Evaluating Traditional and Advanced Machine Learning Approaches for Autism Diagnosis in Toddlers

Ananya Bajpai
Department of Artificial Intelligence and Data science
Indira Gandhi Delhi Technical University for Women
Delhi, India
ananya011btcseai23@igdtuw.ac.in

Ananya Bhardwaj
Department of Artificial Intelligence and Data science
Indira Gandhi Delhi Technical University for Women
Delhi, India
ananya012btcseai23@igdtuw.ac.in

***Abstract*—The impact of autism spectrum disorder (ASD) on a person's social, academic, and occupational functioning have led to its association with disorders such as depression, anxiety, epilepsy, and others. Early intervention is essential for the effective therapy of autism spectrum disorder (ASD), which can significantly improve the developmental results of affected children. Screening children for autism is challenging because of the variety of developmental milestones and complexities of early signs. An early diagnosis is helpful for the development of behavior and language. However, since there is no treatment, diagnosis might be difficult; therefore, the goal is to minimize symptoms while maximizing a person's abilities. This work offers a comprehensive approach to autism screening for children by analyzing a dataset of behavioral and demographic characteristics using machine learning algorithms.**

***Index Terms*—Machine learning, Autism in Toddlers, Autism Spectrum Disorder, Autism Diagnosis, Diagnosis, Toddlers**

I. Introduction

Autism Spectrum Disorder (ASD) is a developmental disorder primarily associated with difficulties within social interaction, communication activities both verbal and nonverbal, and the presence of several repetitive activities and behaviors which normally tend to surface within the first two years of a child's life (Centers for Disease Control and Prevention, 2014). According to the US Centers for Disease Control and Prevention, one out of every sixty-eight children globally is diagnosed with ASD, which translates to more than 68 million individuals. Most of these people reside in America, over 2 million have been reported. The different diversity of symptoms and the severity of symptoms has resulted in a majority of experts categorizing this disease as a spectrum disorder. When Connected with early interventions, they can be beneficial in improving the functioning of an individual with autism and also their quality of life. Typical signs are disorders characterized by a variable degree of poor eye contact or shyness; getting into conversations but being unsuccessful due to an absent expression, or facial expression which does not match the situation or social scene.

Effective screening and diagnostic instruments for autism spectrum disorder (ASD) are desperately needed. Recent researchers are investigating machine learning (ML) to improve the processes of ASD screening and assessment. Recent advancements include customized therapy plans and the provision of several modalities comprising visual attention and voice.

Nevertheless, while illustrating these applications' potential, some will inevitably depend on certain information or specifics of a model's design, making them less reliable and flexible. Our paper employs different modeling tools like XGBoost, Logistic Regression, Decision Tree, Random Forest, SVM. This was absent in related works that only focused on one data modality. In this light, By analyzing several models and selecting the best ones, the study seeks to fill in the gaps in the literature. In doing so, the research aims at finding the most efficient model for diagnosing the disorder in question. While at the same time considering machine learning applications in practice. Once more, this study answers the question of why an early diagnosis is essential. Early diagnosis of autism is important, because interactions between the family and the autistic kid may have an impact on the child's development.

The rest of the paper is structured such that it presents as follows: the review of the literature introduces the progress of earlier studies relevant to the problem, followed by the methods, research findings, and their analysis. In the methodology we have discussed about the dataset as well. We have added the ROC curves for our models to compare their performance efficiently.

II. Literature Review

Due to recent developments in machine learning, diagnosing individuals with autism spectrum disorder is becoming easier. Early studies were limited to certain data types such as visuals or vocals, but are shifting to more advanced methodologies through the integration of multiple sources of data in a single analysis for improved accuracy in diagnostics.

Machine learning tools have also been used to meet the demands of people with autism spectrum disorder. Maenner et al. (2021) for example; studied the use of Applied Behavior Analysis techniques through patient-based and collaborative filtering approaches and recorded an accuracy level of 81-84% in the personalization of ABA treatment for patients. Therefore, this initial study had offered a basis on how individualized treatment approaches for ASD could be further investigated.

With this basis, the further studies carried out later have pointed out the effectiveness of data augmentation techniques such as in eye tracking and coupling it with machine learning. Kohli et al. (2022) pointed out the effectiveness of analyzing social visual attention to detect metabolic disorders at an early stage. They also showed the usefulness of support vector machines and neural networks in their studies. Their results showed the limitations and advantages of SVD based methods in performance gain and cost efficiency. Also, Lau et al. (2022) used a mixed model, combining eye tracking with such deep architectures as GoogleNet and ResNet-18. This allowed them to reach accuracy of 95.5% and 94.5%, accordingly.

To diagnose ASD, recent research has kept pushing the limits of machine learning techniques. In addition to optimization techniques, Rabbi et al. (2023) used a gray wolf optimizer with SVM and feature selection. He did this by utilizing genetic and personal attribute data to attain a very high 99.96% accuracy.Deng et al. (2022) and Cao et al. (2023), investigated the use of complex architectures. For eg transformers for analyzing temporal data and facial expressions.

It shows that while these developments represent significant progress, they also highlight the challenges involved in training sophisticated models. Despite their potential, these methods frequently require more sophistication and longer training times.

III. METHODOLOGY

A. Dataset Description

We initially needed a dataset that included information on toddlers' physical and behavioral characteristics as well as whether they had ASD in order to create machine learning models that could predict when ASD will manifest in young children. This study made use of the Autism Spectrum Disorder dataset from the University of California, Irvine's Machine Learning Repository, which is a reliable source for machine learning datasets (Thabtah 2018). This dataset has important characteristics that are examined when toddlers are screened for autism. The dataset, which has 1,054 entries, comprises 18 variables that relate to various features. Ten of these variables—items A1 through A10 in the table below—are questions that determine whether a youngster exhibits symptoms of ASD.

The Quantitative Checklist for Autism in Toddlers (Q-CHAT-10), a clinically accepted assessment for toddlers, is the source of these ten questions (Booth et al. 2013). Items A1 through A9 were given a value of 1 if the response pointed to a behavior linked to ASD features, and 0 for all other responses.

If a response to question A10 mentioned a behavior linked to a higher risk of ASD, it was given a value of 1. The Q-CHAT-10 score is the sum of these question values. A toddler is significantly more likely to exhibit traits of an ASD if their score is greater than three. There are no obvious symptoms of ASD if the score is 3 or lower.

Variables	Corresponding Q-chat-10-Toddler Features
A1	Child responds to their name with eye contact.
A2	Ease of engaging the child in eye contact.
A3	Child points to show what they want.
A4	Child points to share interest in objects/events.
A5	Child engages in make-believe or pretend play.
A6	Child looks where others are looking.
A7	Child comforts others when they seem upset.
A8	Quality of the child's first spoken words.
A9	Child uses gestures like waving or nodding.
A10	Child stares into space or at objects without purpose.

TABLE I
DETAILS OF VARIABLES MAP PING TO THE Q-CHAT-10 SCREENING METHODS

Specific characteristics of each toddler that help determine the variables influencing the occurrence of autism spectrum disorder make up the remaining elements. These 10 questions, in addition to other features, serve as the predictor variables for identifying whether a toddler shows signs of autism spectrum disorder (ASD). The target variable is the Q-CHAT-10 score, which is determined by utilizing the responses to the questions.

It should be noted that there were five potential responses for each question, each of which represented a different level of agreement with the assertion made in the question. For instance, the response options for A1 are "Always," "Usually," "Sometimes," "Rarely," or "Never." There is a specified set of five responses for each question that are relevant to the question. Responses to questions A1 through A9 were mapped to a 1 if they indicated behaviors associated with ASD traits.

On the other hand, the response was mapped to a 0 if it was one of the replies that represented behaviors not typically associated with ASD. For Question A10, a response was mapped to a 1 if it was one of the answers that indicated behaviors strongly associated with ASD. Otherwise, the response was mapped to a 0 if it indicated less likelihood of ASD traits. It should be noted that this data mapping was carried out by the authors of the dataset (Thabtah 2018), not by the authors of this paper.

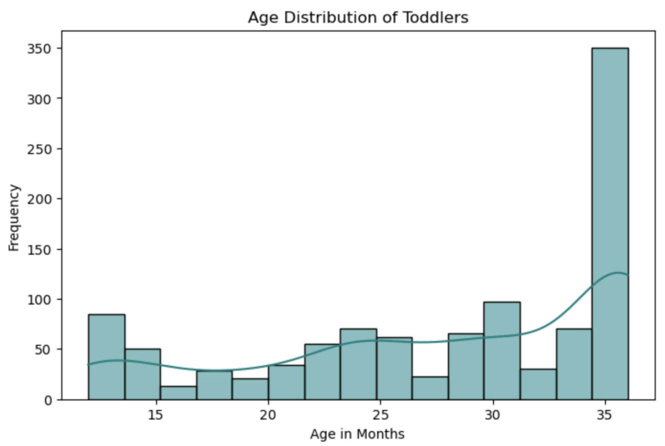

Fig. 1. **Age Distribution of Toddlers in months**

B. Data Prepossessing

It was observed that the dataset did not contain any missing values. This was further confirmed since the 'Non-null Count' accurately matched the 'RangeIndex'. However, this check is vital for correctly preprocessing the data.

Furthermore, Missing Value Handling was carried out using the SimpleImputer. For columns that were numerical, the null values were imputed using the median value. For columns which were categorical, the missing values were imputed with the most frequent value. Target Variable Encoding was carried out where the line: '**data['Class/ASD Traits']=data['Class/ASD Traits'].map({'Yes' :1, 'No' :0})** was used to convert the target variables into numerical form (1 for 'Yes', 0 for 'No'). This was done to provide numerical input to machine learning models.

One-hot encoding was carried out by using the '**pd.get_dummies'** function to convert categorical variables into indicator variables. This step was carried out to convert features like 'Sex', 'Ethnicity', 'Jaundice' and 'Family_mem_with_ASD' into numerical columns.

The class distribution was checked to verify if the dataset was imbalanced. In case of imbalance, SMOTEENN was applied to handle the class imbalance by oversampling class which was minority and cleaning up the majority class.

This data was finally organized into training sets. The training data was split in a fashion where the '**test_size=0.2',** thus the training data comprised of 80% original dataset. The dataset originally had 1054 entries which were reduced to 843 samples.

C. Equations

1) Accuracy: The proportion of accurately predicted cases—both favorable and negative—to all instances is known as accuracy. It provides the model's overall efficiency metric.

Where:

- False Positives (FP) Type 1 error: Strongly positive cases that were inappropriately classified as such.

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{True Positives (TP)} + \text{True Negatives (TN)} + \text{False Positives (FP)} + \text{Fa}}$$

- False Negatives (FN) Type 2 error: Number of negative cases in mislabeling.
- TN (True Negatives): accurately identified as positive cases
- TP (True Positives): Number of instances which were correctly classified as positives.

2) ROC Score: Plotting the True Positive Rate (TPR) versus the False Positive Rate (FPR) at different threshold values is done via the Receiver Operating Characteristic Curve.A visual depiction of the contrast between FPR and TPR (sensitivity). The ROC curve illustrates how the rates change as you adjust the decision threshold.

3) AUC (Area Under the ROC Curve): Calculates the total area under the ROC curve. It offers a performance metric for every classification threshold. An ideal model has an AUC of 1, while a model with no discrimination ability (equal to random guessing) has an AUC of 0.5.

4) F1 Score: Precision and recall are given equal weight by the harmonic mean of the two. When having an uneven class distribution and need to strike a balance between recall and precision, the F1 score is especially helpful.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

- **Precision** $= \frac{\text{True Positives (TP)}}{\text{True Positives (TP)}+\text{False Positives (FP)}}$
- **Recall** (or Sensitivity) $= \frac{\text{True Positives (TP)}}{\text{True Positives (TP)}+\text{False Negatives (FN)}}$

Where:

- *Precision:* Precision is defined as the proportion of accurately predicted positive observations to all predicted positive observations. It provides an answer to the query, "How many of all the cases that were predicted to be positive were actually positive?"
- *Recall:* The proportion of all observations in the actual class that were accurately predicted to be positive. It provides an answer to the query, "How many of all actual positives did we correctly predict?"

D. Machine Learning Techniques

1) Random Forest: This model uses a technique that creates various decision trees while training to give the mean prediction of the individual trees. This method of combining multiple trees is referred to as Bagging. Here, a subset of data is utilised to train every individual tree such that results are voted upon. This model was selected because it reduces the variance of the model, making it less prone to overfitting.

It can be noted that fig. represents the ROC-AUC score of random forest model.

2) K-Nearest Neighbour: KNN is a non-parametric method that keeps track of every case and uses a similarity pattern to categorize the most recent cases. In KNN, the object can be classified by the majority votes of its neighbours. This is thus, instance-based method because of its ability to directly compare instances rather than learning a specific model on its own.

3) Logistic Regression: It is a simple and popular technique for binary classification issues. In order to predict binary outcomes like "yes" or "no" with easy interpretability and computational efficiency, logistic regression predicts the probability of an outcome by fitting data to a logistic curve.

4) Gradient Boosting: It is a technique that constructs models in order, where every fresh model tries rectifies errors that were manufactured by earlier models. The technique involves fitting a model to the residual error off current model, henceforth minimizing the overall error.

5) XGBoost: It is an advanced implementation of Gradient boosting technique, helping in increasing speed and performance. It consists of techniques are regularized like L1,L2 regularizations. This helps to prevent overfitting. XGBoost is used widely in machine learning environments due to its high efficiency and high accuracy.

6) LightGBM: It is a boosting technique that can be distinguished from XGBoost in a way, that it uses histogram-based framework for finding best splits, which in turn speeds up the training process and reduces memory usage. It is also well-suited for large datasets and high-dimensional data.

7) Support Vector Machine: It works by finding the best border, or hyperplane, to divide the dataset's various classes is how the Support Vector Machine (SVM) classification method operates. Since, SVM uses kernels to convert data into higher dimension, separation becomes easier. It works very well with datasets that have complicated or non-linear relationships.

E. ROC-AUC Figures of ML models

The capacity of each model to distinguish between classes is shown by the ROC-AUC curves for Random Forest, K-Nearest Neighbor, Decision Tree, Gradient Boosting, XGBoost, LightGBM, and Ensemble Voting. An important parameter to assess model efficacy in autism diagnosis is AUC, as a higher value indicates better classification performance.

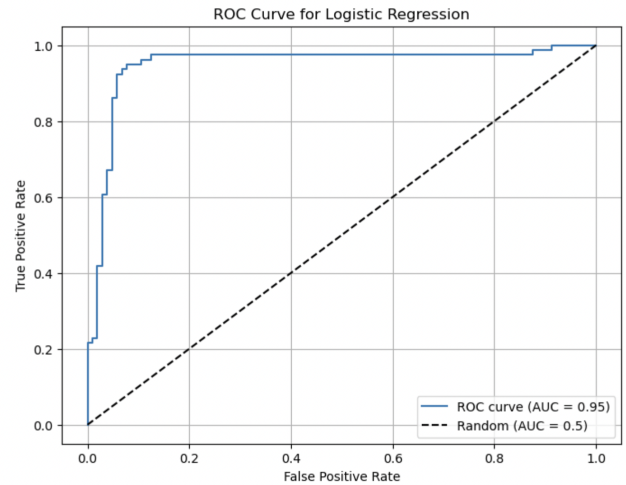

Fig. 2. Logistic Regression's ROC curve

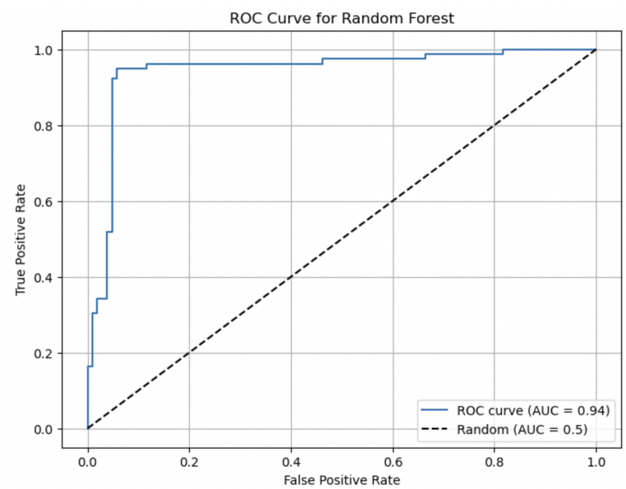

Fig. 3. Random forest's ROC curve

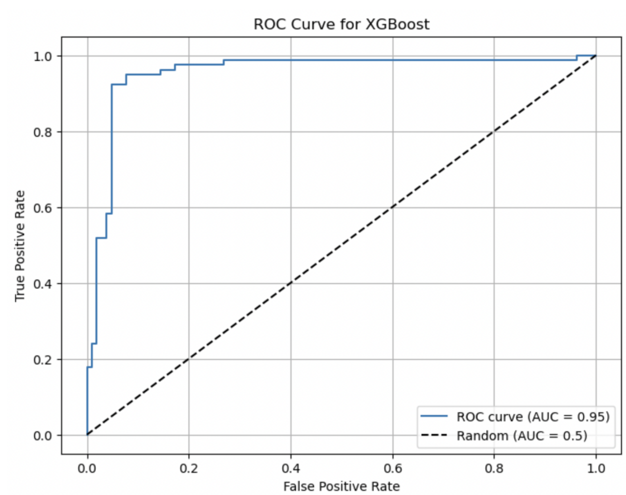

Fig. 4. XGBoost's ROC curve

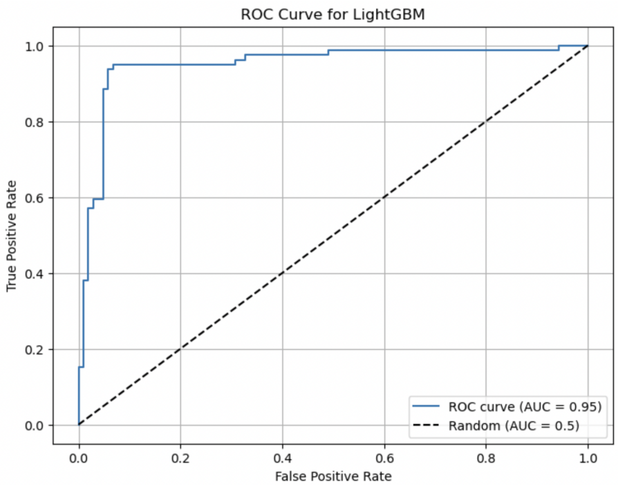

Fig. 5. LightGBM's ROC curve

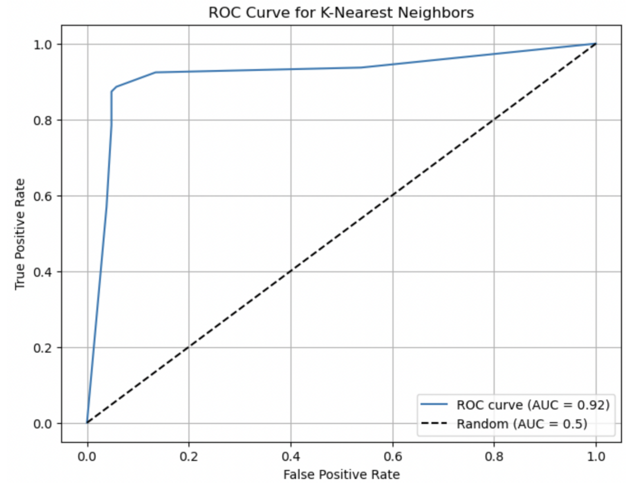

Fig. 8. K-Nearest Neighbour's ROC curve

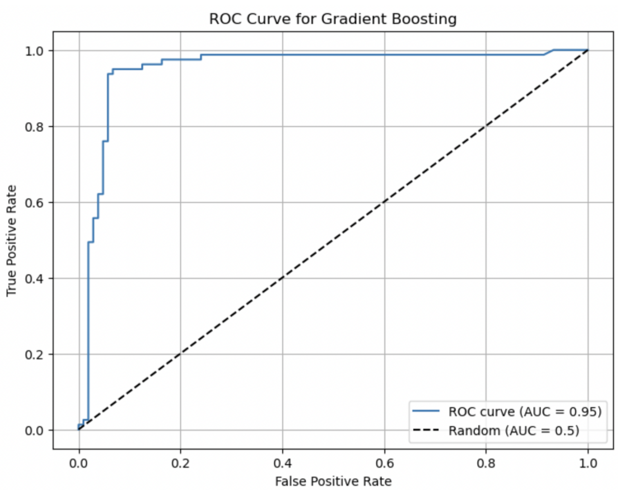

Fig. 6. Gradient Boosting's ROC curve

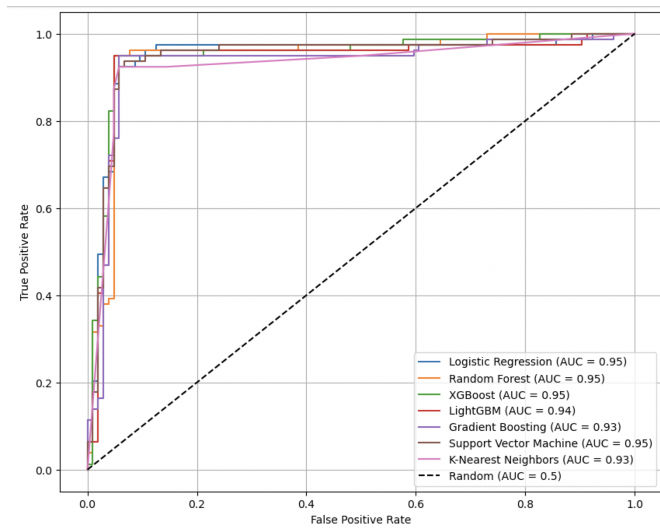

Fig. 9. All models' ROC curves

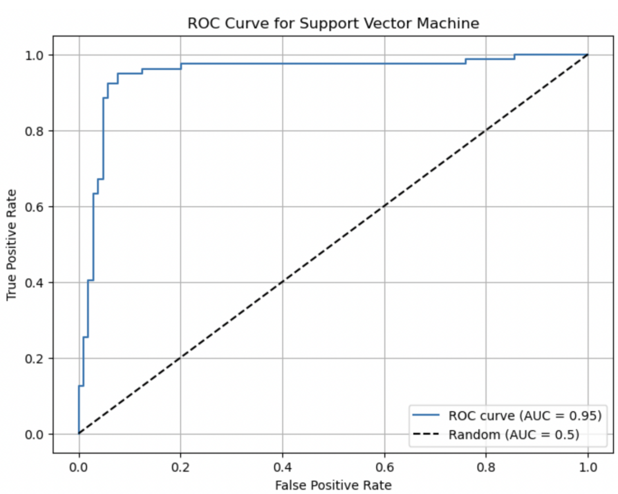

Fig. 7. Support Vector Machine's ROC curve

F. Results and Discussions

In this section, we provide a comparative analysis of the seven models used: Random Forest, K-Nearest Neighbor, Logistic Regression, XGBoost, LightGBM, Support Vector Machine and Gradient Boosting for classification of autistic characteristics present in a toddler. Numerous evaluation metrics were applied. This includes: Accuracy, F1 Score, ROC-AUC Score, Precision, and Recall. Strong performance is necessary in autism screening to avoid missing cases and causing delays in early intervention, as well as to prevent unnecessary stress for families from incorrect identifications. The metric report of all seven models has been compared, analyzed, and presented in this paper. Precise predictions are crucial in evaluating model performance for screening toddlers for autism.

The findings of this study are as follows: Random Forest, Gradient Boosting, and Logistic Regression performed very

Model	Test Accuracy	ROC AUC	Recall	F1 Score
Random forest	94.5355%	0.944133	0.949367	0.937500
Gradient Boosting	93.9891%	0.950463	0.936709	0.930818
Logistic Regression	93.4426%	0.948150	0.924051	0.924051
LightGBM	93.4426%	0.949489	0.949367	0.925926
Support vector machine	92.8962%	0.948759	0.911392	0.917197
XGBoost	92.3497%	0.953627	0.924051	0.912500
K-Nearest Neighbors	91.8033%	0.918208	0.873418	0.901961

TABLE II
MODEL PERFORMANCE METRICS.

well, demonstrating outstanding precision. All three's accuracy reached almost perfect, with Random Forest achieving 94.54% accuracy, Gradient Boosting 93.99%, and Logistic Regression 93.44%. For Random Forest, Gradient Boosting, and Logistic Regression, the corresponding F1-Score, Recall, and Precision were 93.75%, 94.94%, and 93.75%, 93.08%, 93.67%, and 92.59%, respectively. These high numbers demonstrated that the models could accurately identify both positive (possibly autistic) and negative (typically developing) instances. As can be shown, the AUC Scores for Random Forest, Gradient Boosting, and Logistic Regression were 0.94, 0.95, and 0.95, respectively, suggesting approximately flawless accuracy.

The Support Vector Machine (SVM) and LightGBM models also demonstrated strong performance, achieving accuracies of 92.89% and 93.44%, respectively, and AUC Scores of 0.95 for both. .Both recall and F1-scores were high for these models. However, they performed noticeably worse than Gradient Boosting, Logistic Regression and Random Forest.

However, Decision Tree and K-Nearest Neighbors (KNN) models did not perform as well as the others, with accuracies of 91.00% and 91.80%, respectively. Note that with an accuracy of 91% and AUC Score of 0.91, Decision Tree performed the worst out of all of the models. This showcases Decision Tree's limited ability to make precise predictions. A decrease in reliability may result in more misdiagnoses, whether it be failing to detect autism cases or wrongly labeling typically developing children as at risk, affecting early intervention results.

In the context of autism screening, precision plays a crucial role. Therefore, it is to be noted that models like as Logistic Regression, Random Forest, and Gradient Boosting are favored due to their strong ability to accurately identify cases, thus reducing room for misdiagnoses and effectively reducing the risk of untimely care for toddlers.

CONCLUSION AND FUTURE SCOPE

In this research, we explored multiple Machine Learning algorithms in order to screen the presence of autism in toddlers. The models employed - Random Forest, Gradient Boosting, K-Nearest Neighbors (KNN), Logistic Regression, XGBoost, LightGBM, and Support Vector Machine classifier- were trained and evaluated on their performance and prediction tendencies of correctly identifying ASD.

The results indicated that ensemble methods could effectively enhance autism screening efforts. This can help in creating a promising tool for early detection of the condition. The use of multiple classifiers in this study, helped in creating a more comprehensive comparison of the performance metrics. This demonstrated that these models can effectively aggregate the predictions of multiple algorithms to achieve superior results.

The future scope of this study includes integrating more diverse and larger datasets to improve model's robustness and accuracy in screening autism for toddlers. Advanced feature engineering techniques and dimensionality reduction can further enhance model efficiency. A real-time screening tool can also be envisioned, such as a mobile or a web application. Such applications can be developed to help doctors and parents in methods for early detection. In more sense, Explainable AI can be used to make models more interpretable. Future work can also include utilizing neurodevelopmental datasets and genetic data to create more comprehensive screening tool, and expanding this project to recommend doctors and parents on personalized intervention strategies based on each individual cases, which will result in bridging the gap between diagnoses and therapy

REFERENCES

[1] Raj, Suman, and Sarfaraz Masood. "Analysis and detection of autism spectrum disorder using machine learning techniques." *Procedia Computer Science* 167 (2020): 994-1004.

[2] Bone, Daniel, et al. "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion." *Journal of Child Psychology and Psychiatry* 57.8 (2016): 927-937.

[3] Wall, Dennis Paul, et al. "Use of machine learning to shorten observation-based screening and diagnosis of autism." *Translational psychiatry* 2.4 (2012): e100-e100.

[4] Achenie, Luke EK, et al. "A machine learning strategy for autism screening in toddlers." *Journal of Developmental & Behavioral Pediatrics* 40.5 (2019): 369-376.

[5] Omar, Kazi Shahrukh, et al. "A machine learning approach to predict autism spectrum disorder." *2019 International conference on electrical, computer and communication engineering (ECCE)*. IEEE, 2019.

[6] Thabtah, Fadi, et al. "Autism screening: An unsupervised machine learning approach." *Health Information Science and Systems* 10.1 (2022): 26.

[7] Thabtah, Fadi, Neda Abdelhamid, and David Peebles. "A machine learning autism classification based on logistic regression analysis." *Health information science and systems* 7.1 (2019): 12.

[8] Tartarisco, Gennaro, et al. "Use of machine learning to investigate the quantitative checklist for autism in toddlers (Q-CHAT) towards early autism screening." *Diagnostics* 11.3 (2021): 574.

[9] Priyadarshini, Ishaani. ”Autism screening in toddlers and adults using deep learning and fair AI techniques.” Future Internet 15.9 (2023): 292.

[10] Selvaraj, Shanthi Palanisamy, Poonkodi Parveen, Summia Monisha, Monisha. (2020). Autism Spectrum Disorder Prediction Using Machine Learning Algorithms. 10.1007/978-3-030-37218-7_56.

[11] Singh, Arjun, et al. ”Using machine learning optimization to predict autism in toddlers.” Proc. Int. Conf. Ind. Eng. Oper. Manag. 2021.

[12]Maenner, Matthew J. ”Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018.” *MMWR. Surveillance Summaries* 70 (2021).

[13]Kohli, Manu, Arpan Kumar Kar, and Shuchi Sinha. ”The role of intelligent technologies in early detection of autism spectrum disorder (asd): A scoping review.” *IEEE Access* 10 (2022): 104887-104913.

[14]Lau, Joseph CY, et al. ”Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study.” *PloS one* 17.6 (2022): e0269637.

[15] Hasan, SM Mahedy, et al. ”Machine learning-based models for predicting autism spectrum disorders.” *Applied Intelligence for Industry 4.0*. Chapman and Hall/CRC, 2023. 27-38.

[16] Deng, Andong, et al. ”Language-assisted deep learning for autistic behaviors recognition.” *Smart Health* 32 (2024): 100444.

[17] Cao, Xu, and Jianguo Cao. ”Commentary: Machine learning for autism spectrum disorder diagnosis–challenges and opportunities–a commentary on Schulte-Rüther et al.(2022).” *Journal of Child Psychology and Psychiatry* 64.6 (2023): 966-967.

[18] Grainger, Catherine, David M. Williams, and Sophie E. Lind. ”Judgment of learning accuracy in high-functioning adolescents and adults with autism spectrum disorder.” Journal of Autism and Developmental Disorders 46 (2016): 3570-3582.

[19] Kohli, Manu, Arpan Kumar Kar, and Shuchi Sinha. ”Robot facilitated rehabilitation of children with autism spectrum disorder: A 10 year scoping review.” Expert Systems 40.5 (2023): e13204.