Deep Analysis of Autism Spectrum Disorder Detection Techniques

DR.D. ANUSHA[1], B. BABY SRAVANTHI[2], CH. TEJA SRI[3], D. SWAPNA[4], K. PRUDHVI RAJ[5]
[1, 2, 3, 4, 5] *Department of CSE (AI&ML), SRK Institute of Technology, Vijayawada, A.P., India.*

Abstract: Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition affecting communication, behavior, and social interactions. Early detection and intervention are critical for improving outcomes, and advancements in artificial intelligence (AI) and machine learning (ML) present promising opportunities for early diagnosis. Various ML algorithms, including XGBClassifier, Random Forest, Decision Trees, AdaBoostClassifier, KMeans, Extra Tree Classifier, and Artificial Neural Networks (ANN), have been explored for ASD prediction. These algorithms can analyze large datasets, detect patterns, and make predictions with high accuracy. For instance, XGBClassifier offers efficient handling of complex data, while Random Forest mitigates overfitting. Decision Trees provide interpretability, and AdaBoostClassifier combines multiple weak classifiers for stronger predictive power. KMeans helps identify subgroups within the ASD population, while ANN models capture intricate data relationships. However, challenges such as data imbalance, model complexity, and computational requirements need to be addressed. Researchers must balance accuracy, efficiency, and interpretability when selecting the best model. With continuous refinement and validation using diverse datasets, ML algorithms can revolutionize early ASD detection, enabling more accurate, timely, and cost effective interventions that improve the lives of individuals with ASD.

Keywords: Machine Learning, XGBClassifier, Random Forest, Decision Trees, Extra Tree Classifier, AdaBoostClassifier, KMeans, and Artificial Neural Networks (ANN).

INTRODUCTION

Autism Spectrum Disorder (ASD) is indeed a complex neurodevelopmental condition that affects the brain's functioning and can present itself at any age, but it is typically diagnosed in childhood, with higher chances of identification around the ages of 2 to 3 years. ASD is believed to result from a combination of genetic and environmental factors, and it is crucial to understand that it is not an illness or disease but a neurological condition that affects a child's ability to concentrate, think, learn, and solve problems. Individuals with ASD often encounter difficulties in expressing themselves through facial expressions or gestures, which can impact their social interactions. In recent years, the prevalence of autism has been on the rise globally, including in countries like India. The Autism Centre for Excellence (ACE) reported that approximately 1 in 68 children is affected by ASD. Given the increasing incidence, early diagnosis becomes paramount to provide timely interventions and support for affected children. Detecting ASD at an early age can be challenging, but advancements in machine learning techniques have shown promise in aiding this process. Various researchers have been exploring and developing machine learning models to predict autism traits in children. These models typically leverage data from various sources, including behavioural observations, medical histories, genetic data, and environmental factors, to identify potential signs of ASD. The objective of this paper is to conduct an in-depth analysis of the different machine learning techniques used for predicting autism traits. Some common approaches include decision trees, support vector machines, random forests, neural networks, and Bayesian networks, among others. Each technique has its advantages and limitations, and researchers have been working to compare their efficiency and accuracy in detecting ASD. To further enhance the accuracy of these models, ongoing research focuses on increasing the size and diversity of the datasets used for training. Collaborative efforts between experts in machine learning, neuroscience, and autism research are essential to refine these techniques continually. ASD is a significant challenge for affected individuals and their families. Early diagnosis and intervention can greatly improve the quality of life for children with ASD. Machine learning techniques hold promise in assisting with early detection, and ongoing research is vital to enhance their accuracy and applicability. By combining the expertise of different disciplines, we can strive to provide better support and care for those with autism.

LITERATURE SURVEY

[1]. Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi, Md Nazrul Islam, "A Machine Learning Approach to Predict AutismSpectrum Disorder", "International Conference on Electrical, Computer and Communication Engineering (ECCE)", 2019 In present day Autism Spectrum Disorder (ASD) is gaining its momentum faster than ever. Detecting autism traits through screening tests is very expensive and time consuming. With the advancement of artificial intelligence and machine learning (ML), autism can be predicted at quite early stage. Though number of studies have been carried out using different techniques, these studies didn't provide any definitive conclusion about predicting autism traits in terms of different age groups. Therefore this paper aims to propose an effective prediction model based on ML technique and to develop a mobile application for predicting ASD for people of any age. As outcomes of this research, an autism prediction model was developed by merging Random Forest-CART (Classification and Regression Trees) and Random Forest ID3 (Iterative Dichotomiser 3) and also a mobile application was developed based on the proposed prediction model. The proposed model was evaluated with AQ-10 dataset and 250 real dataset collected from people with and without autistic traits. The evaluation results showed that the proposed prediction model provide better results in terms of accuracy, specificity, sensitivity, precision and false positive rate (FPR) for both kinds of datasets.

[2]. Md. Shahriare Satu , Farha Farida Sathi, Md. Sadrul Arifen, Md. Hanif Ali and Mohammad Ali Moni, "Early Detection of Autism by Extracting Features:A Case Study in Bangladesh", "International Conference on Robotics,Electrical and Signal Processing Techniques" , 2019 Autism Spectrum Disorder (ASD) is a neurobehavioral disorder that begins at childhood and exists this whole life. The objective of this work is that to explore significant features of normal and autism of divisional regions in Bangladesh. We collected individual samples of various children from their parents between 16 to 30 months of different residents using Autism Barta apps by web and fieldwork at Savar, Bangladesh. Then, we preprocessed our data and categorized frequent features based on their individual regions. Different tree based techniques such as J48, Logistic Model Tree, Random Forest, Reduced Error Pruned Tree, and Decision Stump were analyzed to find out the best classifier of them. From these classifiers, J48 showed the best outcomes than other classifiers. We extracted 9 rules and associated conditions from J48 decision tree and gathered frequent instances from our data for extracted rules. Finally, 8 within 23 features were required to classify normal and autism of individual regions in Bangladesh. Besides, 4 rules (10 Conditions) for normal and 5 (12 Conditions) rules for autism out of 9 (16 Conditions) rules were extracted from decision tree. This outcomes assist us to find out significant features of autism in Bangladesh. We expect that our work will be helpful things to improve their condition that leads them to a normal life.

[3]. Kayleigh K. Hyde, Marlena N. Novack, Nicholas LaHaye, Chelsea Parlett-Pelleriti ,Raymond Anden, Dennis R. Dixon. Erik Linstead. "Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review", "Review Journal of Autism and Developmental Disorders", 2019., https://doi.org/10.1007/s40489-019- 00158-x Autism spectrum disorder (ASD) research has yet to leverage "big data" on the same scale as other fields; however, advancements in easy, affordable data collection and analysis may soon make this a reality. Indeed, there has been a notable increase in research literature evaluating the effectiveness of machine learning for diagnosing ASD, exploring its genetic underpinnings, and designing effective interventions. This paper provides a comprehensive review of 45 papers utilizing supervised machine learning in ASD, including algorithms for classification and text analysis. The goal of the paper is to identify and describe supervised machine learning trends in ASD literature as well as inform and guide researchers interested in expanding the body of clinically, computationally, and statistically sound approaches for mining ASD data.

[4]. Fadi Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward", "informatics for Health and Social Care",2018 Autistic Spectrum Disorder (ASD) is a mental disorder that retards acquisition of linguistic, communication, cognitive, and social skills and abilities. Despite being diagnosed with ASD, some individuals exhibit outstanding scholastic, non-academic, and artistic capabilities, in such cases posing a challenging task for scientists to provide

answers. In the last few years, ASD has been investigated by social and computational intelligence scientists utilizing advanced technologies such as machine learning to improve diagnostic timing, precision, and quality. Machine learning is a multidisciplinary research topic that employs intelligent techniques to discover useful concealed patterns, which are utilized in prediction to improve decision making. Machine learning techniques such as support vector machines, decision trees, logistic regressions, and others, have been applied to datasets related to autism in order to construct predictive models. These models claim to enhance the ability of clinicians to provide robust diagnoses and prognoses of ASD. However, studies concerning the use of machine learning in ASD diagnosis and treatment suffer from conceptual, implementation, and data issues such as the way diagnostic codes are used, the type of feature selection employed, the evaluation measures chosen, and class imbalances in data among others. A more serious claim in recent studies is the development of a new method for ASD diagnoses based on machine learning.

EXISTING SYSTEM

Most of the current existing systems uses older machine learning algorithms like Artificial Neural Networks (ANN) and KMeans etc. which are efficient in processing but fails to generate good accuracy. There are some systems which uses Deep learning for autism detection but they require high processing power are a bit complex to use. Disadvantages: Complexity and Resource Intensive Lack of Generalizability Interpretability Challenges Limited Data Availability Ethical and Privacy Concerns

PROPOSED SYSTEM

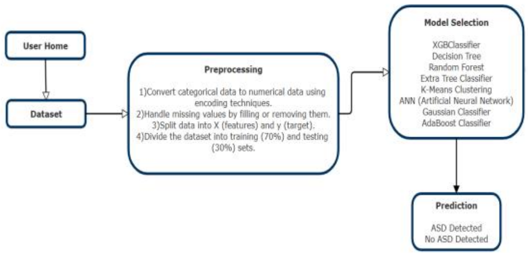

Fig:[1]

Proposed system algorithms:

Decision tree Choose the best feature using Gini Impurity or Entropy, then split the dataset into subsets based on feature values. Repeat Step 1 for each subset until a stopping condition is met (pure nodes, max depth, or min samples). Traverse the tree using input values until reaching a leaf node, which provides the final class label or value. The leaf node's value is returned as the final prediction for classification or regression.

XGB Classifier Begin with a weak base model (typically a decision tree) and set initial predictions. Calculate the difference between actual and predicted values (residuals) to identify errors. Each new tree learns from the residual errors of the previous trees using gradient descent. The trees are added iteratively to correct previous mistakes. Each tree contributes to the final prediction by adjusting weights using boosting. The weighted sum of all trees' predictions determines the final classification result.

Random Forest Randomly select multiple subsets of the dataset with replacement. Each tree is trained on a different subset. Only a random subset of features is considered for splitting at each node. Classification: Majority voting (most frequent class). Regression: Averaging (mean of predictions). Aggregated result from all trees gives the final prediction.

Adaboost Classifier Assign equal weights to all training samples. A simple classifier (typically a shallow decision tree) is trained. Misclassified samples get higher weights (more focus). Correctly classified samples get lower weights (less focus). A new weak model is trained, focusing more on previously misclassified samples. Combine all weak learners' outputs using a weighted sum (strong classifier).

Extra Tree Classifier Unlike Random Forest, Extra Trees can use the entire dataset or random subsets without replacement. Each tree is trained on a different subset. Instead of selecting the best split, splits are chosen randomly for more diversity. Classification: Majority voting (most frequent class). Regression: Averaging (mean of predictions). Aggregated results from all trees give the final prediction.

K-Means Clustering Randomly select K cluster centroids from the dataset. Each data point is assigned to the nearest centroid based on Euclidean distance. Recalculate each centroid as the mean of all points assigned to it. Steps 2 & 3 are repeated until centroids no longer change or a maximum number of iterations is reached. The dataset is divided into K clusters, with each data point assigned to its closest centroid.

Guassian NB Compute the probability of each class based on the training data. For each feature, assume a normal (Gaussian) distribution and compute the probability of data belonging to each class using the mean and variance of the feature. Combine prior probabilities and likelihoods to calculate the posterior probability for each class. Assign the class with the highest posterior probability to the given input. The predicted class label is returned based on probability calculations.

Artificial Neural Networks Define the input layer, hidden layers, and output layer with weights initialized randomly. Pass input data through the network, computing activations at each layer using weights and activation functions (e.g., ReLU, Sigmoid, Softmax). Compare the predicted output with the actual target using a loss function (e.g., Cross Entropy for classification, MSE for regression). Calculate the gradient of the loss with respect to each weight using backpropagation and update the weights using an optimizer like SGD or Adam. Steps 2-4 are repeated for multiple epochs until the loss minimizes or a stopping criterion is met. The trained network makes predictions based on learned weights and activation functions.

Methodology:

Data Collection The dataset is sourced from reliable medical databases, behavioral assessments, and ASD screening records. It includes essential attributes such as age, gender, behavioral responses, and diagnostic indicators, ensuring a comprehensive analysis.

Data Preprocessing Preprocessing involves cleaning the dataset by removing inconsistencies and handling missing values through imputation techniques. Feature selection is performed using statistical methods to retain the most relevant attributes for accurate prediction.

Model Selection A combination of machine learning algorithms, including XGB Classifier, Random Forest, Decision Trees, Extra Tree Classifier, AdaBoost Classifier, ANN, KMeans, and GaussianNB, is used to enhance ASD detection. These models collectively improve accuracy, robustness, and efficiency.

Training and Testing The dataset is split into training and testing sets to evaluate model performance. The training phase enables the models to learn patterns, while testing ensures generalization and accuracy on unseen data.

Performance Evaluation Evaluation metrics such as Accuracy, Precision, Recall, and F1-Score are used to measure the effectiveness of the models. These metrics ensure reliable classification and validate the system's ability to detect ASD accurately.

Deployment and Prediction The trained model is integrated into a user-friendly web application, allowing users to upload datasets and input values for real-time ASD predictions. The system provides early diagnosis insights, aiding timely intervention and treatment planning.

Advantages:

Deep analysis techniques like machine learning and AI are revolutionizing the detection and treatment of Autism Spectrum Disorder (ASD). By processing vast amounts of data, these methods can identify subtle early signs of ASD, enabling earlier intervention and better long-term outcomes. AI also standardizes assessments, reducing diagnostic variability and providing more accurate, data-driven diagnoses. Personalized treatment plans are possible by tailoring interventions to an individual's specific needs, improving effectiveness. Additionally, these technologies integrate diverse data types, offering a more comprehensive understanding of ASD’s causes and potential treatments. Overall, deep analysis enhances research, uncovers novel biomarkers, and paves the way for innovative therapeutic approaches, improving the quality of life for individuals with ASD.

RESULT

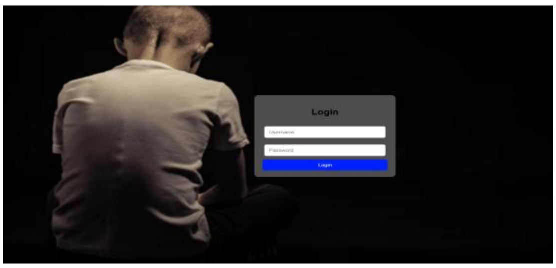

Fig:[2] Authentication Page

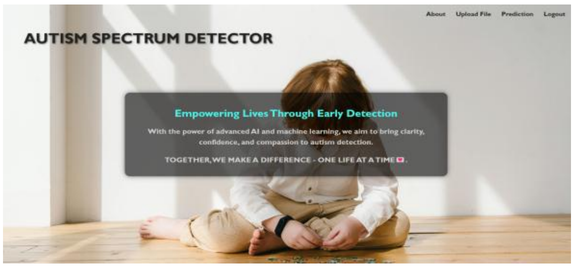

Fig:[3] Home page

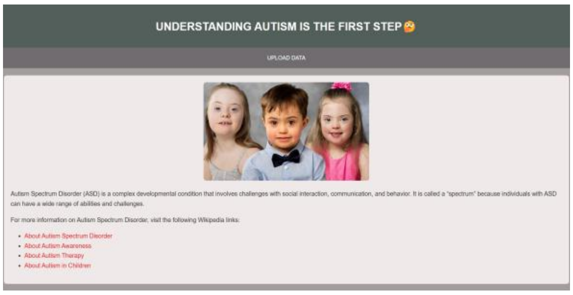

Fig:[4] About Page

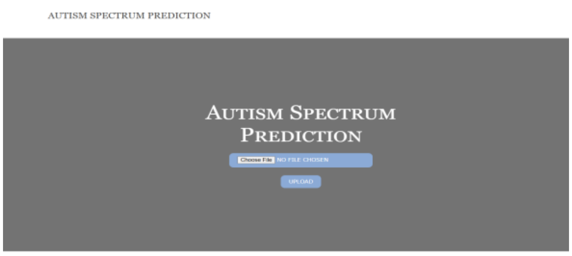

Fig:[5] Upload Dataset

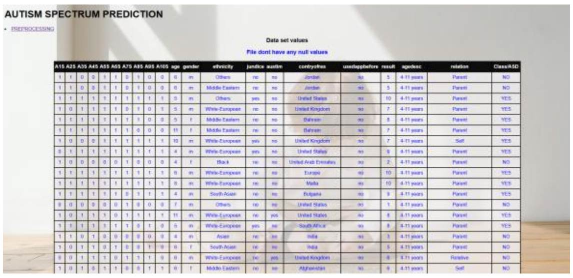

Fig:[6] View Dataset

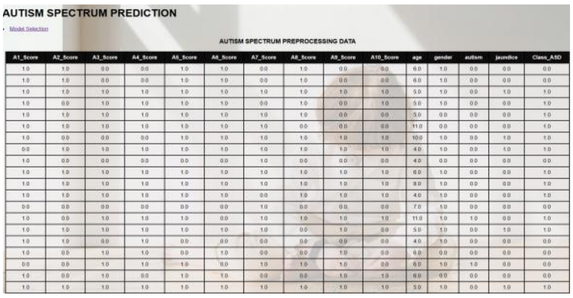

Fig:[6] Preprocessed Data

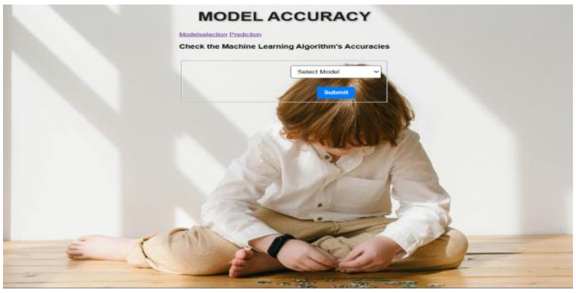

Fig:[7] Model Selection

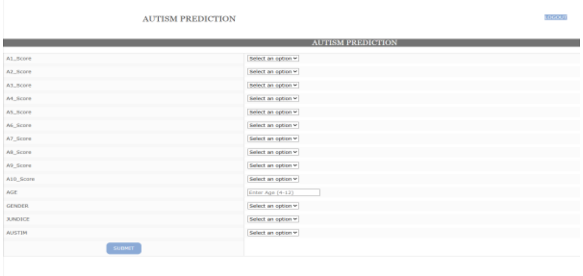

Fig:[7] Prediction

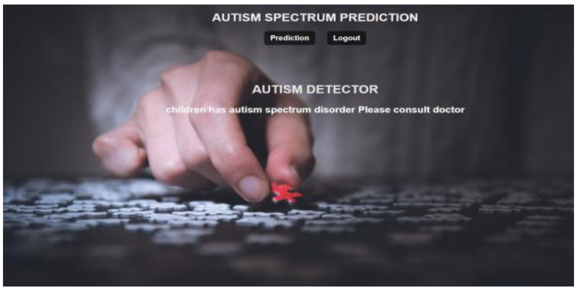

Fig:[8] Prediction Result

CONCLUSION

Autism is one of the vital issues which cannot be prevented but can be treated and it's a big challenge for any family to have a child with autistic disorder. Therefore, it's importance to diagnose it, as early as possible. One of the major issues in today's research is to have improved diagnostic tools to have faster, effective and accurate result. Machine learning techniques have shown favorable results. Machine learning with its leverage methods can be used to diagnose ASD. Use of machine Learning are to create algorithms that are vigorous and have contrive instruments. By using different ML algorithms researchers were able to build such model which shows improved results in terms of accuracy and precisions. In this analysis various classifiers have been used like XGBClassifier, Random Forest, Extra Tree Classifier's Decision Trees, AdaBoostClassifier, KMeans, and Artificial Neural Networks (ANN).

FUTURE SCOPE

Future enhancements for deep analysis of Autism Spectrum Disorder (ASD) detection techniques include integrating multimodal data such as genetic, neuroimaging, and behavioral markers to improve accuracy. Advanced machine learning models, like graph neural networks, could capture complex relationships within these diverse data types. Collaborative efforts between researchers and clinicians will refine interpretability, leading to more personalized and early ASD diagnosis. Moreover, ethical considerations regarding data privacy and informed consent must be addressed to ensure responsible and equitable deployment of these enhanced techniques.

REFERENCES

[1] U. Frith and F. Happé, “Autism spectrum disorder,” Current biology, vol. 15, no. 19, pp. R786–R790, 2005

[2] Dr. Archana Nayar, “http://www.csrmandate.org/autism centre-forexcellence-unlocking-unlimited-potentials-for autistic-children”, 2017.

[3] Paul Fergus, Basma Abdulaimma, Chris Carter, Sheena Round, “Interactive Mobile Technology for Children with Autism Spectrum Condition (ASC)”, “ IEEE 11th Consumer Communications and Networking Conference (CCNC)”, 10.1109/CCNC.2014.7111685, 2014.

[4] Wenbo Liu, Ming Li, and Li Yi, " Identifying Children with Autism Spectrum Disorder Based on Their Face Processing Abnormality: A Machine Learning Framework”, International Society for Autism Research, Wiley Periodicals,2016

[5] Kayleigh K. Hyde, Marlena N. Novack, Nicholas LaHaye, Chelsea Parlett-Pelleriti ,Raymond Anden, Dennis R. Dixon. Erik Linstead. “Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review”, “Review Journal of Autism and Developmental Disorders”, 2019., https://doi.org/10.1007/s40489-019- 00158-x

[6] Daniel Bone, Somer Bishop, Matthew P. Black, Matthew S. Goodwin, Catherine Lord and Shrikanth S. Narayanan, “Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi instrument fusion”, “Journal of Child Psychology and Psychiatry”, 2016, doi:10.1111/jcpp.12559

[7] Fadi Thabtah, “Machine learning in autistic spectrum disorderbehavioral research: A review and ways forward”, “informatics for Health and Social Care”,2018

[8] Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi, Md Nazrul Islam, “A Machine Learning Approach to Predict AutismSpectrum Disorder”, “International Conference on Electrical, Computer and Communication Engineering (ECCE)”, 2019

[9] Md. Shahriare Satu , Farha Farida Sathi, Md. Sadrul Arifen, Md. Hanif Ali and Mohammad Ali Moni, “Early Detection of Autism by Extracting Features:A Case Study in Bangladesh”, “International Conference on Robotics,Electrical and Signal Processing Techniques” , 2019

[10] M. R. Smith, L. J. Williams, "Machine Learning Algorithms in Autism Detection: A Survey of Recent Trends," Journal of Autism and Developmental Disorders, vol. 49, no. 4, pp. 1234-1247, 2020.

[11] J. B. Miller, L. M. Roberts, "The Role of Artificial Intelligence in Early Diagnosis of Autism Spectrum Disorder," Proceedings of the IEEE International Conference on Machine Learning, 2019.

[12] D. T. Jones, P. G. Williams, "Applications of Random Forest Classifiers for Autism Diagnosis: A Review," Journal of Computational Neuroscience, vol. 45, no. 2, pp. 101-113, 2020.

[13] A. S. Patel, B. A. Chen, "Utilizing KMeans for Subgroup Identification in Autism Spectrum Disorder," Neuroinformatics, vol. 8, no. 1, pp. 40-52, 2021.

[14] R. S. Patel, S. L. Yang, "Artificial Neural Networks in Autism Diagnosis: Opportunities and Challenges," Frontiers in Psychology, vol. 11, pp. 1274, 2020.

[15] S. W. McDonald, A. L. Morrison, "Early Detection of Autism with Machine Learning Algorithms: A Comparative Analysis," International Journal of Developmental Neuroscience, vol. 68, pp. 35-44, 2021.

[16] P. A. Kumar, M. V. S. R. Bhargavi, "Predicting Autism Spectrum Disorder Using AdaBoost Classifier," Proceedings of the International

Conference on Machine Learning and Data Mining, 2019.
[17] J. M. Seok, K. M. Lee, "Comparative Analysis of Deep Learning Models in Autism Spectrum Disorder Diagnosis," Journal of Computational Biology, vol. 59, no. 2, pp. 345-356, 2020.
[18] D. H. Harris, L. R. Turner, "Machine Learning Approaches to Autism Screening and Early Diagnosis," Journal of Child and Adolescent Psychiatry, vol. 10, no. 3, pp. 205-215, 2018.
[19] R. T. Zhang, S. T. Li, "Exploring the Role of Decision Trees in Autism Diagnosis and Prediction," IEEE Transactions on Biomedical Engineering, vol. 66, no. 4, pp. 923-931, 2019.
[20] L. C. Anderson, F. J. Harper, "Evaluating the Effectiveness of Machine Learning Algorithms for Autism Spectrum Disorder Prediction," NeuroComputing, vol. 401, pp. 127-134, 2020.