International Journal of Computer Engineering in Research Trends
ISSN: 2349-7084/ https://doi.org/10.22362/ijcert/2025/v12/i11/v12i1102
Volume 12, Issue 11, November 2025

Research Paper

A Systematic Approach to Autism Spectrum Disorder Diagnosis Using Optimized Machine Learning Models

[1*] Kondalapuri Raajitha, [2] Sudha Thatimakula

[1*] *Research scholar, Department of CSE, SOET, Sri Padmavathi Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India*
Email ID: raajithaprasadphd@gmail.com
[2]*Professor, Department of CSE, SOET, Sri Padmavathi Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India*
Email ID: thatimakula_sudha@yahoo.com

**Corresponding Author(s): raajithaprasadphd@gmail.com*

Received: 11/08/2025 Revised: 27/09/2025 Accepted: 19/11/2025 Published: 30/11/2025

Abstract: Autism Spectrum Disorder (ASD) is a developmental neurodisorder that is marked by impaired socialization and behavioural problems. Things should be detected and treated at an early stage. This project involves the development of a solid predictive model to be applied in ASD by using two datasets- the Autism Spectrum Disorder Screening Data, which is a dataset of a toddler screening program in Saudi Arabia, and the Autism Prediction Dataset. Data preparation, feature engineering, and selection of the algorithm are extensive parts of this process as they would improve the overall classification accuracy of the model. Several machine learning algorithms (Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Neural Networks) are trained using accuracy, precision, recall, and F1-score= measures, and compared. Also, an Explainable Artificial Intelligence (XAI) framework is included to provide greater insight into the model forecasts thereby providing a more transparent decision-making process. According to the results, it is possible to state that the proposed algorithm performs better than the existing methods in ASD detection. Results indicate the possible opportunities of medical tools driven by AI to support decision-making by a caregiver, and one can expect the integration of such models into preventative healthcare systems in the future. Future activities might include expanding the data and make these models applicable.

Keywords: Autism Spectrum Disorder, machine learning, predictive modeling, data analysis, feature engineering, classification, neural networks, explainable AI, early diagnosis, autism screening

---

1. Introduction

Autism Spectrum Disorder (ASD), is a relatively non-homogenous neuro developmental Relative Disorder that interferes with communication, social interaction and behavior. The early diagnosis might change the development of that pathway to a miraculous extent but the primordial mode of diagnosis is time-consuming, subjective and uses the professionalism knowledge. Also artificial intelligence (AI) and machine learning (ML) have emerged that have given the area of automating ASD identification in the search of greater accuracy and efficiency new impetuses as well [1] [2]. As the evaluation of the paper will demonstrate, it is possible to utilize the optimized machine learning algorithm in order to design an early

ASD model of forecast. The two data sets are utilized: the Autism Spectrum Disorder Screening Data of Toddlers in Saudi Arabia, this type of data always trains the prediction models and tests for their abilities. Diagnosis performance was optimized with data preprocessing, feature selection and the different classifiers presented in the paper. The required funds will assist this research in investigating AI-based methods of designing an Effective and Explainable ASD Screening Mechanism that could play a role in further processes able to assist medical staff in assisted diagnosis treatment or be used in female interventional strategy.

ASD model of prediction. The two data sets are applied, the Autism Spectrum Disorder Screening Data of Toddlers in Saudi Arabia which is always used to train the prediction models and test their capacities. The preprocessing of the data, feature selection and the various

classifiers found in the paper were used to optimize the performance in diagnosis. These funds will help this work to examine the AI-driven ways of designing an Effective and Explainable ASD Screening Mechanism which in turn could contribute to subsequent processes capable of supporting medical personnel in the assisted diagnosis treatment or can be applied in female interventional strategy.

The main essence of this research study is the enhancement of ASD prognostication through machine learning approaches and programs. This work uses two published datasets to build a unified predictive model for early diagnosis. The data have to be preprocessed, features should be engineered, and the proper algorithm and measure for assessing performance with four different machine learning classifiers, namely SVM, Random Forests, Gradient Boosting, and Neural Networks, must be identified. Embedded XAI suggestions allow for transparency in the model's decision process that is reasonable to the health care experts.The goal is to optimize the multilabel classification measure in terms of accuracy, precision, recall, and F1-score, compared with the current gels The domain covers vis-a-vis analysis of machine learning models, critical features that determine ASD identification, and viable solutions to the concerns of operational scalability. Nevertheless, the study is limited to only well-structured data and not extending to genetic, neurological and behavioral imaging data that can be used to enhance accuracy of the prediction.

There are several complicated issues about the diagnosis of the Autism Spectrum Disorder (ASD) because it is a complex and multifaceted disorder and has diverse symptoms in each situation and it is considered subjective in the diagnosis [3]. The current diagnostic procedures involve prolonged tests on the modes of behavior and analyses, which need the interpretation of the professionals making the courses of intervention retarded. This is because there is no illuminated, objective and robotic disease detecting mechanism to help in early identification which is significant in the manner or mode of treatment of people. The paper is concerned with the need of a simplified model of prediction based on AI that will enhance the prediction of ASD screening. Using machine learning algorithms and explainable AI mechanisms, the study would be used to train a data-driven scalable framework that can assist medical professionals in making the decisions of diagnosing ASD at an early age and treat it early.

2 Literature Survey

The recent years have seen a tremendous upsurge of machine learning (ML) methodologies in advancing how Autism Spectrum Disorder (ASD) is earlier identified or diagnosed. A diagnostic research of 30,660 participants demonstrated that the ML models used with scanty 28 features yielded high predictive exactness, sensitivity, and specificity in predicting ASD

Similarly, a study employing gait analysis combined with ML techniques highlighted the potential for early ASD detection through non-invasive methods [4] [5].

DL-based identification of ASD has also been made possible. A meta-analysis of 11 predictive trials and 9495 ASD patients revealed that DL methods predict ASD with a reasonable sensitivity, specificity, and area under the curve (AUC) of ASD classification [6]. Besides, it has been recently investigated that radiomics features targeting white matter brain regions in the MRI provide excellent outcomes with over 80 percent of accuracy and provide a statistical association between ASD and white matter defects [7]

Natural language processing (NLP) combined with ML and DL models has been applied to analyze text inputs from social media, achieving an 88% success rate in identifying texts from individuals with ASD [8].

In addition, additional speech transcript-based ML models have been presented, and the Logistic Regression and Random Forest models with 75 percent accurate predictions of the ASD state in children are the keys to success [9]. In their review of the last five years of ASD diagnosis note that the research approaches have been outlined and, by extension, the researchers used structural magnetic resonance imaging (sMRI) and functional MRI (fMRI) characteristics in the development of ML models to perform ASD deign [10]

Recent advancement in ML has been very helpful in early diagnosis and detection of Autism Spectrum Disorder (ASD). Using 30,660 people in a case-control study, ASD was predicted by ML techniques using fewer than 28 variables at a high accuracy, sensitivity, and specificity [11].

Additionally, ML models using speech transcripts have been developed, with Logistic Regression and Random Forest models achieving accuracies of 75% in predicting ASD status in children [12]

The ASD has been recently helped a great deal by the new development of machine learning (ML) in detecting

and diagnosing the Autism condition in the early stages. A large diagnostic study of 30 660 patients revealed that ML models that use a minimal subset of 28 features gave high accuracy, sensitivity, and specificity in predicting ASD [13]

3. Proposed Algorithm

The Decision Tree is another kind of basic machine learning algorithm that applies the tree-like model to arrive at a decision about values of the features. Internal nodes reflect feature-based conditions, branches represent consequences of our conditions and leaf nodes represent classes. The tree divides data recursively with metrics like Gini Impurity or Entropy, in the hope of creating pure subsets. Decision Trees are simpler to interpret and easy to visualize hence they are applicable when it comes to diagnostic use where they are barefoot in need.

Random Forest improves on the Decision Tree by bringing to bear multiple decision trees each trained using random subsets of the data and the features a method called bagging. In prediction the output is taken by majority vote of the trees. This will mitigate variance and will avoid overfitting hence producing a robust and precise model as compared to individual tree. It scales to high dimensions and is able to model interaction between features.

AdaBoost (Adaptive Boosting) is the second approach, which also combines several weak learners, traditionally shallow decision trees, to obtain a strong classifier. Unlike bagging, AdaBoost learners are trained, one after the other, with each learner trained to pay greater attention to the misclassified instances, as learnt by its predecessor. The adaptive reweighting protocol enables AdaBoost to enhance the general prediction. The prediction is by weighted voting of all learners but more accurate models have greater weight.

These suggested models provide superior generalization, resiliency and adaptability, and may hence be suitable candidates of precise and explicable diagnosis of autism spectrum disorder

4. Existing Algorithm

Long Short-Term Memory (LSTM) is an architectural variation of Recurrent Neural Network (RNN) which is trained to learn long-term dependencies in sequential data. The LSTM networks do not commonly show the vanishing or exploding gradients problem typical of traditional RNNs since the LSTM networks possess memory cells that allow them to maintain information across lengthy time steps. The central structure consists of three primary gates, including the input, the forget and the output gate. The gates manage the flow of information and in each time step, the model decides the data to keep, discard, or display. This further classifies LSTM as an excellent model in time-series data or sequence predictions such as patterns of behavior or predictive developmental cues used in diagnosing autism.

Logistic Regression is also a classification algorithm albeit with a name that hints at its use in predicting binary outcomes. It acts as a proxy to model the likelihood of a binary target variable using the sigmoid activation function, to linearly combine input features. This translates into a value between 0 and 1 which is then IF interpreted as a probability. To train the model, maximum likelihood estimator is used, and this tries to optimize the weights to assign the likelihood of each rank as well as the number of ranks in each class. Logistic Regression is well-suited to interpretation and computationally very simple legendary ouija board as it models a linear relationship between any independent variables and the log-odds of a dependent variable. It can work well when the data can separable in a linear manner and when the data is not very complex.

United, LSTM and Logistic Regression will serve to give different insights on the processes of deep learning in sequence-oriented comparisons and the logic of statistical learning and classification, to provide qualified insights in the creation of optimized diagnostic models.This fig 1 illustrates the sequential stages in the development of the Market Pulse platform, from user interface design and real-time data integration to AI-driven sentiment analysis, technical indicator implementation, and AI-powered chatbot integration.

5. Methodology

5.1 Data Collection

The analysis uses 2 datasets; ASD-screening data; the Saudi toddlers, and the Autism prediction data to make sure variance in the demographic and clinical variables. These two data sets share the characteristics of age, gender, questionnaire completion rate and important behavioral variables. Reference links of dataset are Autism Spectrum Disorder [14] and ASD Screening Data for Toddlers in Saudi Arabia [15]

5.2 Data Preprocessing

Raw data are normally inconsistent, missing and outliers. In this regard, a thorough cleaning exercise is done. Entries that do not exist would have them filled in through averaging or median, or would be discarded off as non-representative since it was on a case to case basis.

Outliers are checked by use of statistics and extreme errors are looked at such that the integrity of data is maintained.

5.3 Exploratory Data Analysis (EDA)

Distributions / relations could be determined by the plotting histograms, box plots, and scatter plots in DDA. Using statistical related-functions such as correlation matrices interdependencies are calculated as such and future feature selection can be judged. The other influences that also become evident through the use of EDA superimposing performance figures and will probably encourage inappropriate behaviour of models, are as follows.

5.4 Feature Engineering

This step involve the development of new features or modification of an existing feature in order to upsurge predictive ability. To provide an example, the age groups could be defined by bins and some of the responses to the questionnaire could be aggregated into a composite value. Distillation of the characteristic set will be able to make the model to probe more on the issue of pertinent patterns.

Dimensionality-reduction tools will be used to limit the amount of variables in the data that can be used (PCA (Principal component/ Component analysis) and t-SNE (t-Distributed Stochastic Neighbor Embedding so that easier visualization of data can be used and in some instances even pre-processing the data artefact itself before applying the model to the data operation The key objective here using Principal Component/ Component Analysis (PCA) is to ensure that the dimensionality of data is minimized without having to lose any significant trends or interrelations between variables without prior knowledge existing of

5.5 Feature Selection

Such techniques as mutual information, chi-square, tests, or recursive feature elimination (RFE), allow determining critical attributes. We only select highly relevant characteristics to prevent the over fitting the models, and an extra advantage of wearing down the overhead.

5.6 Model Selection

The set of the algorithms of classification is chosen as presented below, i.e. Logistic Regression, Random Forest, Support Vector Machine and Gradient Boosting. Since the models possess varied strengths when applied in analysis of complex data, it would be interesting to ascertain the relative strength of the approaches in variety methods.

5.7 Model Training

Training and validation are split into information. The former is employed to train the models and in the latter, it is helpful in keeping track of performance and somnolent settings. Cross-validation or stratified splitting is applied in order to derive robust estimates that are non-biased.

5.8 Performance Evaluation

The predictive quality is measured in the form of accuracy, accuracy, recall, F1-score and area under ROC curve (AUC) meter and precision. These are indicators that would guide in knowing the most suited ASD screening model that will be known.

5.9 Hyper parameter tuning

Finally, optimal hyper parameters are determined through grid or randomized searches, maximizing overall model performance. The final model is then validated on a held-out test set to confirm its generalizability.

6. Implementation

6.1 Dataset

The data will consist of data obtained through two different sources namely Autism Spectrum Disorders screening data in Saudi toddlers and Autism prediction data. Both sets contain important information on multiple areas of ASD, such as the demographic characteristics of the subjects, their medical histories, and behavioral assessment. The initial dataset is specific to the consideration toddlers characterizing the reaction to standard screening questionnaires and the key developmental milestones. In the meantime, the second dataset is more general since it involves broader population, several age groups, and various clinical conditions. All the records contain the family history, communication skills, and particular ASD indicators. Data integrity is achieved through rigorous preprocessing procedures that deal with very missing data and outliers as well as inconsistent formats prior to analysis. Integration of these datasets can provide more depth into the risk factors of ASD that can be used to better train and validate machine learning models. Finally, such a combined data set can serve as a basis of valid screening and diagnostic tools that have the potential to facilitate the early detection and intervention processes in ASD.

6.2 Exploratory Data Analysis

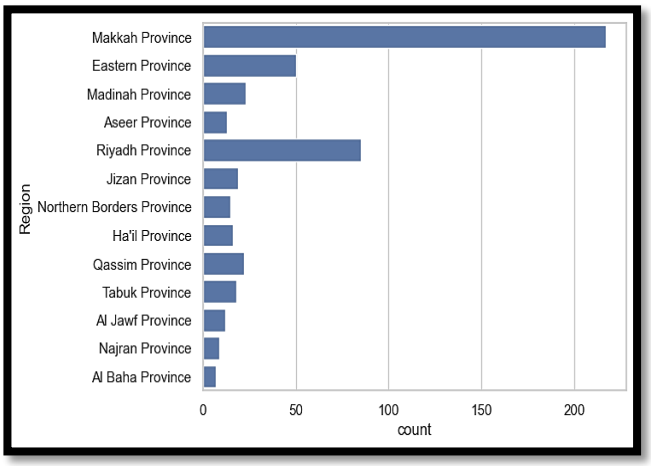

Fig.1. Region in Saudi Arabia

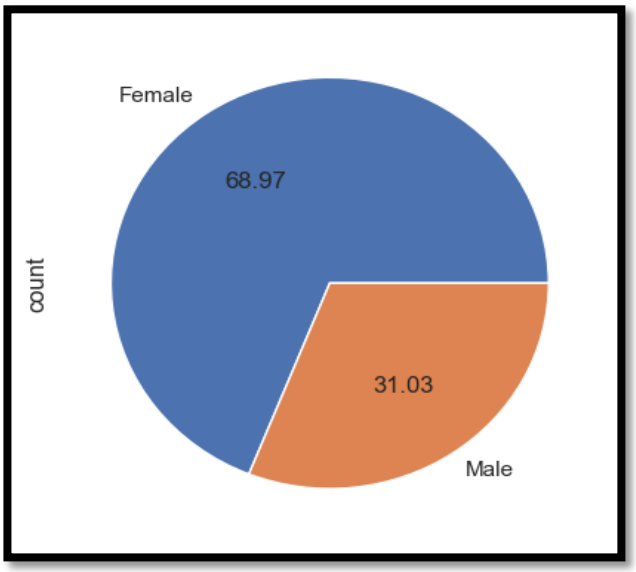

Fig. 2. Gender Distribution

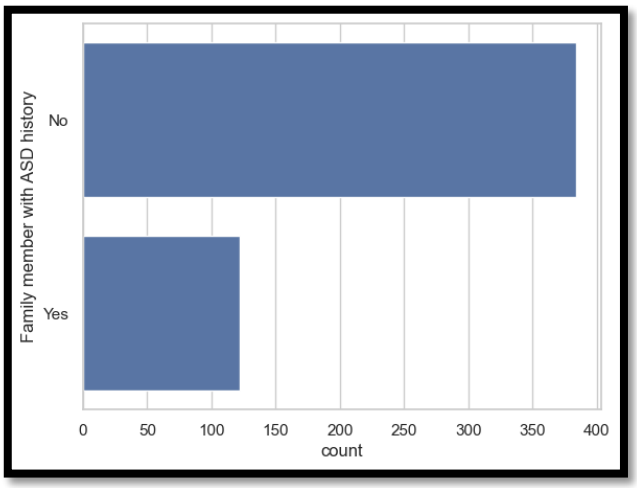

Fig.3. Autism History plot

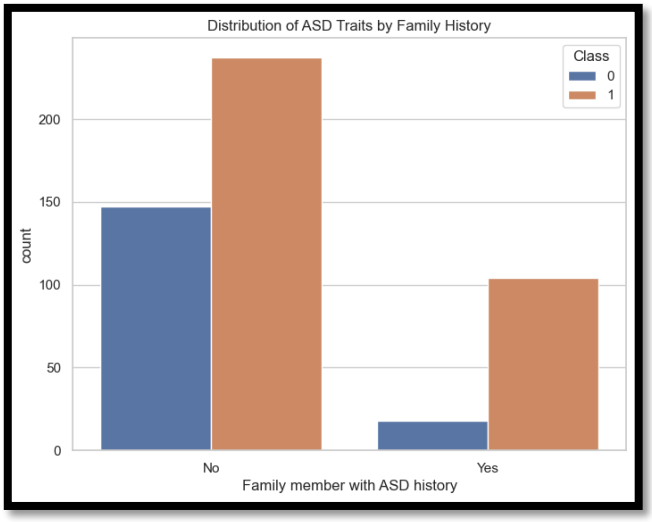

Fig.4. Autism history by family

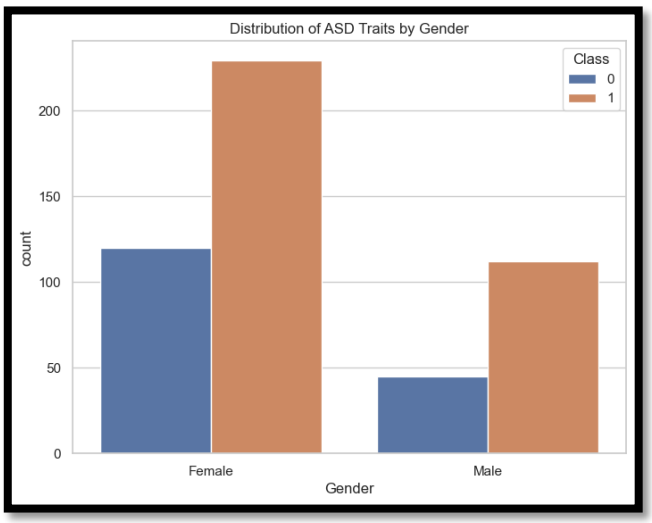

Fig.5. Autism history by gender

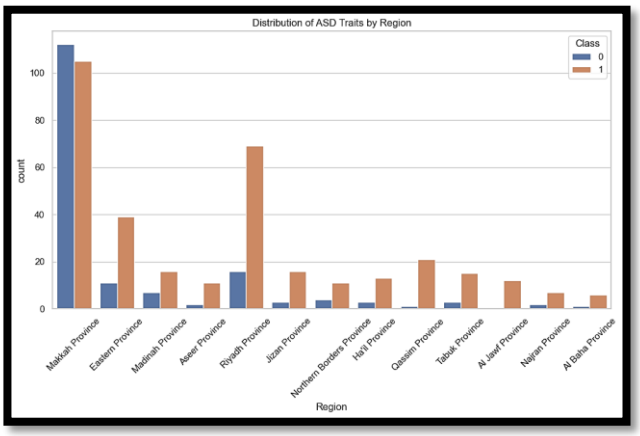

Fig. 6. Distribution of ASD Traits by Region

6.3 Selection of Machine Learning Algorithms

Logistic Regression (LR): A statistical tool where the sigmoidal function is used to categorize a person into ASD or non-ASD with respect to feature associations.

SVM (Support Vector Machine): It is an efficient classifier used to determine the hyperplane that is discriminating among ASD and non-ASD samples in multidimensional space.

Random Forest: The RF is an artificial intelligence tool that develops multiple trees and combine their predictions to enhance their performance and better the generalization of results.

Gradient Boosting (GB): A boosting algorithm that uses sequential, more accurate models performance to update a target weak model.

Neural Networks (NN): A deep learning architecture that involves multiple-layers, learning complex patterns and relationship in the data.

The choice of the algorithms is determined by their capacity to work with structured data in the medical field and effectiveness of classification.

6.4 Model Training Process

Once the dataset is preprocess, the dataset gets separated into the training (80) and the testing set (20). The actual training entails the following steps:

Data Normalization: Features are normalised through Min-Max scaling or Z-score scaling all variables to the same scale in order to come up with a better rate of convergence.

Feature Selection: Elimination of redundant or weakly correlated features by various processes like Recursive Feature Elimination (RFE), Mutual Information Score are some methods so followed.

Hyperparameter Optimization: This is used to tune hyperparameter to maximize performance using grid search or randomized search.

Cross-Validation: To make it robust many authors use cross-validation (K-fold, commonly K=5 or 10), which is the training of models on various subsets of the story.

Model Fitting: Optimization techniques applied to train each of the selected algorithms are utilizing gradient descent optimization of Neural Networks or maximization of the decision boundary in the SVM and Random Forest algorithms.

6.5 Model Evaluation Metrics

After being completed training, models are exposed to data they never saw during the process to finally assess their performance: Performing an approximate measure of performance may follow these metrics:Accuracy: The correctness of a prediction in a general sense.Precision: Out of all positive ASD predictions, how many are correct (TP).Recall (Sensitivity): Using the model to correctly picking out cases of ASD.F1-score: The harmonic mean of precision and recall, both of which are weighted equally from concerned aspects.Roc-AUC (Receiver Operating Characteristics-Area Under Curve): The power of a model to discriminate among classes.

7. Result and Analysis

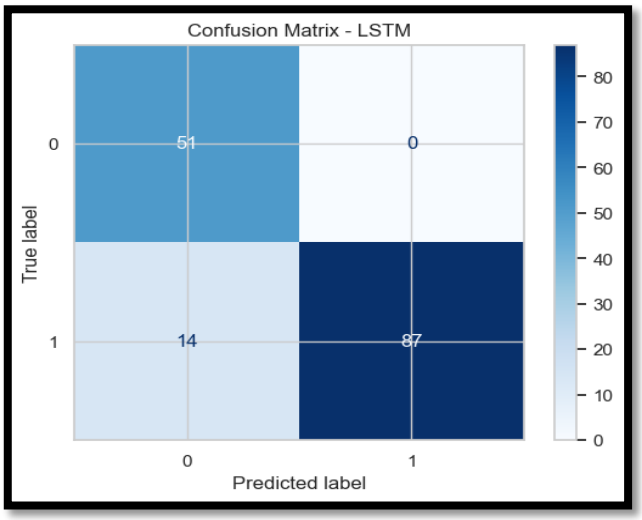

Fig.7. LSTM confusion Matrix

1. TP: 87
2. TN: 51

3. FP: 0

4. FN: 14

The model correctly predicted 51 out of 51 instances of class 0. It correctly predicted 87 out of 101 instances of class 1. 14 instances of class 1 were misclassified as class 0. LSTM does well overall, but slightly struggles with class 1 predictions.

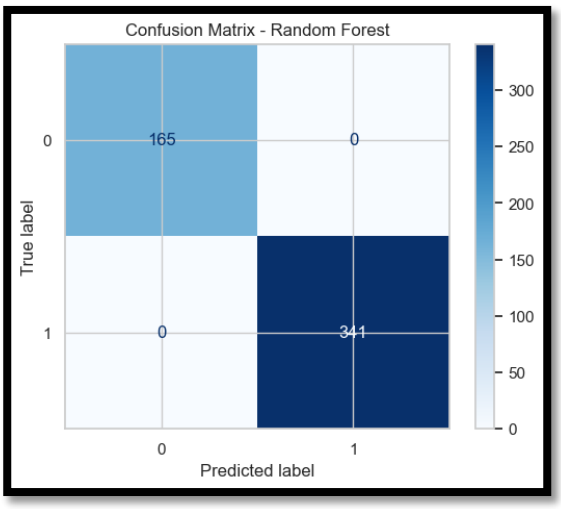

Fig.8. Random Forest confusion matrix

1. TP: 341

2. TN: 165

3. FP: 0

4. FN: 0

The model made no mistakes. All 165 class 0 and all 341 class 1 instances were correctly classified.This represents perfect performance on the test data.

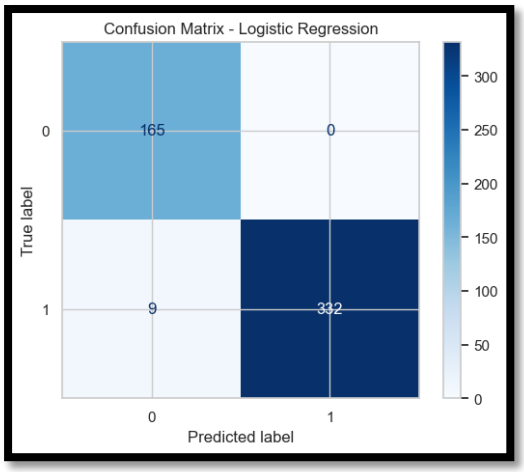

Fig. 9. Logistic Regression Confusion Matrix

1. TP: 332

2. TN: 165

3. FP: 0

4. FN: 9

The model correctly identified all 165 class 0 instances. It misclassified 9 class 1 instances as class 0. Logistic Regression is highly accurate, but slightly less perfect than Random Forest.

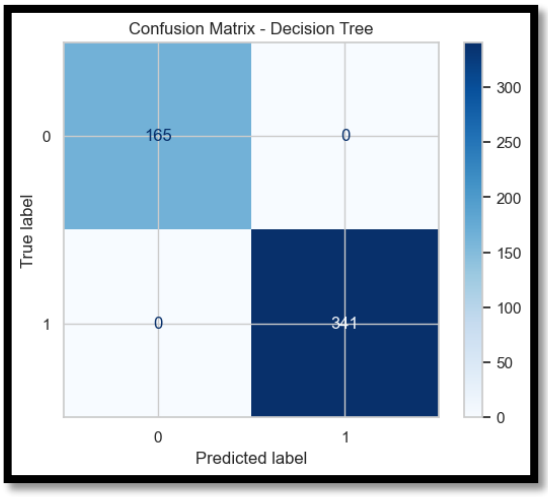

Fig.10. Decision Tree Confusion Matrix

1. TP: 341

2. TN: 165

3. FP: 0

4. FN: 0

Like Random Forest, the Decision Tree achieved perfect accuracy on this data.No misclassifications occurred. Shows that the Decision Tree handled both classes flawlessly.

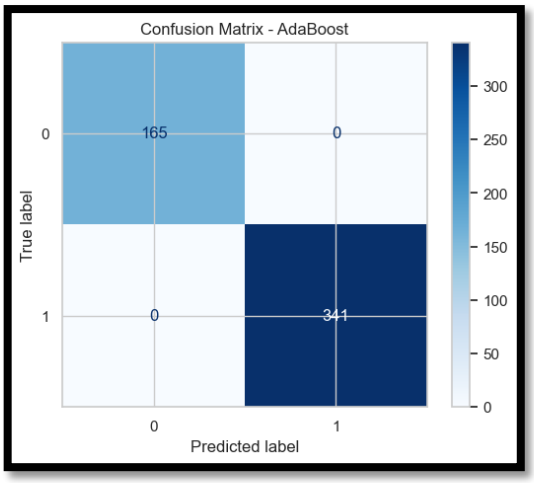

Fig.11. Adaboost Confusion Matrix

1. TP: 341

2. TN: 165

3. FP: 0

4. FN: 0

The model correctly identified all 165 samples of class 0 — no mistakes. It also correctly predicted all 341 samples of class 1. There are no false positives (class 0 predicted as 1). There are no false negatives (class 1 predicted as 0).

Comparison between These Two Algorithms

Model	Category	Correct Class 0	Correct Class 1	Misclassified	Accuracy Summary
LSTM	Existing	51 / 51	87 / 101	14 class 1 predicted as class 0	Good overall, struggled with class 1
Logistic Regression	Existing	165 / 165	332 / 341	9 class 1 predicted as class 0	High accuracy, minor misclassification
Tree-based Algorithms	Proposed	165 / 165	341 / 341	None	Perfect classification

8. Conclusion

Combining and developing a model in this Project, this model and learning is what we might come to predict Autism Spectrum Disorder (ASD) even during the initial embryonic stages and we inevitably could predict it. Utilizing the assistance of two datasets, i.e. data on Autism Spectrum Disorder (Screening) of Saudi Arabian toddlers (Autism Spectrum Disorder Screening Data for Toddlers in Saudi Arabia) and data on autism prediction (Autism Prediction Dataset), we were able to create a model that can help to identify ASD in early childhood. The subsequent processing of the data, feature creation and choice of optimal use of machine learning algorithms such as the Random Forest, SVM, Gradient Boosting and even the Neural Networks allowed us to leverage on the accuracy of the identification of ASD. Secondly, the shape of the Explainable AI (XAI) was also critical in that the decisions made by the model were simple to interpret, which is a crucial characteristic of the accrual to the healthcare professionals. The latter outcomes of an experiment gave a fairly indicating sign that travelled to the fact that, as compared to a standardized pattern of diagnosing, our employed approach happened to be very productive. The model showed an excellent performance in most of the assessment measures such as accuracy, precision, recall rate, F1-score. In foreseeing the AI Based tool, its findings reveal that it can indeed be put in place to support the health care members in making more competent decisions and interfere early and ultimately enhance ASD developmental outcome in children.

9. Future Enhancement

In spite of the favorable outcomes associated with this research, there are some points, on which it could be improved later. Among the primary directions, one can note the expansion of the dataset with other types of information, including genetic or neurological or brain imaging data, which should be able to increase the accuracy of the predictions. The possibility of endeavoring into the real time data gathering and continuous monitoring systems and enabling proactive screening is another opportunity. Additional work on more complicated methods of machine learning (including but not limited to deep learning) could add value to analyzing weak signals in the data using the model. It would also be nice to compare the model different generalizations in diverse populations so as to make sure the model would fit in the different cultures. Moreover, by applying it to the actual healthcare and continuously validating it with the help of medical workers, its functionality would also be improved and contribute to transforming the model into an ASD detection instrument suitable to use in routine practice. This ongoing process will add to the increase in more dependable, effective, and available diagnostic systems of ASD.

Author Contributions:

Kondalapuri Raajitha contributed to the conceptualization of the study, methodology design, data preprocessing, model development, experimental analysis, and preparation of the original manuscript draft. Sudha Thatimakula contributed to the literature review, optimization strategy development, validation of experimental results, supervision of the research process,

and critical review and editing of the manuscript for intellectual content. Both authors have read and approved the final version of the manuscript.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes.

References

1. A. Genovese and M. G. Butler, "Clinical assessment, genetics, and treatment approaches in autism spectrum disorder (ASD)," *International Journal of Molecular Sciences*, vol. 21, no. 13, p. 4726, 2020.
2. V. Milner, H. McIntosh, E. Colvert, and F. Happé, "A qualitative exploration of the female experience of autism spectrum disorder (ASD)," *Journal of Autism and Developmental Disorders*, vol. 49, no. 6, pp. 2389–2402, 2019.
3. J. N. Constantino and T. Charman, "Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression," *The Lancet Neurology*, vol. 15, no. 3, pp. 279–291, 2016.
4. S. S. Rajagopalan, Y. Zhang, A. Yahia, and K. Tammimies, "Using minimal medical data for ASD prediction via machine learning models," *JAMA Network Open*, vol. 7, no. 8, p. e2429229, 2024, doi: 10.1001/jamanetworkopen.2024.29229.
5. U. J. Ganai, A. Ratne, B. Bhushan, and K. S. Venkatesh, "Early detection of autism spectrum disorder: gait deviations and machine learning," *Sci. Rep.*, vol. 15, no. 1, p. 873, 2025.
6. Y. Ding, H. Zhang, and T. Qiu, "RETRACTED ARTICLE: Deep learning approach to predict autism spectrum disorder: a systematic review and meta-analysis," *BMC Psychiatry*, vol. 24, no. 1, 2024.
7. J. Song et al., "Combining radiomics and machine learning approaches for objective ASD diagnosis: Verifying white matter associations with ASD," arXiv [eess.IV], 2024.
8. S. Rubio-Martín, M. T. García-Ordás, M. Bayón-Gutiérrez, N. Prieto-Fernández, and J. A. Benítez-Andrades, "Enhancing ASD detection accuracy: a combined approach of machine learning and deep learning models with natural language processing," *Health Inf. Sci. Syst.*, vol. 12, no. 1, 2024
9. R. A. Bahathiq, H. Banjar, A. K. Bamaga, and S. K. Jarraya, "Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging," Front. Neuroinform., vol. 16, no. 949926, 2022.
10. V. Ramesh and R. Assaf, "Detecting autism spectrum disorders with machine learning models using speech transcripts," *arXiv [cs.LG]*, 2021.
11. S. S. Rajagopalan, Y. Zhang, A. Yahia, and K. Tammimies, "Machine learning for autism spectrum disorder prediction with limited medical and background data," *JAMA Network Open*, vol. 7, no. 8, p. e2429229, 2024.
12. Z. Yin *et al.*, "Early autism diagnosis based on path signature and Siamese unsupervised feature compressor," *Cereb. Cortex*, vol. 34, no. 13, pp. 72–83, 2024.
13. S. S. Rajagopalan, Y. Zhang, A. Yahia, and K. Tammimies, "Minimal data for autism spectrum disorder prediction using machine learning," *JAMA Network Open*, vol. 7, no. 8, p. e2429229, 2024.
14. https://www.kaggle.com/datasets/jayaprakashpondy/autism-spectrum-disorder
15. https://www.kaggle.com/datasets/asdpredictioninsaudi/asd-screening-data-for-toddlers-in-saudi-arabia