias term. A data point is categorized as fake or true news depending on which side of the hyperplane it lies on.

$$y_i(w * x_i + b) \geq 1 \quad (3)$$

A kernel function, like the radial basis function (RBF) kernel, translates the data into a higher-dimensional space where a hyperplane can be identified when a linear hyperplane is unable to separate the data adequately.

4) *Naive Bayes*: It is a probabilistic classification technique that is used to predict Autism Spectrum Disorder (ASD). Naive Bayes employs Bayes' theorem to determine the likelihood of an individual having ASD based on their observed characteristics. It makes the calculations easier by assuming feature independence. Naive Bayes evaluates the chance of ASD presence based on feature patterns by examining behavioral, demographic, or clinical data. While the simplicity of Naive Bayes may not capture complicated associations, it does provide a rapid and interpretable solution for preliminary screening, complementing more advanced ASD prediction tools.

IV. RESULTS AND DISCUSSION

A clinical dataset of (1054, 19) is taken, preprocessed and split into Train and Test Dataset and then used by different Models to find the prediction and accuracy.

```
LogisticRegression() :
Training Accuracy : 100.0
Validation Accuracy : 100.0
Cross Validation Score : 97.62 %
classification_report:
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	57
1	1.00	1.00	1.00	154
accuracy			1.00	211
macro avg	1.00	1.00	1.00	211
weighted avg	1.00	1.00	1.00	211

Fig. 4. Classification report and accuracy of Logistic Regression

The logistic regression model performs admirably, with a 100% accuracy rate. This indicates that for all 211 cases, it properly predicts the absence (0) or presence (1) of Autism Spectrum Disorder (ASD). Both classes have flawless precision and recall values (0 and 1), indicating no false positives or false negatives. A harmonic mean of accuracy and recall, or F1-score, of 1.00 indicates the model's balance between precision and recall. In conclusion, the model has excellent predictive accuracy, making it a highly dependable tool for ASD categorization based on the available data.

```
XGBClassifier(base_score=None, booster=None, callbacks=None,
colsample_bylevel=None, colsample_bynode=None,
colsample_bytree=None, device=None, early_stopping_rounds=None,
enable_categorical=False, eval_metric=None, feature_types=None,
gamma=None, grow_policy=None, importance_type=None,
interaction_constraints=None, learning_rate=None, max_bin=None,
max_cat_threshold=None, max_cat_to_onehot=None,
max_delta_step=None, max_depth=None, max_leaves=None,
min_child_weight=None, missing=None, monotone_constraints=None,
multi_strategy=None, n_estimators=None, n_jobs=None,
num_parallel_tree=None, random_state=None, ...) :
```

```
Training Accuracy : 100.0
Validation Accuracy : 100.0
Cross Validation Score : 100.00 %
classification_report:
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	57
1	1.00	1.00	1.00	154
accuracy			1.00	211
macro avg	1.00	1.00	1.00	211
weighted avg	1.00	1.00	1.00	211

Fig. 5. Classification report and accuracy of XGBoost

The XGBoost classifier produces outstanding results with a 100% accuracy. It accurately forecasts the absence (0) or presence (1) of Autism Spectrum Disorder (ASD) in both the training and validation datasets. Precision, recall, and F1-score for both ASD groups are also perfect, indicating that no false positives or false negatives occurred. The strong performance of the model, as evidenced by the cross-validation score, reveals its excellent capabilities for precise and reliable ASD categorization.

The Support Vector Classifier (SVC) obtains a training accuracy of 100%, correctly predicting ASD for the training dataset. The validation accuracy of 98.10% is outstanding, suggesting strong generalization to previously unseen data. Its reliability is reinforced by a cross-validation score of 96.21%. High precision, recall, and F1-score values in the classification report for both ASD classes (0 and 1) suggest a compromise between avoiding false positives and false negatives. The model's weighted average F1-score of 0.98

```
SVC() :
```

Training Accuracy :	100.0			
Validation Accuracy :	98.10426540284361			
Cross Validation Score :	96.21 %			

```
classification_report:
```

	precision	recall	f1-score	support
0	0.96	0.96	0.96	57
1	0.99	0.99	0.99	154
accuracy			0.98	211
macro avg	0.98	0.98	0.98	211
weighted avg	0.98	0.98	0.98	211

Fig. 6. Classification report and accuracy of SVM

indicates its high overall performance, making it a reliable tool for ASD prediction.

```
GaussianNB() :
```

Training Accuracy :	97.12543554006969			
Validation Accuracy :	97.6303317535545			
Cross Validation Score :	97.16 %			

```
classification_report:
```

	precision	recall	f1-score	support
0	0.93	0.98	0.96	57
1	0.99	0.97	0.98	154
accuracy			0.98	211
macro avg	0.96	0.98	0.97	211
weighted avg	0.98	0.98	0.98	211

Fig. 7. Classification report and accuracy of Naive Bayes

In predicting Autism Spectrum Disorder (ASD), the Naive Bayes model produces satisfying results. It displays robust generalization to fresh data with a training accuracy of 97.13% and a validation accuracy of 97.63%. The cross-validation score of 97.16% adds to its dependability. According to the classification report, the model provides a fair mix of precision and recall for both ASD classes (0 and 1), with high F1 scores. Notably, its overall high performance is reflected in the weighted average F1-score of 0.98. The Naive Bayes model is a powerful and dependable technique for ASD classification, making it useful in healthcare and early screening.

CONCLUSION

This work employed a variety of machine learning techniques, including logistic regression, xGBoost, support vector classifier (SVC), and naive bayes, to predict autism spectrum disorder (ASD) based on important characteristics connected to the DSM-5 criteria. The results are quite promising, with accuracy levels above 97% across all models, XGboost being the most accurate one. This study contributes to the realm of healthcare by proving the feasibility of employing machine learning techniques for early ASD screening and identification. These findings could greatly help healthcare practitioners identify suspected cases of ASD, allowing for early intervention and support. This can lead to alleviation of ASD until any cure is developed. However, it is critical to acknowledge that additional study, validation, and clinical evaluation are required before adopting these models in real-world medical practice. As we move forward, ethical issues, interpretability, and constant model refinement will be critical.

REFERENCES

- [1] C. Allison, B. Auyeung, and S. Baron-Cohen, "Toward brief "red flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist for autism in toddlers in 1,000 cases and 3,000 controls [corrected]," *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 51, no. 2, pp. 202–212.e7, Feb. 2012.
- [2] M. J. Maenner, K. A. Shaw, J. Baio, EdS1, A. Washington, M. Patrick, M. DiRienzo, D. L. Christensen, L. D. Wiggins, S. Pettygrove, J. G. Andrews, M. Lopez, A. Hudson, T. Baroud, Y. Schwenk, T. White, C. R. Rosenberg, L.-C. Lee, R. A. Harrington, M. Huston, A. Hewitt, PhD-7, A. Esler, J. Hall-Lande, J. N. Poynter, L. Hallas-Muchow, J. N. Constantino, R. T. Fitzgerald, W. Zahorodny, J. Shenouda, J. L. Daniels, Z. Warren, A. Vehorn, A. Salinas, M. S. Durkin, and P. M. Dietz, "Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, united states, 2016," *MMWR Surveill. Summ.*, vol. 69, no. 4, pp. 1–12, Mar. 2020.
- [3] F. Thabtah, F. Kamalov, and K. Rajab, "A new computational intelligence approach to detect autistic features for autism screening," *Int. J. Med. Inform.*, vol. 117, pp. 112–124, Sep. 2018.
- [4] M. S. Mythili and A. M. Shanavas, "(2014) "A study on Autism spectrum disorders using classification techniques," " *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 4, pp. 88–91, 2014.
- [5] C. Lord, M. Rutter, and A. Le Couteur, "Autism diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," *J. Autism Dev. Disord.*, vol. 24, no. 5, pp. 659–685, Oct. 1994.
- [6] C. Lord, S. Risi, L. Lambrecht, E. Cook, B. Leventhal, P. DiLavore, A. Pickles, and M. Rutter, "The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of Autism and Developmental Disorders*, vol. 30, pp. 205–223, 06 2000.
- [7] C. NV., *Data mining and knowledge discovery handbook*. Boston: Springer; 2009, 2009.
- [8] G. D. S. J. and L. C., *Letters to the Editor 463*. Am J, 2001.
- [9] F. G and L. C., "The Simons Simplex Collection: a resource for identification of autism genetic risk factors," *Neuron*. 2010;68(2):192–5. doi:, vol. 10., 2010.
- [10] P. A, K. CS, V. R, and P. KV., "Soft computing models for the predictive grading of childhood autism—a comparative study," Tech Report 2014;, 2014.
- [11] P. Baldi, *Bioinformatics: The machine learning approach*. Cambridge, MA: The MIT Press, 2001.
- [12] M. P. Black, A. Katsamanis, B. R. Baucom, C. C. Lee, A. C. Lammert, A. Christensen *et al.*, "(2013)," *Toward automating a human behavioural coding system for married couples' interactions using speech acoustic features*. *Speech Communication*, vol. 55, no. 1, 2013.
- [13] D. Bone, C. C. Lee, T. Chaspari, M. P. Black, M. E. Williams, S. Lee, P. Levitt, and S. Narayanan, "Acoustic-prosodic, turn-taking, and language cues in child-psychologist interactions for varying social demand," in *INTERSPEECH*, 2013, p. 2400–2404.
- [14] Z. Ye, Y. Li, A. Fathi, Y. Han, A. Rozga, G. D. Abowd, and J. M. Rehg, "Detecting eye contact using wearable eye-tracking glasses," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, p. 699–704.
- [15] R. W. Picard, *Affective computing*. Cambridge, MA: MIT press, 2000.
- [16] L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, 2005.
- [17] M. B, L. C, Y. D, F. F, and D. P., *Machine learning classification of adults with autism spectrum disorder*. <http://shorturl.at/hEFJY>. Accessed 25, Jun. 2020. [Online]. Available: <http://shorturl.at/hEFJY>
- [18] J. A. Kosmicki, V. Sochat, M. Duda, and D. P. Wall, "(2015) "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning," " *Translational psychiatry*, 2015.
- [19] F. Thabtah, (2017). "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment, 2017.
- [20] H. Abbas, F. Garberson, E. Glover, and D. P. Wall, ""machine learning approach for early detection of autism by combining questionnaire and home video screening," *Informat. Assoc.*, vol. 25, no. 8, p. 1000–1007, 2018.

- [21] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *Neuroimage Clin*, vol. 17, 2018.
- [22] M. Duda, R. Ma, N. Haber, and D. P. Wall, "use of machine learning for behavioral distinction of autism and ADHD," " *Transl. Psychiatry*, vol. 6, no. 2, p. e732, 2016.
- [23] B.-C. S, W. S, S. R, M. J, and C. E., "The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians," *J Autism Dev Disord*. 2001;31(1):5. doi:, vol. 10., 2001.