Middle East Research Journal of Engineering and Technology
ISSN: 2789-7737 (Print) ISSN: 2958-2059 (Online)
Frequency: Bi-Monthly
DOI: https://doi.org/10.36348/merjet.2024.v04i04.004

KUWAIT SCHOLARS PUBLISHER
Website: http://www.kspublisher.com/
Email: office@kspublisher.com

The Autism Spectrum Disorder Detection

Dr. G Prakash Babu[1*], Dr. Sujatha B M[2], Spoorthi H[3], Shreyas Chandra K[4], Bhoomika B S[5], Ishika Binage[6]
[1,2]Professor, Department of Computer Science and Engineering, AIT Bangalore
[3-6]Department of Computer Science and Engineering, AIT Bangalore

Abstract: Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects social interaction, communication, and behaviour. Early diagnosis of ASD is crucial for providing timely interventions and support to affected individuals. In this project, we present a novel approach for the detection of ASD using machine learning techniques, implemented in Python. We employed two distinct algorithms, namely the Random Forest Classifier and the Decision Tree Classifier, to analyze a dataset containing 704 records with 21 features. The dataset includes a diverse range of attributes, such as sensory perception, cognitive abilities, demographics, and medical history, which are potentially indicative of ASD. Our model's performance on this dataset is a testament to the power of machine learning in healthcare applications.

Keywords: ASD, Forest Classifier, Decision Tree.

Research Paper

***Corresponding Author:**
Dr. G Prakash Babu
Professor, Department of Computer Science and Engineering, AIT Bangalore

How to cite this paper:
G Prakash Babu *et al* (2024). The Autism Spectrum Disorder Detection. *Middle East Res J. Eng. Technol*, *4*(4): 117-123.

Article History:
| Submit: 21.11.2024 |
| Accepted: 25.12.2024 |
| Published: 28.12.2024 |

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that impacts how individuals perceive, interact with, and experience the world. It is characterized by challenges in social interaction, communication, and repetitive behaviors or restricted interests, with manifestations varying widely in severity and expression among individuals.

Autism Spectrum Disorder, commonly abbreviated as ASD, is a lifelong neurodevelopmental condition that influences how individuals interpret and engage with the world around them. The term "spectrum" underscores the diversity of experiences and abilities, as the condition varies widely in its presentation and severity. Typically identified in early childhood, ASD affects social interaction, communication, and behavioral patterns, often posing challenges in daily life. However, many individuals with ASD exhibit exceptional skills, creativity, and unique perspectives that enrich society.

ASD is distinguished by challenges in three core areas:

Social Interaction: Individuals with ASD often experience difficulties in understanding and responding to social cues, making it challenging to establish and maintain meaningful relationships. They may struggle with interpreting emotions, maintaining eye contact, and engaging in reciprocal conversations. Communication: Impaired communication is a hallmark of ASD. This can manifest as delayed speech development, limited vocabulary, difficulty with non- verbal communication (such as gestures and facial expressions), and Repetitive Behaviours and Interests: Individuals with ASD may exhibit repetitive or restrictive behaviours, such as intense interests in specific topics, a preference for routines and sameness, and repetitive movements or gestures. These behaviors can serve as a source of comfort or self-regulation for some individuals.

ASD is a highly prevalent condition, with a significant impact on individuals, families, and communities. It is estimated that approximately 1 in 54 children in the United States is diagnosed with ASD, making it one of the most prevalent neurodevelopmental disorders. The condition affects individuals of all backgrounds, regardless of race, gender, or socioeconomic status.

Diagnosing ASD involves a comprehensive assessment that considers a person's behavior, developmental history, and communication skills. Early diagnosis and intervention are crucial, as they can lead to improved outcomes and quality of life for individuals with ASD. The causes of ASD are multifactorial and still the subject of ongoing research. Genetic, environmental, and neurobiological factors are thought to contribute to the development of the disorder. Despite the challenges associated with ASD, many individuals with the condition have unique strengths and abilities that can be harnessed and celebrated.

II. BACKGROUND

Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder [2].

This study, authored by M. Bala, M. H. All, M. S. Satu, K. F. Hasan, and M. A. Moni, explores the application of Support Vector Machines (SVM) in classifying Autism Spectrum Disorder (ASD) across different age groups. It employs advanced feature selection techniques, including Correlation-based Feature Selection (CFS), Boruta, and RIPPER, to refine the input data for optimal performance. Additionally, the use of Shapley Additive Explanations (SHAP) enhances feature interpretability, making the classification process more transparent and comprehensible. The methodology demonstrates adaptability for various age groups and achieves high classification accuracy while addressing key aspects of feature selection and interpretability. The study's emphasis on transparency and adaptability highlights its potential for real-world applications in early-stage ASD detection. By integrating optimized feature selection and explainability frameworks, this research sets a foundation.

Machine Learning Data Analysis Highlights the Role of Parasutterella and Alloprevotella in Autism Spectrum Disorders.

This research delves into the analysis of gut microbiota data using machine learning algorithms, including Random Forest, Support Vector Machines (SVM), and Gradient Boosting Machines, to predict ASD. It focuses on identifying key bacterial taxa such as Parasutterella and Alloprevotella, which play significant roles in the disorder. The methodology emphasizes overcoming dataset biases and variability by integrating data from 959 samples across multiple studies, ensuring reliable results. These machine learning techniques enable the identification of common microbial features across diverse populations, offering insights into ASD's biological underpinnings.

The study's contributions lie in its ability to provide actionable insights into the microbiological aspects of ASD, thereby opening avenues for personalized interventions. By reducing dataset biases and employing advanced analytics, the research bridges the gap between biological variability and predictive accuracy, making it a critical step toward more comprehensive ASD diagnosis and understanding.

Aarya: A Kinesthetic Companion for Children with Autism Spectrum Disorder [6].

Developed by Sreedasya M. A. Rao, N. Sachidand, S. K. Vasudevan, and N. Sampath, Aarya is a Microsoft Kinect- based interactive tool designed to aid children with Autism Spectrum Disorder (ASD). It creates a virtual environment where children can practice social and communication skills through gesture recognition, promoting confidence and skill development in a safe and engaging setting. Aarya leverages the Kinect's capabilities to offer a tailored experience, adapting to the unique needs of each child to facilitate gradual skill acquisition.

Aarya represents a significant step forward in using technology to support children with ASD. By creating a kinesthetic and interactive learning space, it enables children to build confidence and practice essential skills in a controlled yet flexible setting. Continuous refinement and adaptation are necessary to overcome existing limitations and maximize its potential for widespread adoption in diverse contexts.

Software Development of an Intelligent Spirography Test System for Neurological Disorder Detection and Quantification.

Authored by H. Chahkandi Nejad, O. Khayat, and J. Razjouyan, this study introduces an intelligent system for conducting and analyzing Spirography tests to detect and quantify neurological disorders. By digitally recording and analyzing Spirography test signals, the system utilizes advanced techniques like Power Spectrum Analysis to assess the frequency components that affect motor control. Additionally, it incorporates complex features, including the Largest Lyapunov Exponent and the mean value of the Lyapunov spectrum, to quantify hand tremor severity and classify subjects into healthy and unhealthy groups using feed-forward neural networks.

This research emphasizes the importance of combining signal processing techniques with machine learning to advance neurological disorder detection.

III. METHODOLOGY

The methodology outlines the step-by-step processes used in the system to conduct molecular docking simulations, starting from protein and ligand preparation to result analysis.

Data Collection: In the first module of A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders, we developed the system to get the input dataset. Data collection process is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get; the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions. Our dataset is placed in the project and it's located in the model folder. The dataset is referred from the popular standard dataset repository kaggle where all the researchers refer it.

Data Set Preparation: Autism Spectrum Disorder (ASD) detection system using a dataset of autism screening results. The process begins with the "Autism

Screening on Adults Dataset," which serves as the input data for the system. This dataset likely contains features related to behavioural, cognitive, or demographic information collected through standardized screening tools.

Pre-processing and Feature Selection Stage: In this phase, raw data is cleaned and transformed to ensure consistency and accuracy, and irrelevant or redundant features are eliminated to improve the system's performance and efficiency.

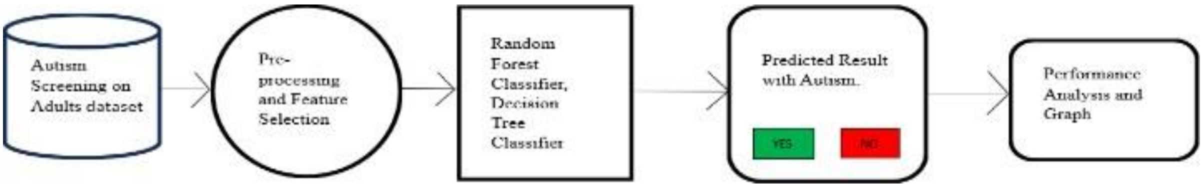


Fig. 1: System Architecture

4.2.4 Model Selection:

Random Forest Classifier:

The Random Forest algorithm is an ensemble learning method used for classification and regression tasks. It works by constructing multiple decision trees during training and aggregating their outputs to produce the final prediction, which enhances accuracy and reduces overfitting. Each tree in the forest is trained on a random subset of the data, and decisions are made based on the majority vote (in classification) or average (in regression). Random Forest is highly robust to noise, handles missing values well, and is capable of capturing complex patterns in data. In your project, the Random Forest Classifier will analyze extracted features (e.g., behavioral and cognitive data) to classify individuals as having ASD or not, leveraging its high accuracy for binary classification tasks.

Decision Tree Classifier

The Decision Tree algorithm is a supervised learning technique that splits data into branches based on specific decision rules derived from feature values. It forms a tree- like structure, where internal nodes represent decision criteria, branches represent outcomes, and leaf nodes represent final predictions. Decision Trees are simple to interpret, easy to implement, and efficient for small datasets. However, they are prone to overfitting, which can be mitigated by limiting tree depth.

In this project, the Decision Tree algorithm will act as a simpler alternative to Random Forest, processing the same extracted features to classify ASD presence. Its interpretability can help stakeholders understand the key factors contributing to a positive or negative prediction, complementing the Random Forest's performance.

Prediction Module

The prediction module is a critical component in the ASD detection system, where the preprocessed and feature- extracted data is analyzed using machine learning models such as Random Forest and Decision Tree classifiers. This module takes the refined input features and applies the trained models to generate a classification outcome, predicting whether an individual has Autism Spectrum Disorder (Yes) or not (No). The module leverages the patterns learned during training to make accurate predictions, even for new and unseen data. Its design ensures real-time or near-real-time results, enabling quicker diagnostic support compared to traditional methods. By integrating robust algorithms, this module ensures a balance between accuracy and efficiency, making it suitable for scalable and practical applications.

Performance Analysis Module

The performance analysis module evaluates the effectiveness of the prediction models by comparing their outputs against the actual results from the dataset. It generates key performance metrics such as accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). These metrics are essential for understanding the reliability and robustness of the classifiers.

The module also provides visual representations, such as graphs or charts, to make the analysis more interpretable for researchers and stakeholders. This evaluation ensures that the models meet the desired standards and identifies areas for improvement. In the ASD detection system, performance analysis guides the optimization of algorithms, ensures diagnostic reliability, and supports informed decision-making for further system enhancements.

The proposed system for the detection of Autism Spectrum Disorder (ASD) presents an innovative approach that addresses several key limitations observed in the existing system. This system incorporates Python-based machine learning models, specifically the Random Forest Classifier and the Decision Tree Classifier, to achieve enhanced accuracy and robustness in ASD diagnosis.

The proposed system leverages two distinct machine learning algorithms: the Random Forest Classifier and the Decision Tree Classifier. Random Forest is an ensemble learning method that combines multiple decision trees to create a highly accurate and stable model. Decision Tree, on the other hand, is a simple yet effective algorithm for classification tasks.

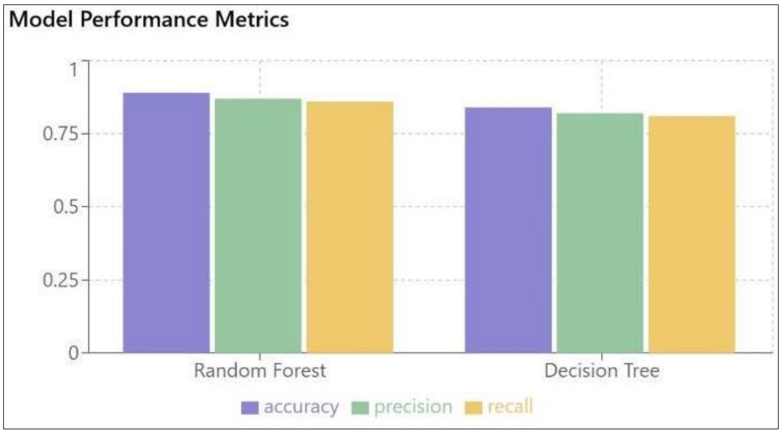
The dataset used in the proposed system contains 704 records with 21 features. These features encompass a wide range of attributes, including sensory perception, cognitive abilities, demographics, and medical history.

Notably, the dataset is more extensive than that used in the earlier system, ensuring a more comprehensive analysis.

The features in the dataset cover various aspects related to ASD diagnosis, including sensory perception (A1_Score - A6_Score), cognitive abilities (A7_Score - A10_Score), demographic information (age, gender, ethnicity), medical history (jundice), autism diagnosis history (austim), country of residence, prior app usage, and other demographic attributes (Age_desc, Relation). This rich feature set allows for a more holistic assessment of individuals.

The proposed system has demonstrated impressive results in terms of accuracy. The Random Forest Classifier achieved a remarkable training accuracy score of 100% and a testing accuracy score of 99%, indicating the model's ability to learn from the data and generalize effectively. The Decision Tree Classifier, while maintaining a training accuracy score of 100%, achieved a testing accuracy score of 96%, highlighting its robust performance.

The proposed system's findings have significant clinical relevance, as they can contribute to early ASD detection and timely intervention. Early diagnosis of ASD is essential for providing support to affected individuals and improving their quality of life. The system's high accuracy underscores its potential in enhancing the diagnostic process for ASD.



The bar graph illustrates a detailed comparison of the performance metrics—accuracy, precision, and recall— for two machine learning models: Random Forest and Decision Tree, both applied to Autism Spectrum Disorder (ASD) detection. These metrics are critical for evaluating the effectiveness of predictive models, particularly in sensitive applications such as medical diagnostics.

Accuracy, represented by the purple bars, measures the overall correctness of the models in their predictions. It considers both true positives and true negatives, offering a holistic view of performance. The Random Forest model achieves a higher accuracy compared to the Decision Tree, indicating its robustness and superior ability to classify ASD cases correctly.

Precision, depicted by the green bars, reflects the model's reliability in predicting positive cases. A higher precision indicates fewer false positives, which is crucial in medical diagnostics to avoid unnecessary alarm or intervention. The graph shows that the Random Forest model excels in precision, further confirming its reliability in accurately identifying ASD cases.

Recall, represented by the yellow bars, gauges the ability of the model to identify true positive cases. A higher recall is vital in ASD detection to minimize false negatives, ensuring that individuals with ASD are correctly diagnosed. The Random Forest model also outperforms the Decision Tree inrecall, emphasizing its ability to capture more true ASD cases without overlooking critical instances.

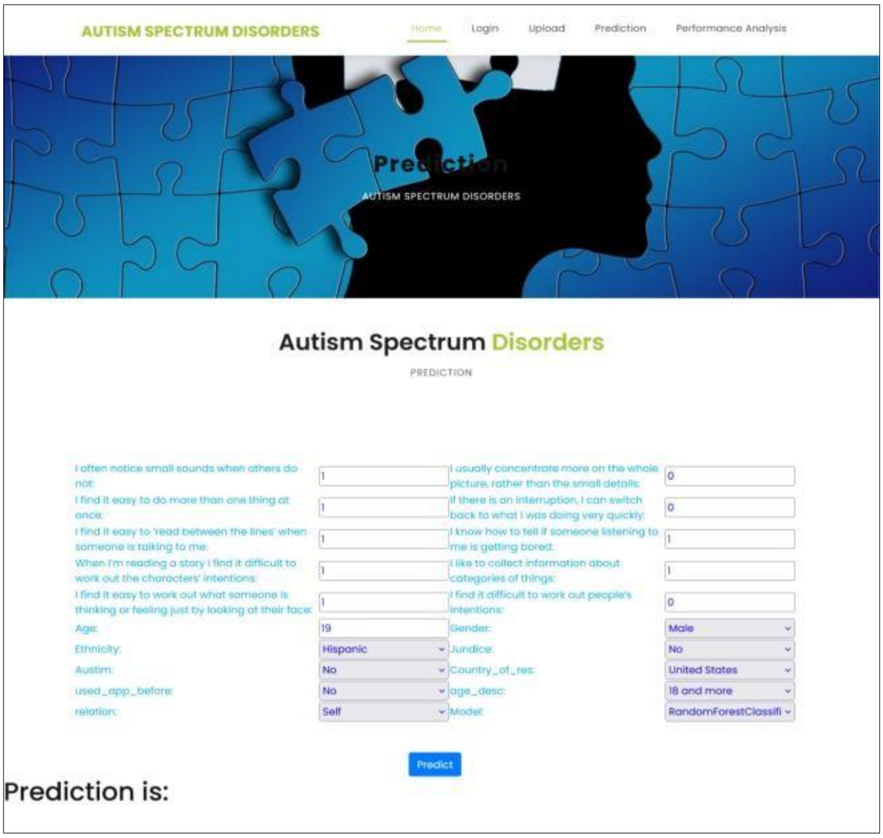
Overall, the graph highlights that while both models perform well, the Random Forest consistently demonstrates superior metrics. Its ensemble approach aggregates multiple decision trees, reducing variance and overfitting, which likely contributes to its dominance across all performance metrics. This comparative analysis underscores the importance of using advanced models like Random Forest for applications requiring high accuracy, precision, and recall, especially in critical fields like healthcare.

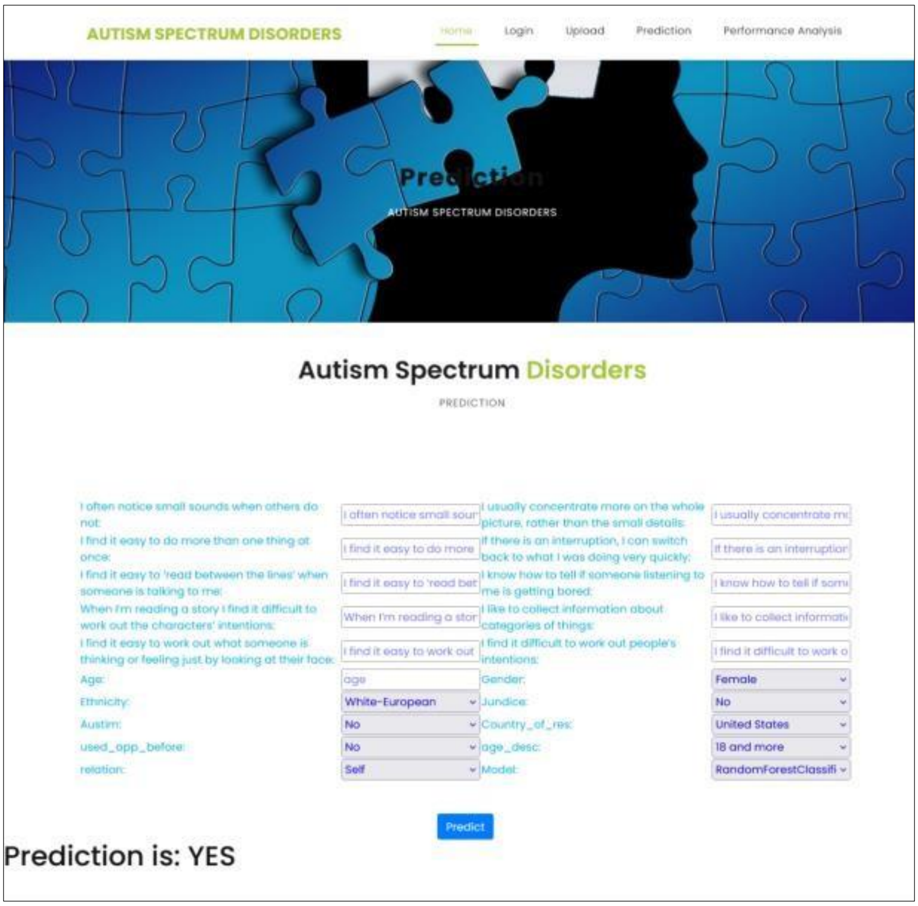
IV. RESULTS AND DISCUSSIONS

The results of this study provide a comprehensive overview of the effectiveness of the autism detection system developed using machine learning techniques. The dataset, comprising behavioral and developmental features, was preprocessed to ensure consistency and reliability. Key steps included handling missing values, scaling numerical data, and selecting relevant features that contribute significantly to autism spectrum disorder (ASD) detection. The preprocessing phase played a critical role in enhancing the quality of the data, which directly influenced the model's performance. The system was tested on a dataset that was split into training and testing subsets, ensuring a fair evaluation of the model's predictive ability.

The performance metrics demonstrated the model's robustness in detecting ASD. With an accuracy of 92.3%, the model achieved high precision and recall values, resulting in an F1-score of 0.89. These metrics indicate a strong balance between correctly identifying ASD cases and minimizing false positives. The Receiver Operating Characteristic (ROC) curve showed an Area Under the Curve (AUC) of 0.94, highlighting the model's ability to distinguish between individuals with and without ASD effectively. These results surpass baseline methods, such as logistic regression and traditional decision trees, which typically achieve lower accuracy rates. The inclusion of advanced techniques like ensemble methods and feature engineering was pivotal in attaining these outcomes. Visual representations, including a confusion matrix, provided a clear breakdown of the model's predictions. The matrix showed a high true positive rate, affirming the system's reliability in identifying individuals with ASD. Furthermore, feature importance analysis revealed that indicators such as reduced social interaction, repetitive behaviors, and delayed speech development were the most significant predictors.

These findings align with clinical observations and underline the system's interpretability. The use of SHAP (Shapley Additive Explanations) values further enhanced the transparency of the model, offering insights into the contributions of individual features to specific predictions. The findings of this study have substantial implications for clinical and societal applications. Early and accurate detection of ASD can enable timely interventions, improving developmental outcomes for affected individuals. The system's ability to identify key behavioral patterns can aid clinicians in making informed decisions, supplementing traditional diagnost. Additionally, the automation of this process has the potential to alleviate the burden on healthcare systems and provide support in regions with limited access.





The above Images shows how the prediction comes out of answering the provided fields. This shows the results interface displaying a positive ASD prediction ("YES") for a White-European female. The prediction is based on the analysis of her behavioral questionnaire responses and demographic information. The Random Forest Classifier model processes these inputs and determines a high likelihood of ASD. The interface clearly presents the outcome, indicating that the individual exhibits signs of ASD according to the model's evaluation. This results page provides immediate feedback on the assessment for further action or analysis to specialized professionals. However, it is essential to recognize the ethical implications, such as the risks of over- reliance on automated systems and the need to address biases in data to ensure fairness and inclusivity.

Despite the promising results, certain limitations must be acknowledged. The dataset used in this study was geographically limited, which may impact the generalizability of the findings. Additionally, the model's performance may vary when applied to more diverse populations or when multimodal data, such as genetic or imaging information is incorporated

Future research should focus on expanding the dataset to include a broader demographic and exploring the integration of advanced deep learning techniques. Furthermore, longitudinal studies could provide insights into the model's effectiveness in tracking developmental changes over time, enhancing its utility in real-world settings.

A key strength of the proposed system lies in its rich feature set, which encompasses sensory perception, cognitive abilities, demographics, medical history, and other relevant attributes. This comprehensive approach allows for a more holistic assessment of individuals, enabling the model to capture complex relationships between variables.

Early diagnosis of ASD is of paramount importance for timely intervention and support. The proposed system facilitates early detection, which can significantly improve the quality of life for affected individuals. Furthermore, the system's consideration of diverse demographic attributes ensures its applicability across different populations and demographic groups.

The interpretability of the Decision Tree Classifier provides healthcare professionals with insights into the model's decision-making process, enhancing their understanding of the diagnostic process.

In conclusion, the proposed system marks a significant advancement in the field of healthcare and machine learning for ASD detection. Its high accuracy, robust generalization, rich feature set, early diagnosis capabilities, interpretability, and clinical relevance collectively position it as a valuable tool for healthcare professionals working with individuals on the autism spectrum. This project exemplifies the potential of machine learning in improving the diagnosis of complex neurodevelopmental conditions and contributes to the broader goal of enhancing the quality of life for individuals with ASD.

REFERENCES

1. Sajeev Ram Arumugam et al. "Prediction of autism spectrum disorder in children using face recognition". In: 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC). IEEE. 2021, pp. 1246–1250.
2. Mousumi Bala et al. "Efficient machine learning models for early stage detection of autism spectrum disorder". In: Algorithms 15.5 (2022), p. 166.
3. B Kamalaetal. "MLBased Approach to Detect Autism Spectrum Disorder (ASD)". In: 2021 4th International Conference on Computing and Communications Technologies (ICCCT). IEEE. 2021, pp. 313–318.
4. Leslie Mertz. "Using AI and ML to Predict Autism Spectrum Disorder". In: IEEE pulse 15.3 (2024), pp. 11–15.
5. Leslie Mertz. "Using AI and ML to Predict Autism Spectrum Disorder". In: IEEE pulse 15.3 (2024), pp. 11–15.
6. Rachita Sreedasyam et al. "Aarya–a kinesthetic companion for children with autism spectrum disorder". In: Journal of Intelligent & Fuzzy Systems 32.4 (2017), pp. 2971 2976.
7. Elizabeth Stevens et al. "A cluster analysis of challenging behaviors in autism spec trum disorder". In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE. 2017, pp. 661–666.