Machine Learning Approaches for Predicting Autism in Children: A Comparison of AdaBoost and Other Algorithms

G. Divya*, V. Maniraj

Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur (Dt)

*Corresponding author's email address: gdivya19.mca@gmail.com

doi: https://doi.org/10.21467/proceedings.173.16

ABSTRACT

This study presents a Autism Spectrum Disorder (ASD) in children is a neurodevelopmental disorder characterized by challenges in social interactions, communication, and behavior. Early identification and diagnosis of ASD, especially between the ages of 20 and 60 months, are essential for effective treatment. Without early detection, intervention becomes significantly more difficult. While a variety of machine learning (ML) techniques have been employed to predict ASD, the accuracy of predictions for younger age groups is still limited. This study investigates the use of three machine learning algorithms—Support Vector Machine (SVM), Random Forest, and AdaBoost—to predict and identify autism in children. The AdaBoost classifier, which integrates multiple weak learners to form a more robust classifier, is proposed as the primary approach. To assess the performance of these algorithms, key metrics such as accuracy, precision, F-score, and the confusion matrix are computed. The algorithm with the highest accuracy is then utilized to forecast autism in children.

Keywords: Machine Learning, Autism, SVM, Random Forest, AdaBoost

1 Introduction

Autism is a developmental disorder that affects social interactions, communication, and behaviour. It primarily impacts children's responses to cognitive functions. The condition is marked by impairments in both verbal and nonverbal communication, as well as the repetition of stereotyped behaviours. Unfortunately, autism often develops rapidly, and although it can be diagnosed at any age, its symptoms typically manifest within the first two years of life [3,4]. Children with autism face numerous challenges, including poor response to stimuli, learning disabilities, difficulty focusing, sensory sensitivities, anxiety, depression, and motor coordination issues. The impact of autism spectrum disorder (ASD) varies significantly among children, with differences in family history, co morbidities, and associated costs. Research suggests that autism may result from a combination of genetic, environmental, and non-genetic factors in a child's life. Early signs of autism can often be identified when children fail to respond to their parents, peers, or social interactions [5]. To address the challenges faced by children with autism, we propose using machine learning techniques and algorithms for effective diagnosis and prediction of autism. Machine learning methods are valuable tools for extracting meaningful insights from long-term stored data [6-8]. These algorithms help uncover hidden patterns within large datasets and facilitate the extraction of relevant information for practical applications. By implementing these techniques, we can process and analyze data to predict the presence of autism and inform potential treatment strategies.

2 Background Information

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that enables systems to learn from data and improve automatically without human intervention. By providing training data, the system gains experience, allowing it to make predictions or decisions based on patterns and inferences, rather than relying

Proceedings DOI: 10.21467/proceedings.173; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-970666-x-x

on predefined rules. With sufficient experience, the system can predict outcomes for new inputs without any human assistance. ML is typically classified into three primary types of learning: supervised learning, unsupervised learning, and reinforcement learning [1,2]]

Supervised learning involves using patterns and parameters learned from past data to process new input instances, based on labeled data. Many machine learning algorithms can be used with supervised learning, including SVM, Random Forest, and Naïve Bayes, among others. These algorithms create a predictive function by analyzing the training data to forecast outcomes. Labeled data for the instances can be time-consuming to obtain, but it provides targets for input instances through multiple rounds of training. Ultimately, the algorithm compares its predictions with the expected results and evaluates metrics such as accuracy, sensitivity, specificity, and error rates. These errors are then adjusted within the model to minimize them and improve accuracy.

2.1 Dataset

For predicting autism in children, our study utilizes the Autism Screening Data for Children (Toddler Dataset). This dataset includes 1,054 records of children aged 12 to 36 months. Each record contains 15 features, comprising both binary and string values. Feature engineering is applied to transform the string values into binary format, making them suitable for training and classification purposes. The dataset can be utilized for text classification tasks and can be processed using algorithms designed for text-based classification.

3 Proposed Work and Results Analysis

In our proposed approach, we employ three machine learning algorithms—SVM, AdaBoost, and Random Forest—to compare their performance and identify the algorithm that delivers the highest accuracy for predicting outcomes based on any given input. The proposed architecture is illustrated in Fig. 1 below.

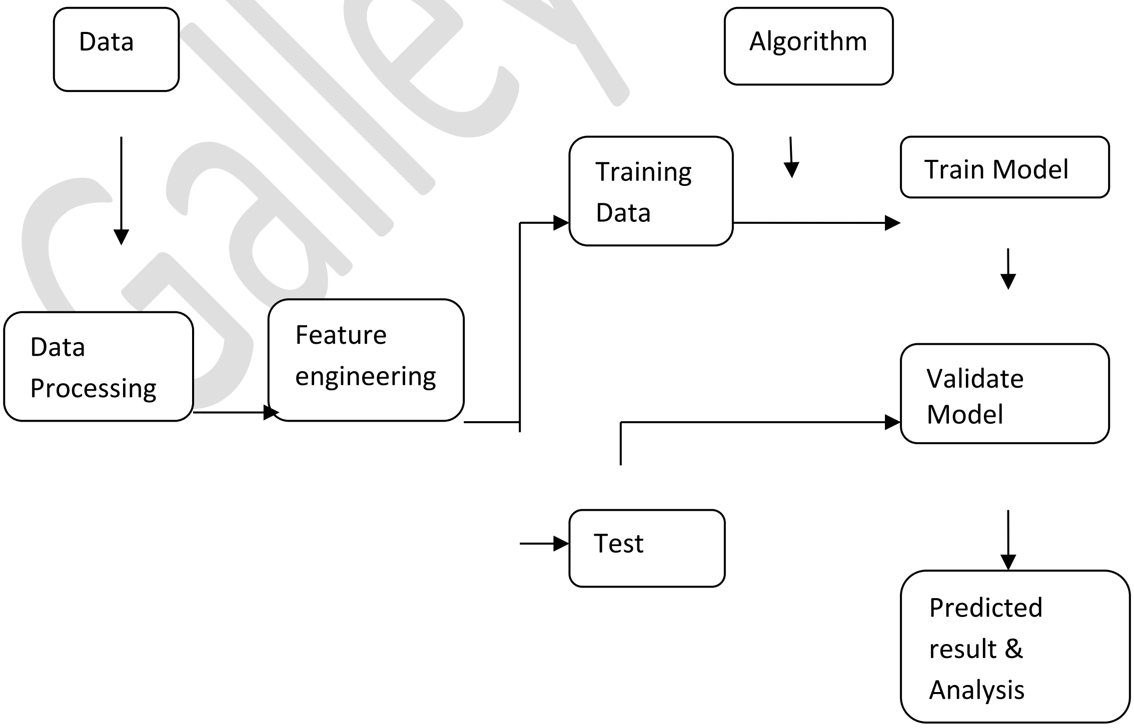

Fig1: *Proposed architecture*

The basic criterion in comparing the performance of the classifiers is to measure the effectiveness of the algorithms.

A. Precision

Precision gives the output quality of the model by evaluating the below mentioned formula.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

Precision can be calculated by dividing the true positive to the summation of true positive and false positive values. It is a measure of result relevancy.

B. Recall

Recall is also another metric to find the output quality to find how many true relevant results are obtained. Recall is sensitivity.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

In mathematical form, the true positive values are divided by the summation of true positive and false negative values of the instances, both of which are correctly classified.

C. F1 score

F1score is the weighted average of recall and precision. It gives the single score that balances of precision and recall.

$$\text{F1 score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

D. Accuracy

Accuracy is the overall classification validation with overall classification ratio

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

The confusion matrix is the prescribed general tool to measure the classification performance. It is measured against the true cases and the predicted cases with positive and negative outcomes.

Proceedings DOI: 10.21467/proceedings.173
ISBN: 978-81-970666-x-x

Series: AIJR Proceedings
ISSN: 2582-3922

		True Class	
		Positive	Negative
Predicted class	Positive	True positive count(TP)	False positive count (FP)
	Negative	False Negative Count (FN)	True Negative Count(TN)

Fig2: *Confusion matrix*

In **Fig. 2**, the actual cases are compared with the predicted cases, resulting in four possible outcomes. **True Positive** refers to children who are correctly identified as autistic by the classifier. **False Positive** denotes children who are not autistic but are incorrectly classified as autistic. **False Negative** refers to children who are actually autistic but are misclassified as non-autistic. **True Negative** indicates children who are correctly identified as non-autistic by the classifier. The confusion matrix is computed for three different algorithms, with the aim of accurately detecting autistic children through the identification of true positives.

Table 1: *Performance measures obtained using confusion matrix for three algorithms*

Algorithms	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Random Forest	96.20	0.97	0.95	0.96
SVM	96.68	0.97	0.96	0.96
AdaBoost	100	1.00	1.00	1.00

Table 1 presents the performance metrics—accuracy, precision, recall, and F1 score—for the Random Forest (RF), Support Vector Machine (SVM), and AdaBoost algorithms. According to the table, the accuracy of the Random Forest classifier is 96.20%, while SVM achieves an accuracy of 96.68%. In contrast, the AdaBoost classifier achieves a perfect accuracy of 100%, making it the top performer compared to both SVM and RF. When considering precision, recall, and F1 score, the results show that the recall values for both Random Forest and SVM are 0.95 and 0.96, respectively, which are lower than the AdaBoost algorithm's recall score of 1.00. The F1 score for both RF and SVM is 0.96, but AdaBoost reaches a perfect F1 score of 1.00. Based on accuracy, recall, and F1 score, it is evident that AdaBoost outperforms both Random Forest and SVM classifiers in predicting autism from children's data.

4 Conclusion

In recent years, boosting algorithms have gained significant popularity in the fields of Machine Learning and Data Science. These algorithms are often employed in accuracy-driven competitions to achieve superior performance. The experimental results demonstrate that the proposed algorithm delivers excellent accuracy and performs better overall, making it effective for predicting autism traits in children. The AdaBoost algorithm is particularly efficient when handling large datasets and selecting high-dimensional features. Moreover, AdaBoost and other boosting methods are less prone to issues like overfitting. In conclusion, it is evident that the AdaBoost classifier achieves a perfect accuracy of 100%, outperforming the SVM and Random Forest classifiers, which achieved 96%. For any given set of input data, the presence or absence of autism can be accurately predicted by the trained model, with AdaBoost being the top-performing classifier. In future work, the performance of these algorithms could be evaluated on larger datasets, and other boosting techniques could be explored in subsequent studies.

5 Declarations

5.1 Competing Interests

The debate here is whether simpler models like AdaBoost offer more interpretability and faster training times, or whether more complex models like deep neural networks provide superior accuracy at the cost of explainability.

5.2 Acknowledgements

Pioneering Work on Autism Diagnosis Using ML**:** Many researchers have paved the way for applying machine learning (ML) algorithms, including AdaBoost, in the field of autism diagnosis. Their efforts in exploring various data types (e.g., behavioral, genetic, neuroimaging) and testing different ML models have laid the foundation for further studies. Acknowledging the work of these pioneers is crucial, as their contributions directly influence current methodologies and algorithms.

5.3 Study Limitations

The limitations of studies comparing AdaBoost and other machine learning algorithms for predicting autism in children stem from factors such as limited and imbalanced data, generalization issues, ethical concerns, and the complexity of implementing these models in real-world clinical environments. These limitations highlight the need for more robust, diverse, and longitudinal datasets, as well as continued efforts to improve model interpretability and fairness. Addressing these challenges will be crucial in making machine learning models for autism prediction more reliable, ethical, and applicable in clinical practice.

5.4 Funding source

Financial Support: Acknowledging the support provided by funding agencies is essential, as many of these studies and algorithmic advancements rely on grants and funding. Organizations like the National Institutes of Health (NIH), National Science Foundation (NSF), and private foundations contribute to advancing autism research by providing the necessary financial resources for data collection, algorithm development, and clinical trials.

Philanthropic Contributions: Private organizations and philanthropic efforts have also played a role in advancing research into autism prediction, particularly in funding innovative approaches like machine learning.

5.5 Warning for Hazard

While machine learning algorithms, including AdaBoost, offer exciting potential for predicting autism in children, there are significant hazards and risks associated with their use. These include the potential for misdiagnosis, data bias, overfitting, privacy concerns, ethical implications, and challenges related to the implementation of these systems in real-world clinical settings. To mitigate these risks, machine learning models should always be used as complementary tools to clinical judgment, and proper safeguards must be put in place to ensure their responsible, ethical, and effective use in healthcare.

References

[1] Fatiha Nur Büyükoflaz, Ali Öztürk, "Early Autism Diagnosis of Children with Machine Learning Algorithms," Bilgisayar Mühendisli˘gi Bölümü, KTO Karatay Üniversitesi Konya, Türkiye, 2018 IEEE, pp. 1-4, doi: 10.1109/SIU.2018.8404223.

[2] Alanoud Bin Dris, Abdulmalik Alsalman, Areej Al-Wabil, Mohammed Aldosari, "Intelligent Gaze-Based Screening System for Autism," 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2019, pp. 1-5, doi: 10.1109/CAIS.2019.8769452.

[3] Kazi Omar, Nabila Khan, Prodipta Mondal, Md. Rezaul, Md Nazrul Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder," International Conference on ECCE, February 2019, pp. 1-6, doi: 10.1109/ECACE.2019.8679454.

[4] Khondaker Abdullah Al Mamun et al, "Smart Autism –A mobile, Interactive and integrated framework for screening and confirmation of autism," Conference IEEE engineering in medicine and biology society, Aug. 2016, pp. 5989-5992. doi: 10.1109/EMBC.2016.7592093. PMID: 28269617.

[5] Osman Altay, Ulas, "Prediction of the Autism Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and K Nearest Neighbor in Children," Firat University, Turkey 2018, pp. 1-4, doi: 10.1109/ISDFS.2018.8355354.

[6] Sushama Rani Dutta, Sujoy Datta, Monideepa Roy, "Using Cogency and Machine Learning for autism detection from a preliminary symptom," KIIT University, Odisha, India, 2019 IEEE, pp. 331-336, doi: 10.1109/CONFLUENCE.2019.8776993.

[7] 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, 2012 Feb, vol. 51, no. 2, pp. 202-212, doi: 10.1016/j.jaac.2011.11.003.

[8] Armin Lawi, Firman Aziz, Comparison of Classification Algorithms of the ASD Diagnosis. The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT) 2018, pp. 218-222, doi: 10.1109/EIConCIT.2018.8878593.