Interpreting Likelihood of Autism Spectrum Disorder in Adults from Screening Results

Fahim Muntasir[1], Syed Md. Ahnaf Hasan[2], Asif Jawaad[3]
Dept. of Computer Science and Engineering
BRAC University, Dhaka, Bangladesh
Email: muntasirfahim.niloy@gmail.com

***Abstract*—Autism Spectrum Disorder(ASD) consists of a diverse group of mental conditions. Although there have been awareness practices for ASD in children and adolescents, the situation among adults is largely overlooked. In this study, a framework based on verbal screening and analysis of the response with explainable AI has been demonstrated for detecting the likelihood of generating ASD in adults. Despite numerous machine learning approaches in the medical field, the black box nature of the algorithms have been the obstacle in adapting them among mass clinical practitioners. This study focuses on the interpretability of the result with the help of both global and local explanations of the features effect on association with ASD. An accurate model combined with an explained result has the ability to become a useful tool in ASD screening in Adults.**

***Index Terms*—Autism Screening, Explainable AI, SHapley Additive exPlanations (SHAP), LIME, Ada boost, XGBoost**

I. Introduction

Autism spectrum disorder (ASD) is a term used to refer to a set of neurodevelopmental disorders that impede natural behavioural activities in human beings. Individuals with autism are born with the neural condition and retain it for the rest of their lives. People with autism have difficulties in expressing themselves, interacting with others, and accomplishing day-to-day tasks. However, studies show that early detection of autism can help to start early treatment that can assist them in better coping with their conditions and adapting accordingly [1], [2]. The diagnosis methods consist of questionnaires based on which scores are calculated which estimate whether a person has autism or not. Some widely used methods are ADI, ADOS-R, Q-CHAT, etc. [3]. Most of the prevalent screening tools compute the scores via handcrafted rules designed by clinical experts. The operating of the tools also requires prior expertise. Consequently, the diagnosis process becomes tedious and time-consuming, making the task of early detection very difficult. An alternative approach that can replace the handcrafted rules and also automate the process is machine learning [4]. Machine learning is a part type of AI technique that is able to learn patterns from a given dataset. The most common uses of machine learning are regression and classification. The case of medical diagnosis can is treated as a classification problem in machine learning. The dataset is divided into two parts: feature variables and a target variable. The algorithm learns the relationship between the features and the target variable and generates the appropriate prediction when a new instance is provided. Use of machine learning techniques are already showing promising results in various fields of medical diagnosis. Likewise, for the case of early autism screening, the inclusion of machine learning can speed up and improve the performance of the already established tools [5]. For autism screening, the task is a binary classification. The target variable 'ASD' can have either of two values for an instance: 'YES' or 'NO'. The questionnaires can provide relevant feature variables based from which the machine learning algorithm can develop a relationship with the target variable for prediction. The machine learning approach is significantly faster compared to the traditional rule-based approach, causing timely intervention for early treatment. Performance-wise machine learning algorithms are very efficient in generating accurate predictions, however, it is not readily understandable why a particular prediction was made. This is due to the complexity of the model. Recently, an algorithm called Explainable AI (XAI) has been making significant progress in this regard. It can help to explain why a particular prediction has been made based on the features of the dataset [6]. The interpretation of a prediction in this regard is important, as a faulty prediction can have precarious consequences. It can improve our faith in the model used, as well as help us in improving an untrustworthy model. In this research work, we primarily analyze various machine-learning models for the purpose of estimating the likelihood of autism. Furthermore, we use some XAI models to gain some insight into the predictions made by the model.

II. Literature Review

There are few diagnosis tools for ASD such as Autism Diagnostic Interview (ADI), Observation Schedule Revised (ADOS-R). Clinical diagnosis methods have been shown a significant success in cases related to ASD. Researchers recently started to use a machine learning approach for diagnosis of ASD [7]. In 2018, Vaishali et al. proposed a swarm intelligence based feature selection with an aim to select fewer features and obtain maximum accuracy. They applied binary firefly feature selection and filtered 10 feature subset among 21 features in the dataset.They obtained 87.67% to 99.66% accuracy on the original dataset for different machine learning algorithms. They got 87.67% accuracy in the KNN algorithm with k=5 and 99.66% with support vector machines. After reducing the feature to 10, they got 93.84% accuracy on KNN [8]. Thabtah et al. in 2019 used logistic regression for autism classification. They used 10 validation methods. They trained the dataset in ten different subsets in order to produce error rate. Finally, all the error rates were counted into average to produce one global error rate for the classifier. They got 87.74% accuracy in their different subset of the dataset [9]. In 2021, Sujatha et al. applied PCA to extract the feature and used different types of machine learning models including SVM, KNN on their data set. They divided the dataset into different age groups and trained on each subset of the dataset separately. They got the highest accuracy in the adult dataset, with accuracy of 99.7% [10]. Tartarisco et al. applied machine learning and computation intelligence to improve classification accuracy of the Q-CHAT. They proposed to select features before approaching to train their model with SVM, KNN, LR and other machine learning algorithms. They identified that SVM showed the highest accuracy of 90% compared to other models like KNN and NB which obtained accuracy of 84% and 89% respectively [11]. Alteneiji et al. applied Ada Boost, CV boosting and neural network to classify ASD in different ages of people including child, adolescent, and toddler. The Ada Boost results with 95.41% accuracy in child-database, on the other hand 75% accuracy on adolescent people. However, they get 94% accuracy on the toddler. For CV boosting, they got 92% accuracy on children, 82% for adolescents and 94% for the toddler [12].

III. Methodology

A. Dataset

The dataset used here consists of different information of 704 people participating in a questionnaire survey [13]. Among those 704 people, 189 were classified as likely to be ASD prone. The question set is clinically approved and provided by National Institute of Health Research (NHS) [14]. There are a set of ten questions with 4 options to choose from for each one. Different questions had different levels of numbering. There is only one point available for every question. For 1, 7, 8, and 10, one point is given if the answerer agreed with them definitely or somewhat. For 2, 3, 4, 5, 6, and 9, one point is assigned if the answerer disagrees slightly or strongly. A sum of score over 6 is considered to be likelihood of having ASD. Apart from the 10 question answers, there were also the age, gender, ethnicity, and some other demographic information in the dataset.

B. Data Preprocessing

Before feeding the data to a machine learning algorithm, the dataset had to be preprocessed to ensure quality [15]. The preprocessing steps for this dataset were cleaning, feature scaling and label encoding. The cleaned dataset was then explored by different data analysis techniques and several visualizations.

1) Data Cleaning and Feature Engineering: Firstly, the rows with missing values were dropped. There were two missing values in the age column, so those rows were dropped. However, the missing values in ethnicity column were mapped to the country of residence column for avoiding data loss, as it had about 100 missing values. A new column of "age group" was created by taking cuts based on age column. From the age distribution, it was seen that one particular point had an age of 383, which is clearly an outlier. It was dropped as well. Some other irrelevant features were also dropped. Finally, there were 701 data points with 16 feature columns.

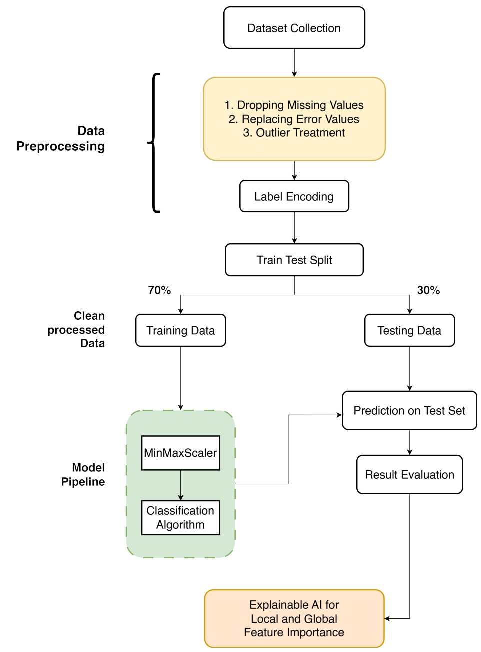

Fig. 1. Methodology

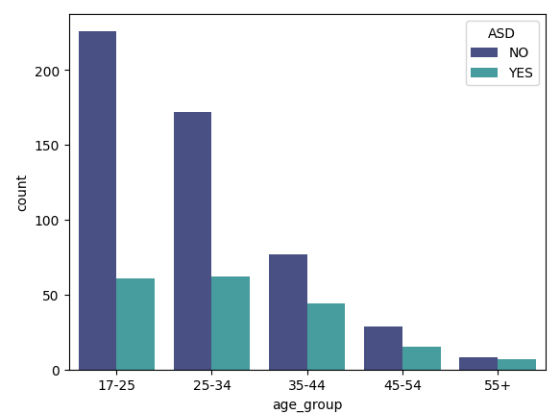

Fig. 2. ASD Distribution in Different Age group

2) Label Encoding: Apart from the ten question columns, all other columns had categorical values. To fit the models, the algorithms require numerical values [16]. Label encoding was opted to solve this problem. Gender, Jaundice, Previous history of autism and the target column "ASD" were fitted to the label encoder. This changed their values to numeric representation.

3) Statistical Analysis: Correlation analysis and Chi^2 test were opted to understand how the features associated with the target column. These tests allowed us to determine which features are statistically significant without applying any prior machine learning algorithms. Correlation heatmap showed that the result column had a correlation of 0.82 with ASD, which is much higher than limit 0.7. This is expected as it is stated in NHS questionnaire [14]. So the result column was dropped. The Chi^2 chart showed all the selected columns, specially the ten questions, were statistically significant here.

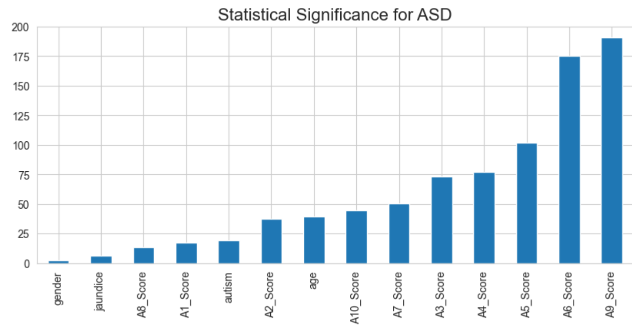

Fig. 3. Statistical Significance of Different Features

4) Feature Scaling: As the different features here had different range of values, scaling was opted. The objective of scaling is to transform values of different features to a single scale. The algorithms using distance for convergence, such as support vector machines, K nearest neighbours, logistic regression greatly benefits from scaling. For scaling, Min Max Scaling was implemented in the dataset. The scaled dataset was then fed to the machine learning models.

C. Machine Learning models

For the purpose of this classification problem, 8 different machine learning models have been tested: Decision Tree, Random Forest, Adaptive boosting, Gradient Boosting, eXtreme Gradient Boosting, Logistic Regression, Support Vector Machines and K-Nearest Neighbour.

1) Decision Tree: Decision trees are rule based machine learning algorithms. They have two types of nodes: decision nodes and leaf nodes. The rule based decisions take place within the decisions nodes. Further nodes are generated from a decision node until a leaf node is reached which generates a prediction [17].

2) Random Forest: A random forest is an ensemble of multiple decision trees. It generates various decision trees by bootstrapping the main dataset. The final prediction is made from all the decision trees constructed via bagging. This ensemble approach increases the robustness of the model [18].

3) Adaptive Boosting (Ada Boost): Adaptive Boosting is an ensemble of numerous consecutively constructed weak learners. Each weak learner is a decision stump constructed by focusing on the wrong predictions of the previous stump. This ensemble of weak learners generalizes the data very well [19].

4) Gradient Boost (GB): Gradient boosting is another popular machine learning algorithm in the boosting family . It is an ensemble of weak learners where the later weak models are fit on the pseudo residuals of the previous weak models rather than being fit on the actual labels. This increases the generalization capability of the model significantly [20].

5) eXtreme Gradient Boosting (XGBoost): XGBoost is an extended version of GB that can employs regularization techniques such as L1 and L2 to reduce over-fitting. XGBoost also uses greedy algorithm for increasing the speed of decision splits and enabling scalability [21]. .

6) Logistic Regression: Logistic regression uses the logistic function to predict the probabilities of the target variable by taking a linear combination of the features as its input. The optimal weights for the linear combination are found by using an optimization function. This model works simple and works quite well for classification [22].

7) Support Vector Machines: This model generates a hyperplane to separate the instances of various classes. The data points which have the minimum distances from the hyperplane are called support vectors. The model tries to maximize the distances of the support vectors so that it can generalize better [23].

8) K-Nearest Neighbour: K-Nearest Neighbour is a simple machine learning algorithm which generates the label for a prediction based on the distance from K nearest data points. It uses various distance functions such as Euclidean distance, Manhattan distance, etc. After finding out the closest data points, it uses voting to classify the instance [24].

D. Evaluation Metrics

As this is a classification problem, the trained algorithms were then tested using accuracy, precision, recall and f1-score. The test set was also viewed in a confusion matrix to see if the algorithms were making any type 1 or type 2 errors in classifying. Lastly, as the dataset was imbalanced, roc auc score was also calculated. Accuracy is the percentage of correctly classified samples. Precision is the percentage of predicted positive samples that are actually positive. Recall is the percentage of actual positive samples that are correctly predicted. F1 score is the harmonic mean of precision and recall. It is a measure of overall performance that takes into account both precision and recall. ROC-AUC is a measure of the ability of a classifier to distinguish between positive and negative samples. It is a value between 0 and 1, with a higher value indicating better performance.

E. Explainable AI

Explainable AI's are used to disclose the black box nature of the machine learning algorithms. This paper puts an effort in understanding how each feature can impact the classification made by the algorithms, thus interpreting using both global and local explanations. Local explanation shows how features are associating with a particular prediction. Global explanation shows the impact of all the features in overall prediction made by an algorithm. SHapley Additive exPlanations (SHAP) is such an AI that can show both local and global explanations [25] SHAP takes a game theoretic approach in determining the SHAP values for all the features and instances. These values then can be easily visualized for an easy understanding. Another explainable AI called ELI5 (Explain like I am 5) was also used to see both local and global explanations [26].

This one also showed local and global results like SHAP, but in a very simpler form. Lastly, another explainable AI that focused on local explanations, LIME, was also implemented [27]. LIME is a model-agnostic interpretability technique that can be used to explain any machine learning model. It works by locally approximating the model with a linear interpretable model. To our knowledge, this comparison of explainable AI results to this extent is a unique approach.

IV. Result Analysis

The table I show that the Ada Boost algorithm is able to correctly classify all the samples in the test set, and it has a high precision, recall, and F1 score. This indicates that the Ada Boost algorithm is a very good classifier for this task.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT ALGORITHMS

Algorithm	Accuracy	Precision	Recall	F1	ROC-AUC
LR	0.99	0.99	0.98	0.99	0.98
DT	0.86	0.82	0.85	0.83	0.85
GB	0.95	0.94	0.95	0.94	0.95
Ada Boost	**1**	**1**	**1**	**1**	**1**
RF	0.94	0.93	0.91	0.92	0.91
SVC	0.96	0.96	0.95	0.95	0.96
KNN	0.94	0.91	0.94	0.92	0.94
XGB	0.95	0.93	0.94	0.93	0.94

The Logistic Regression and Support Vector Classifier algorithms also have a good performance, but they are not as good as the Ada Boost algorithm. The Logistic Regression algorithm has a slightly lower accuracy and F1 score than the Ada Boost algorithm, and the Support Vector Classifier algorithm has a slightly lower recall than the Ada Boost algorithm. The other algorithms in the table have a lower performance than the Ada Boost, Logistic Regression, and Support Vector Classifier algorithms. However, they still have a good performance overall, with accuracy scores above 0.90.

TABLE II
RESULT FROM CONFUSION MATRIX

Algorithm	TP	FP	FN	TN
LR	153	1	4	53
DT	136	18	11	46
GB	148	6	4	53
Ada Boost	**154**	**0**	**0**	**57**
RF	146	8	8	49
SVC	154	0	57	0
KNN	141	13	15	42
XGB	148	6	5	52

The following are some key observations from the table II:

1) All the algorithms have a high true positive rate (TP), which means that they are able to correctly identify the actual class of most of the samples in the test set.
2) The Ada Boost and SVC algorithms have the highest true positive rates, with a TP of 154 (100%) for both algorithms. However, SVC falls short as it fails to detect any True Negative values.
3) The tree based algorithms outperformed distance based algorithms by a big margin.

From these two tables, it is clear that, accuracy can not be trusted as the only metric for a machine learning model as a high accuracy model can perform poorly in test set, as it is seen from the confusion matrix. This finding worked as the inspiration towards model explanation.

V. Explainable AI Analysis

A. Global Explanation

XGBoost model was fitted with all the explainable AIs. SHAP and ELI5 provided with global explanations. The SHAP Bee swarm plot can show how features affect the target column or ASD here.

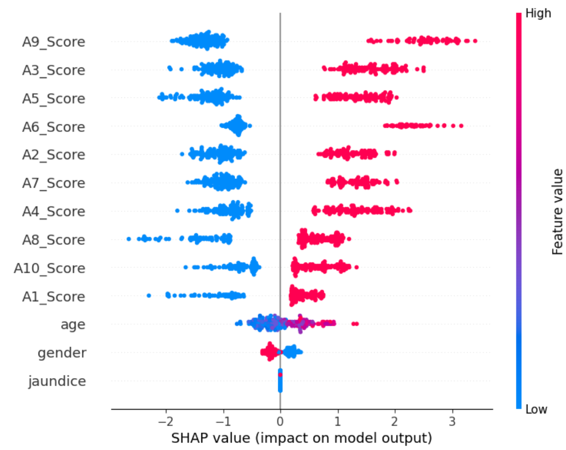

Fig. 4. SHAP Bee Swarm Plot

The partial SHAP summary plot shows that the A9_Score feature has the biggest impact on the prediction, followed by the A3_Score and A5_Score features. These features all have a positive impact on the prediction, meaning that a higher value (1) for these features will lead to a higher prediction (likelihood of growing ASD). The A6_Score, A2_Score, A7_Score, and A8_Score features also have a positive impact on the prediction, but their impact is smaller than the impact of the A9_Score, A3_Score, and A5_Score features. This shows that the question set had the significant effect on the target column other than gender or age. It is also seen that higher age increases the risk of ASD.

The ELI5 global explanation shows the weights of different features while fitting the XGBoost on training data. The weights are numbers between 0 and 1, and they indicate how important each feature is for determining the overall score. The A9_Score feature has the highest weight (0.4507), which means that it is the most important feature for determining the prediction outcome. The age and gender features have the lowest weights (0.0181 and 0.0052, respectively), which means that they are the least important features for determining the prediction.

B. Local Explanation

For the local explanation, a single instance from the test set was chosen. Here it was the instance number seven.

Weight	Feature
0.4507	A9_Score
0.0917	A4_Score
0.0732	A3_Score
0.0707	A6_Score
0.0646	A5_Score
0.0613	A1_Score
0.0569	A8_Score
0.0437	A10_Score
0.0395	A7_Score
0.0244	A2_Score
0.0181	age
0.0052	gender
0	jaundice

Fig. 5. ELI5 Global Explanation

The actual target value was 1 (Likely to have ASD), and the prediction was also 1. XGBoost made this prediction with 67% possibility.SHAP, LIME and ELI5 all had different representations when it comes to local explanation.

1) SHAP Plots: The SHAP provided with waterfall plot and force plot, which made it easy to understand. The SHAP values represent the contribution of each feature to the model's prediction, and the waterfall plot shows how the prediction changes as each feature is added or removed. The waterfall plot starts at the baseline prediction, which is the average prediction of the model across all data points. Then, each feature is added to the model one at a time, and the waterfall plot shows how the prediction changes with each addition. The

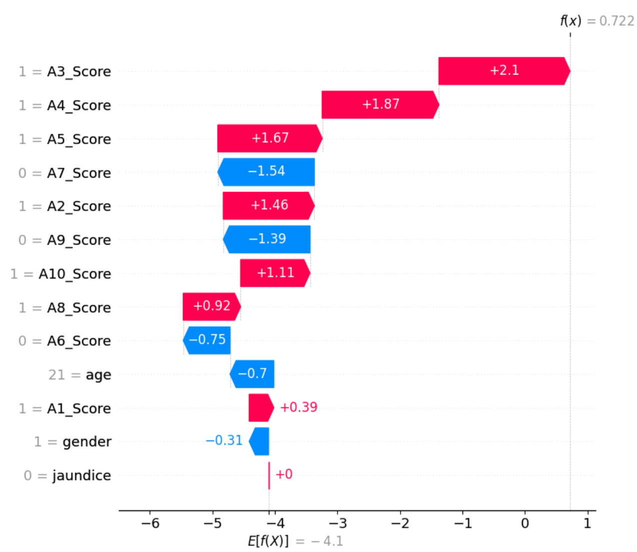

Fig. 6. Feature Importance Waterfall Plot for Instance No: 7

height of each bar in the waterfall plot represents the SHAP value for that feature. Positive SHAP values indicate that the feature increases the prediction, while negative SHAP values indicate that the feature decreases the prediction. The same can also be viewed by the force plot. It shows how each feature was pushing the prediction towards or away from the baseline prediction. As the plots show, A7, A9, A6 scores along with

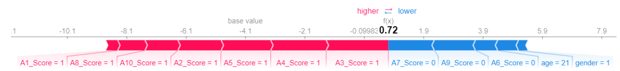

Fig. 7. Feature Importance Force Plot for Instance No: 7

age and gender were pushing the prediction to 0 (Not ASD). But the rest of the features pushed the prediction to be 1.

2) LIME Plot: The LIME plot shows the coefficients of the linear interpretable model, which represent the importance of each feature in the prediction. Here A9, A6 and A7 Scores coefficient sum suggested that the prediction should be Not ASD, but coefficient sum of other 7 features predicted instance 7 to be ASD. The bars on the left also show this possibility. It reveals, however, that SHAP and LIME have utilized nearly identical attributes in determining the forecast.

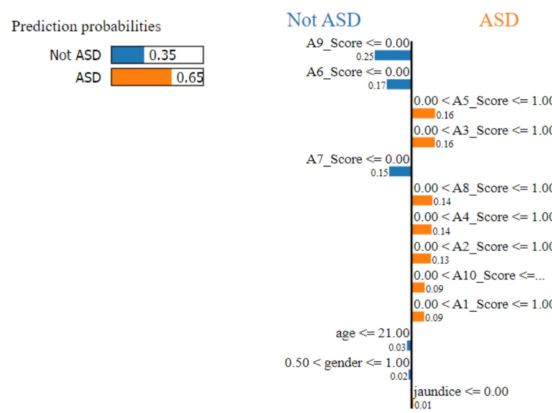

Fig. 8. LIME Interpretation for Instance No: 7

y=1 (probability **0.647**, score **0.607**)

Contribution?	Feature
+1.869	A4_Score
+1.757	A2_Score
+1.515	A3_Score
+1.340	A5_Score
+1.245	A8_Score
+1.188	A10_Score
+0.311	A1_Score
-0.494	gender
-0.965	A6_Score
-1.366	age
-1.685	<BIAS>
-1.934	A7_Score
-2.174	A9_Score

Fig. 9. ELI5 Interpretation for Instance No: 7

3) ELI5 Plot: As seem before, ELI5 takes a more strimmed down approach when showing feature contributions. The most important feature in the prediction is A4_Score, which increases the probability of participation by 1.869. The next most important feature is A2_Score, which increases the probability of participation by 1.757. Some other features in the table have negative contributions, which means that they decrease the probability of participation. For example, the A6_Score feature decreases the probability of participation by -0.965.

The overall probability of participation for this person is 0.647. This probability is calculated by adding up the contributions of all the features.

VI. Conclusion

The significance of early screening for autism cannot be overstated, as it can commence early intervention. Machine learning is an approach that accelerates the screening process with automation and effective results. In this research work, we have evaluated 8 machine learning models for an autism screening dataset. The final results show that the Ada Boost model outperforms all the other models, achieving a score of 1 in all the performance metrics. Logistic regression attained slightly lower scores than the Ada boost algorithm with a score of 0.99 for Accuracy, Precision, and F1 and a score of 0.98 for Recall, Precision respectively. The decision tree attained the lowest scores across all the metrics due to its simplistic approach. The rest of the models performed similarly, with slight deviation among the scores. The results, therefore, corroborate the efficiency of machine learning models in detecting ASD accurately. Thereafter, we used three different explainable AI models to analyse the predictions from XGBoost. Global explanations by SHAP and ELI5 show that A9 score is the most important feature in detecting ASD, whereas jaundice is observed as the least important feature for the same. In fact, the explanations suggest that the features other than the screening scores, such as age, gender, and jaundice contribute the least for the prediction. This work can be further extended by building a more robust dataset, conducting an age based study. The inclusion of other machine learning or deep learning models may also provide more discerning knowledge about the features likely raising and answering new questions.

Acknowledgment

We would like to convey our gratitude to Professor Dr. Md. Golam Rabiul Alam for his supervision and guidance on this project.

References

[1] A. S. Rabba, C. Dissanayake, and J. Barbaro, "Parents' experiences of an early autism diagnosis: Insights into their needs," *Research in Autism Spectrum Disorders*, vol. 66, p. 101415, 2019.

[2] N. Gabbay-Dizdar, M. Ilan, G. Meiri, M. Faroy, A. Michaelovski, H. Flusser, I. Menashe, J. Koller, D. A. Zachor, and I. Dinstein, "Early diagnosis of autism in the community is associated with marked improvement in social symptoms within 1–2 years," *Autism*, vol. 26, no. 6, pp. 1353–1363, 2022.

[3] S. Soto, K. Linas, D. Jacobstein, M. Biel, T. Migdal, and B. J. Anthony, "A review of cultural adaptations of screening tools for autism spectrum disorders," *Autism*, vol. 19, no. 6, pp. 646–661, 2015.

[4] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," *Health informatics journal*, vol. 26, no. 1, pp. 264–286, 2020.

[5] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets," *Health Information Science and Systems*, vol. 9, pp. 1–13, 2021.

[6] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, 2020.

[7] S. Raj and S. Masood, "Analysis and detection of autism spectrum disorder using machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020.

[8] R. Vaishali and R. Sasikala, "A machine learning based approach to classify autism with optimum behaviour sets," *International Journal of Engineering & Technology*, vol. 7, no. 4, p. 18, 2018.

[9] F. Thabtah, N. Abdelhamid, and D. Peebles, "A machine learning autism classification based on logistic regression analysis," *Health information science and systems*, vol. 7, pp. 1–11, 2019.

[10] R. Sujatha, S. Aarthy, J. Chatterjee, A. Alaboudi, and N. Jhanjhi, "A machine learning way to classify autism spectrum disorder," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 6, pp. 182–200, 2021.

[11] G. Tartarisco, G. Cicceri, D. Di Pietro, E. Leonardi, S. Aiello, F. Marino, F. Chiarotti, A. Gagliano, G. M. Arduino, F. Apicella *et al.*, "Use of machine learning to investigate the quantitative checklist for autism in toddlers (q-chat) towards early autism screening," *Diagnostics*, vol. 11, no. 3, p. 574, 2021.

[12] M. R. Alteneiji, L. M. Alqaydi, and M. U. Tariq, "Autism spectrum disorder diagnosis using optimal machine learning methods," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 9, 2020.

[13] F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment," in *Proceedings of the 1st International Conference on Medical and health Informatics 2017*, 2017, pp. 1–6.

[14] 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 in 1,000 cases and 3,000 controls," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 51, no. 2, pp. 202–212, 2012.

[15] M. F. Kilkenny and K. M. Robinson, "Data quality:"garbage in–garbage out"," pp. 103–105, 2018.

[16] M. K. Dahouda and I. Joe, "A deep-learned embedding technique for categorical features encoding," *IEEE Access*, vol. 9, pp. 114 381–114 391, 2021.

[17] C. Kingsford and S. L. Salzberg, "What are decision trees?" *Nature biotechnology*, vol. 26, no. 9, pp. 1011–1013, 2008.

[18] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[19] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, "Advance and prospects of adaboost algorithm," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2013.

[20] X. Zhu and J. Chen, "Risk prediction of p2p credit loans overdue based on gradient boosting machine model," in *2021 IEEE international conference on power, intelligent computing and systems (ICPICS)*. IEEE, 2021, pp. 212–216.

[21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[22] S. Sperandei, "Understanding logistic regression analysis," *Biochemia medica*, vol. 24, no. 1, pp. 12–18, 2014.

[23] W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[24] J. Laaksonen and E. Oja, "Classification with learning k-nearest neighbors," in *Proceedings of international conference on neural networks (ICNN'96)*, vol. 3. IEEE, 1996, pp. 1480–1483.

[25] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[26] M. Korobov and K. Lopuhin, "Eli5. 2016," *Available online: eli5. readthedocs. io/(accessed on 5 November 2022)*.

[27] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 2016, pp. 1135–1144.