AUTISM SPECTRUM DISORDER ANALYSIS AND DETECTION SYSTEM

PHASE 1 REPORT

Submitted by

MOHANAPRIYA E (210701164)

MUKHILAN S S (210701169)

in partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

NOVEMBER 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this project titled “**AUTISM SPECTRUM DISORDER ANALYSIS AND DETECTION SYSTEM**” is the bonafide work of “**MOHANAPRIYA E (210701164) , MUKHILAN S S(210701169)**” who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other project or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Signature	**Signature**
Dr. P. Kumar,M.E.,Ph.D.,	Dr. K. Ananthajothi,M.E.,Ph.D.,
HEAD OF THE DEPARTMENT	**SUPERVISOR**
PROFESSOR	**PROFESSOR**
Department of Computer Science and Engineering	Department of Computer Science and Engineering
Rajalakshmi Engineering College	Rajalakshmi Engineering College
Chennai – 602 105	Chennai – 602 105

Submitted to Project Viva-Voce Examination held on ______________

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through everywalk of our life and showering his blessings through the endeavorto put forth this report. Our sincere thanks to our respected Chairperson **Dr.(Mrs.) Thangam Meganathan**, **Ph.D.,** Our respected Chairman**Mr. S. Meganathan, B.E, F.I.E.,** and our respected ViceChairman **Mr. Abhay Shankar Meganathan, B.E., M.S**., for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution. Our sincere thanks to **Dr. S.N. Murugesan, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr.P.Kumar,M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr.K,Ananthajothi, M.E.,Ph.D.,** Department ofComputer Science and Engineering. Rajalakshmi Engineering College for her valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator, **Dr.K.Kumaragurubaran, M.E.,Ph.D.,** Department of Computer Science and Engineering for his useful tips during our review to build ourproject.

MOHANAPRIYA E 210701164

MUKHILAN S S 210701169

ABSTRACT

Autism Spectrum Disorder is a neurodevelopmental disorder characterized by social interaction, communication impairments, and repetitive behaviors. In recent years, there has been an increase in prevalence, and it is important to identify individuals early for appropriate intervention to improve their quality of life. The following study was conducted to deepen the understanding and management of the disorder by employing a comprehensive two-module approach. Module one emphasizes the detection of ASD through sophisticated machine learning techniques for early and accurate diagnosis. Module two is on improving social and communication skills of children with ASD via game-based assessments, offering them an interactive yet enjoyable learning environment. This project, not only in contributingto the scientific community, but also through practical applications contributes and offers easy solutions for caregivers and educators. Integration of technology with both detection and intervention holds huge advancement in this area, promising better outcomes for ASD individuals.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder that affects individuals' ability to communicate, interact socially, and adapt to changes in behavior or routine. According to the World Health Organization (WHO), ASD affects 1 in 100 children worldwide, making it a significant public health concern. Early diagnosis and intervention are crucial in mitigating the long-term effects of ASD, as they can significantly improve cognitive and social outcomes for individuals.

However, diagnosing ASD poses several challenges. Traditional diagnostic methods, such as the Autism Diagnostic Observation Schedule (ADOS) and clinical assessments, are resource-intensive, time-consuming, and heavily reliant on subjective interpretations by trained professionals. Additionally, these methods require the availability of skilled personnel, making early detection inaccessible in underserved or remote areas.

Advancements in technology, particularly machine learning and artificial intelligence, present an opportunity to address these challenges. By analyzing large datasets containing demographic, behavioral, and clinical features, machine learning algorithms can identify patterns indicative of ASD with high accuracy and consistency. This automated approach reduces the dependency on human expertise and acceleratesthe diagnostic process.

The goal of this project is to develop a machine learning-based diagnostic tool that can detect ASD efficiently and accurately. The system uses structured datasets, feature selection techniques, and state-of-the-art machine learning models to classify individuals as having ASD or not. By leveraging models like XGBt, Random Forest, and Support Vector Machines (SVM), the project aims to bridge the gap between traditional diagnostic methods and modern technology.

1.2 OBJECTIVE

The primary objective of this project is to develop an efficient, accurate, and scalable system for the early detection of Autism Spectrum Disorder (ASD) using advanced machine learning algorithms. By analyzing structured data such as demographic information, behavioral assessments, and family history, the system aims to identify patterns and correlations indicative of ASD. The project seeks to overcome the limitations of traditional diagnostic methods, which are often time-consuming, resource- intensive, and reliant on subjective clinical observations.

1.3 EXISTING SYSTEM

Standardized Questionnaires and Surveys : Existing systems for assessing ASD often rely on standardized questionnaires and surveys, such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R). These tools involve a series of structured questions and observational tasks designed to evaluate specific behaviors and developmental milestones associated with ASD.

Medical and Psychological Evaluations : Comprehensive medical and psychological evaluations are part of the existing system for diagnosing ASD. These evaluations may include physical examinations, neurological assessments, and psychological testing to rule out other conditions and gain a thorough understanding of the child's cognitive functioning. While these evaluations are essential for a complete diagnostic picture, they are often conducted in clinical settings that can be intimidating for the child..

1.4 PROPOSED SYSTEM

Autism, or autism spectrum disorder (ASD), refers to a broad range of conditions characterized by challenges with social skills, repetitive behaviors, speech and nonverbal communication.This project aims to significantly enhance the assessment and development of key skills in children with ASD through an innovative game-based approach.By participating in various game-based scenarios, children will have opportunities to practice and improve essential skills such as communication, social interaction, and problem-solving in a supportive and enjoyable environment.

CHAPTER 2

LITERATURE SURVEY

[1] Self-imposed methods of dietary control in children with autism lead to poor nutrition and food deficiencies. This study examines the dietary habits of children with Autism Spectrum Disorder (ASD) and other neurodevelopmental conditions. A survey conducted on 141 children revealed patterns of self-imposed dietary restrictions, such as gluten-free and casein-free diets. While such diets are often perceived as beneficial, they lead to significant nutritional deficiencies and potential health issues. The study highlights the urgent need for nutritional guidance tailored to children with ASD,ensuring balanced diets without compromising health.

[2] A comprehensive solution to the problem of longer waiting periods for the diagnosis of Autism Spectrum Disorder (ASD) is presented, which is the development of a system that utilizes VR for screening as well as classifying autism from non-verbal behavior analysis.The research proposes a virtual reality (VR) screening system toaddress delays in ASD diagnosis. Participants interact with a virtual shop assistant in a simulated shopping scenario, during which non-verbal behaviors like gaze direction and head movements are captured. Using machine learning techniques, the system analyzes these behaviors to classify individuals as autistic or non-autistic. This novel approach combines immersive technology with advanced analytics, significantly reducing diagnostic timelines.

[3] A Platform for Autism Home-Based Therapeutic Intervention highlights the importance of providing cost-effective home-based interventions for children with Autism Spectrum Disorder (ASD) because conventional ABA therapy is quite expensive.This study introduces a platform comprising a mobile and desktopapplication to facilitate in-home therapeutic interventions for children with ASD. The platform provides customized content and guidelines for parents to administer therapy sessions, reducing the financial burden of traditional Applied Behavior Analysis (ABA) therapy. This innovation ensures greater accessibility and empowers families to activelyparticipate in their child’s developmental journey.

[4] Incorporating the Help of a Computer Program With Tangible Interfaces in Teaching Emotions to Children With Autism Spectrum Disorder.This research focuses on addressing the challenges children with ASD face in recognizing and expressing emotions. The study introduces "Tangible Emotions," a gamified computer program that incorporates tangible user interfaces. Children use physical representations of emotions, such as joy or sadness, to interact with the system, which provides immediate feedback. This interactive learning approach enhances emotional intelligence and social integration for children with ASD.

[5] Autistic spectrum conditions (ASC) present a variety of challenges, one of which is the acquisition of an appropriate emotion and its expression which impedes socialization.The study develops a serious game integrated with an emotion recognition system to help children with high-functioning autism recognize and express emotions. The system uses RGB-D sensors to capture body movements and employs a linear Support Vector Machine (SVM) for emotion classification. This innovative approach promotes socialization and emotional learning in children with ASC.

[6] The research tackles issues related to the diagnosis and treatment of Autism Spectrum Disorder (ASD).The study combines Eye Tracking (ET) and Electroencephalography (EEG) to explore reliable diagnostic metrics for ASD. By analyzing visual attention patterns and brain activity, the researchers identify unique neural markers associated with ASD. This integrated approach enhances the precision and objectivity of ASD diagnosis, paving the way for new diagnostic tools.

[7] Autistic Children Probability Estimation Using Hidden Markov Models reveals the genetic heritability of Autism Spectrum Disorder (ASD).Using Hidden Markov Models (HMMs), this study predicts the likelihood of ASD in children based on genetic data. It identifies probabilities of approximately 33% for female children and 80% for male children inheriting ASD from autistic parents. This work highlights the strong genetic component of ASD and its implications for early diagnosis.

[8] Using statistical information about the heritability of autism and the sister-brother recurrence of autism, the authors establish an HMM that predicts if a child is likely to be autistic based on their parent's features.Building on previous studies, this research delves deeper into genetic inheritance patterns using HMMs. It focuses on

sibling recurrence rates and parental features to provide probabilistic models for ASD prediction, aiding in early interventions and genetic counseling.

[9] Diagnosing Adults with High-Functioning Autism through Eye Tracking and Machine Learning.This study evaluates the potential of eye-tracking technology to diagnose high-functioning autism (HFA) in adults. Participants' eye movements were recorded while they interacted with web pages, and the data was analyzed using machine learning classifiers. With a detection accuracy of 74%, the study demonstrates the viability of combining eye tracking and machine learning for adult ASD diagnosis.

[10] Discriminative Few Shot Learning of Facial Dynamics in Interview Videos for Autism Trait Classification.This research addresses the labor-intensive process of analyzing Autism Diagnostic Observation Schedule (ADOS) videos. By employingfew-shot learning methods, the study extracts spatio-temporal features from video recordings to classify individuals into Autism, Autism Spectrum, or Non-Spectrum categories. This automated approach reduces reliance on manual annotation while improving the accuracy of ASD trait classification.

[11] The Cognitive and Neural Basis of Systemizing.The study introduces the "hyper-systemizing" theory, which explains the exceptional pattern-recognition abilities often observed in individuals with ASD. These abilities are linked to an enhanced genetic predisposition for systemizing, making individuals with ASD adept at logical reasoning and problem-solving in fields like mathematics, music, and engineering.

[12] Characterizing Autism Spectrum Disorder Through Fusion of Local Cortical Activation and Global Functional Connectivity Using Game-Based Stimuli and a Mobile EEG System.This study uses a game-based EEG system to analyze brain activity during social cognitive tasks. By examining cortical activation and functional connectivity, the research identifies biomarkers for ASD diagnosis. This innovative approach combines neuroscience and machine learning to offer objective and data- driven diagnostic tools.

[13] Prediction of Symptom Severity in Autism Spectrum Disorder Using EEG Metrics.Using EEG data, this study predicts the severity of ASD symptoms, aiming to reduce the subjectivity of clinical assessments. By analyzing features such as power spectral density, spatial patterns, and network properties, the researchers provide quantitative biomarkers for symptom severity, aiding personalized treatment planning.

[14] The role of the CNTNAP2 gene in the development of Autism Spectrum Disorder (ASD).This research investigates the association of the CNTNAP2 gene variant with ASD and explores the effects of prenatal exposure to valproic acid (VPA) on gene expression in animal models. Findings suggest that genetic and environmental factors jointly influence ASD development, providing insights into its biological underpinnings.

[15] Less frequent face looking at infancy is related to the likelihood status of autism but not the diagnosis.This study examines early social attention patterns in infants at risk for autism. By analyzing video recordings of parent-infant interactions, the researchers found that reduced face-looking behavior is linked to the likelihood of ASD diagnosis. These findings highlight early behavioral markers for ASD.

[16] Relationship between Anxiety, Repetitive Behavior, and Parenting Stress: A Cross-Comparative Study between Spanish and Colombian Individuals with Autism.This study explores cultural differences in anxiety, repetitive behaviors, and parenting stress in families of children with ASD in Spain and Colombia. Data from 118participants reveal significant correlations between these factors, emphasizing the importance of culturally tailored interventions.

[17] Oral Health and Quality of Life among Autistic Spectrum Disorder Individuals.This study examines the impact of oral health on the quality of life in children with ASD. Using tools like the Oral Health Assessment Tool (OHAT) and EQ-5D-Y, the researchers assessed the dental health and overall well-being of 163 children, highlighting the need for better oral hygiene practices in this population.

[18] Vestibular Function and Postural Control in Children with Autism Spectrum Disorder.This research investigates the link between vestibular function and motor balance in children with ASD. Using vestibular evoked myogenic potentials and static posturography, the study identifies sensory processing deficits that affect postural control, contributing to motor coordination challenges in ASD.

[19] Stories of South Asian parents around the diagnosis of autism in their children and early services in Australia.This study explores the challenges faced by South Asian parents in Australia when seeking autism diagnosis and early intervention services. Through participatory research, the study highlights cultural barriers, long

waiting times, and a lack of community-specific resources, emphasizing the need for inclusive healthcare solutions

[20] Early Auditory Temporal Processing Deficit in Children with Autism Spectrum Disorder: The Research Domain Criteria Framework.This study analyzes auditory temporal processing deficits in children with ASD. Using event-related potentials (ERP), the researchers compared mismatch negativity (MMN) responses between children with ASD and typically developing peers. The findings suggest that deficits in processing sound durations and inter-stimulus intervals contribute to social and communication challenges in ASD.

[21]. Tariq, Q., Fleming, S.L., Schwartz, J.N., Dunlap, K., Corbin, C., Washington, P., Kalantarian, H., Khan, N.Z., Darmstadt, G.L., Wall, D.P., 2019. Detecting developmental delay and autism through machine learning models usinghome videos of Bangladeshi children: Development and validation study. Journal of Medical Internet Research, 21, e13822.This study explores the use of machine learning models to detect developmental delays and autism by analyzing home videos of Bangladeshi children. The researchers collected and annotated home videos, focusingon behavioral markers associated with autism and developmental delays. Machine learning algorithms were then trained on these annotations to identify behavioral patterns indicative of neurodevelopmental disorders. The approach is cost-effective and scalable, making it suitable for low-resource settings. The study demonstrates the feasibility of using non-invasive, video-based assessments for early autism detection, particularly in regions with limited access to professional diagnostic services.

[22]. Thabtah, F., 2017a. Autism Screening Adult. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5F019.This paper introduces a dataset for autism screening in adults, hosted in the UCI Machine Learning Repository. The dataset consists of responses to a series of screening questions designed to identify autistic traits in adults. The study aims to facilitate the development of machine learning models that can predict autism spectrum disorder (ASD) in adult populations. By making the dataset publicly available, the research encourages further exploration and benchmarking of predictive models. The work provides a foundation for developing automated tools to assist in adult autism diagnosis.

[23]. Thabtah, F., 2017b. Autistic Spectrum Disorder Screening Data for Children. UCI Machine Learning Repository. DOI:

https://doi.org/10.24432/C5659W.This paper presents a publicly available dataset for screening autism in children, curated and hosted in the UCI Machine Learning Repository. The dataset includes demographic and behavioral features derived from screening questionnaires. The goal is to enable researchers to develop and test machine learning models for early autism detection in children. The dataset supports comparative analysis of different algorithms and serves as a benchmark for evaluating the accuracy and efficiency of predictive models. This contribution plays a critical role in advancing automated ASD screening tools for pediatric populations.

[24]. Thabtah, F., Peebles, D., 2019. Early autism screening: A comprehensive review. International Journal of Environmental Research and Public Health, 16, 3502.This review comprehensively examines the existing methods for early autism screening, highlighting their strengths, limitations, and opportunities for improvement. The paper discusses traditional diagnostic tools such as Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R), as well as emerging machine learning-based approaches. The authors emphasize the potential of data-driven models in improving diagnostic accuracy, reducing reliance on subjective assessments, and increasing accessibility. They also discuss the ethical and practical challenges associated with implementing machine learning in clinical settings. This review serves as a valuable resource for understanding the state-of-the-art in autism screening research.

[25]. Thabtah, F., Peebles, D., 2020. A new machine learning model based on induction of rules for early autism screening.This study proposes a novel machine learning model that uses rule induction for early autism screening. The model is designed to extract interpretable decision rules from data, making it easier for clinicians and researchers to understand the underlying patterns. By leveraging both behavioral and demographic features, the model achieves high accuracy while maintaining simplicity and transparency. The authors demonstrate the effectiveness of their approach through experiments on publicly available datasets, highlighting its potentialto support clinicians in making timely and accurate ASD diagnoses.

[26] Joudar, S.S., Albahri, A., Hamid, R.A., 2022. Triage and priority-based healthcare diagnosis using artificial intelligence for autism spectrum disorder and gene contribution: A systematic review. Computers in Biology and Medicine, 146, 105553.This paper conducts a systematic review of artificial intelligence (AI)-driven

approaches for triage and prioritization in healthcare diagnosis, specifically focusing on Autism Spectrum Disorder (ASD) and gene contribution. The study evaluates existing AI models that incorporate genetic data to predict ASD susceptibility and severity. By examining a range of triage systems, the review highlights the potential of AI in optimizing diagnostic workflows, enabling faster and more accurate identification of ASD cases. This work emphasizes integrating AI with genomic data to develop personalized diagnostic and therapeutic solutions for ASD

[27] Jung, Y., 2018. Multiple predicting k-fold cross-validation for model selection. Journal of Nonparametric Statistics, 30, 197–215.This study introduces anovel variation of k-fold cross-validation, called "Multiple Predicting k-Fold Cross-Validation," designed to enhance model selection in machine learning. The method improves upon traditional k-fold cross-validation by minimizing overfitting and providing more robust performance estimates, especially for small datasets. The paper demonstrates the utility of this approach in selecting optimal models for autism-related diagnostic tasks, ensuring higher generalization accuracy. This method is particularly relevant for datasets in autism studies, which often suffer from class imbalances and limited sample sizes.

[28] Kashef, R., 2022. ECNN: Enhanced convolutional neural network for efficient diagnosis of autism spectrum disorder. Cognitive Systems Research, 71, 41–49.This paper proposes the Enhanced Convolutional Neural Network (ECNN), a deep learning architecture tailored for ASD diagnosis. ECNN leverages convolutional layers optimized for extracting features from structured clinical data and neuroimaging datasets. The model achieves superior accuracy and computational efficiency compared to traditional CNNs and other baseline models. The study underscores the potential of ECNN to automate ASD diagnosis with minimal manual intervention, making it a valuable tool for both clinicians and researchers.

[29] Landa, R.J., 2008. Diagnosis of autism spectrum disorders in the first 3 years of life. Nature Clinical Practice Neurology, 4, 138–147.This paper reviews methods for diagnosing ASD in children within the first three years of life, focusing on behavioral and developmental markers. The author discusses early diagnostic criteria, including delays in communication, repetitive behaviors, and atypical social interactions.

CHAPTER 3

SYSTEM DESIGN

3.1 GENERAL

The proposed system is a machine learning-based diagnostic tool designed to detect Autism Spectrum Disorder (ASD) with high accuracy and efficiency. Unlike traditional diagnostic approaches, which are heavily reliant on subjective assessments and extensive clinical evaluations, this system leverages structured data and advanced algorithms to provide an objective and scalable solution for ASD detection.

3.2 SYSTEM ARCHITECTURE DIAGRAM

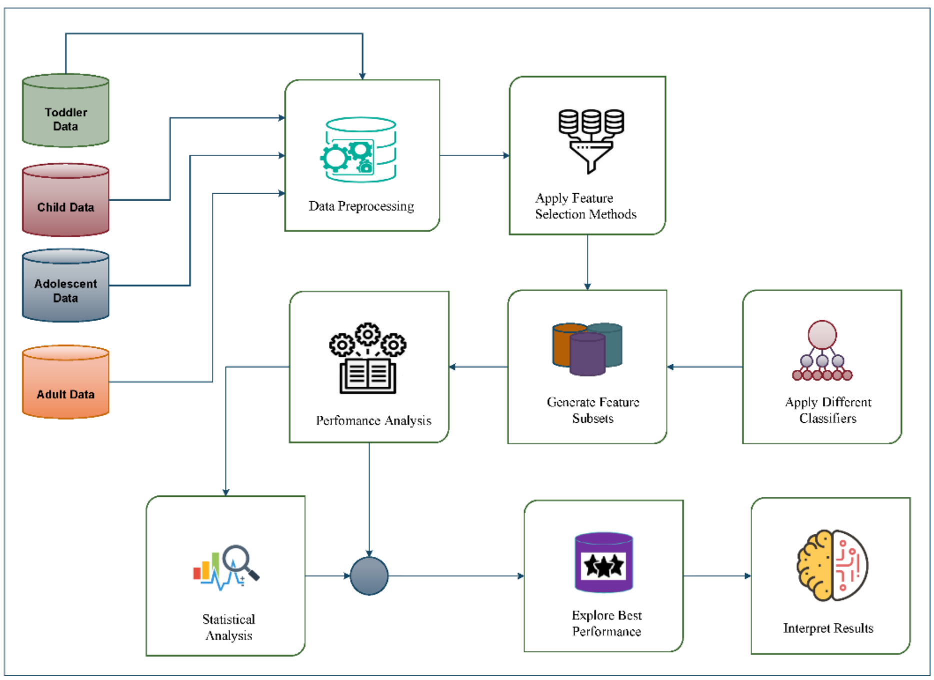

Fig. 3.1 System Architecture

The figure 3.1 shows the architecture of the system is divided into two main packages: Data Processing and Model Training. In the Data Processing package, raw data is collected from various sources and undergoes preprocessing to handle missing values, remove duplicates, perform one-hot encoding, and apply Min-Max normalization. This results in preprocessed data, which is ready for the next stage. In theModel Training package, the preprocessed data is used to train various machine learningmodels. These models are then evaluated based on their performance metrics. The final step involves validating the models to ensure they perform well on unseen data, ensuring the robustness and reliability of the system.

3.3 SYSTEM REQUIREMENTS

3.3.1 HARDWARE REQUIREMNETS

These hardware requirements form the basis for the implementation of the university management system, which embodies a complete and consistent specification which is essential for proper deployment of the software.

The system setup includes:

- Processor: Intel i5 or higher
- RAM: Minimum 8GB RAM
- Storage: 500GB - 1TB hard drive
- Graphics Processing Unit (GPU): Integrated or dedicated graphics card

3.3.2 SOFTWARE REQUIREMENTS

The specification of the system is the software requirements document. There should be a definition and a list in which needs would be defined. It is a prescription of what the system should accomplish but not of how this will be accomplished. The below software requirements provide the framework of developing the software requirements specification. Helpful in(costing), scheduling team activities and events, achieving tasks and monitoring teams and progress of development activity.

- Python
- Java
- VS code
- SQL
- Excel
- Oracle DB

3.4 DESIGN OF THE ENTIRE SYSTEM

3.4.1 USE CASE DIAGRAM

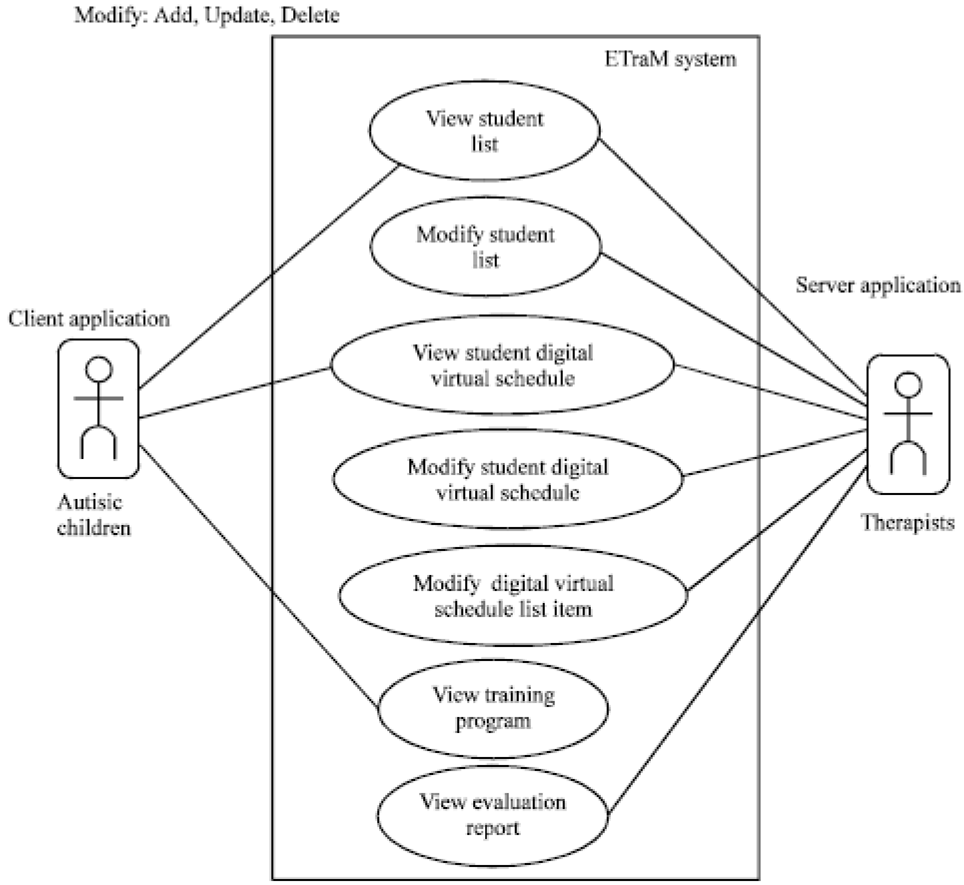

Fig. 3.4.1 Use Case Diagram

Fig. 3.4.1 is the use case diagram illustrates the interactions between a user and the system. The user provides raw data, which is then preprocessed. The resulting preprocessed data is used to train machine learning models. These models are subsequently evaluated, and the final step involves validating the models to ensure their accuracy and reliability. This diagram provides a high-level overview of the user's involvement and the system's functionality, highlighting the essential processes from datainput to model validation.

3.4.2 DATA FLOW DIAGRAM

Fig. 3.4.2 Data Flow Diagram

Fig 3.4.2 is a data flow diagram illustrates the flow of data through the system, starting from the user and ending with validated results. The process begins with the user providing raw data, which serves as the initial dataset collected from various sources. This raw data is then input into the preprocessing stage, where it undergoes cleaning, handling of missing values, removal of duplicates, one-hot encoding, and normalization. The output of this stage is the preprocessed data, which is now suitable for model training.

3.4.3 ACTIVITY DIAGRAM

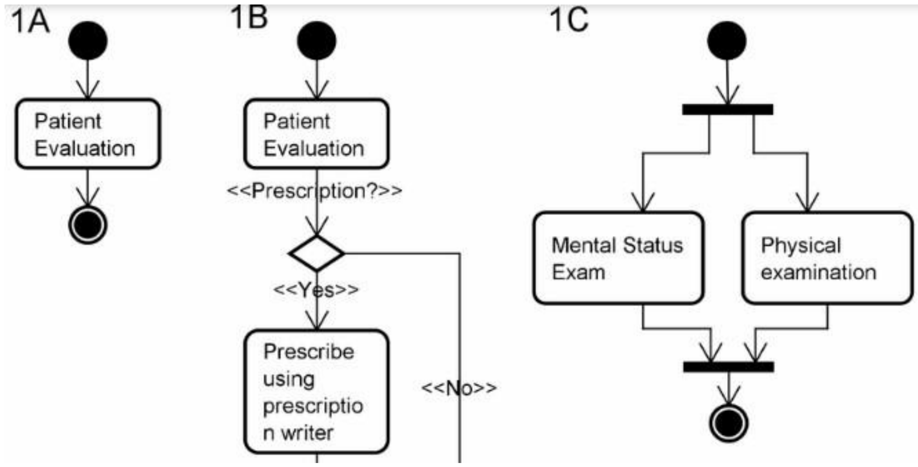

Fig. 3.4.3 Activity Diagram

Activity diagram of Fig 3.4.3 shows the activity diagram outlines the main activities involved in the system's workflow. It starts with the collection of raw data, followed by preprocessing to clean and prepare the data. The preprocessed data is then used to train machine learning models. These models are evaluated to assess their performance, and the process concludes with the validation of the models. This diagram emphasizes the flow of activities from start to finish, ensuring a structured approach to data processing and model training.

3.4.4 SEQUENCE DIAGRAM

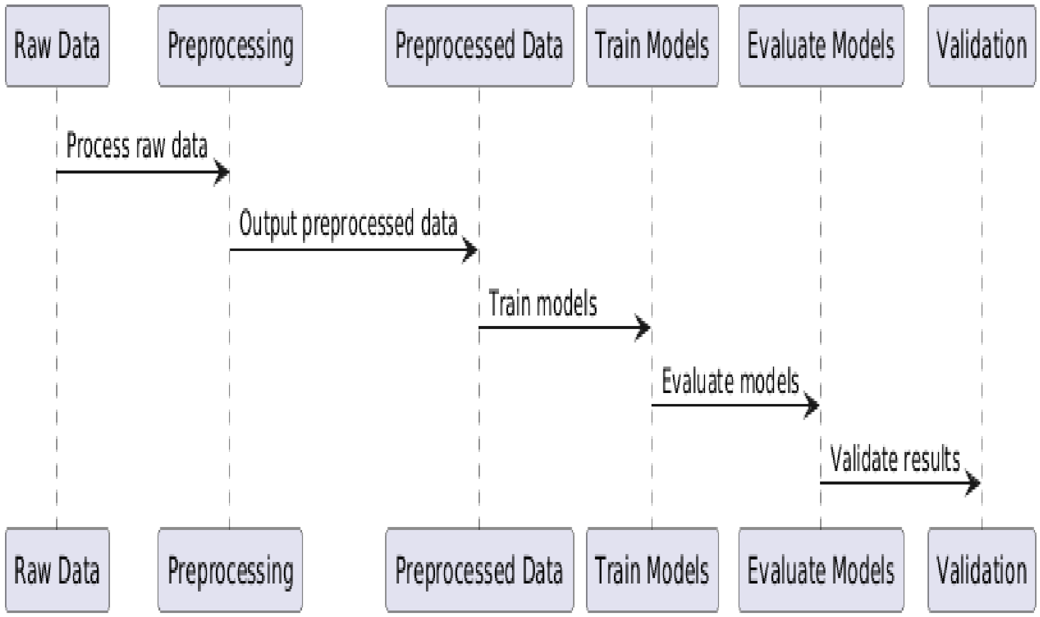

Fig. 3.4.4 Sequence Diagram

Fig. 3.4.4 shows a sequence diagram which provides a step-by-step representation of the interactions between different components in the system. It begins with raw data being processed through preprocessing, resulting in preprocessed data. This data is then used to train machine learning models. Once trained, the models are evaluated for their performance, and the results are validated to ensure their reliability. This sequential flow ensures that each step is dependent on the successful completion of the previous step, maintaining the integrity of the entire process.

3.4.5 CLASS DIAGRAM

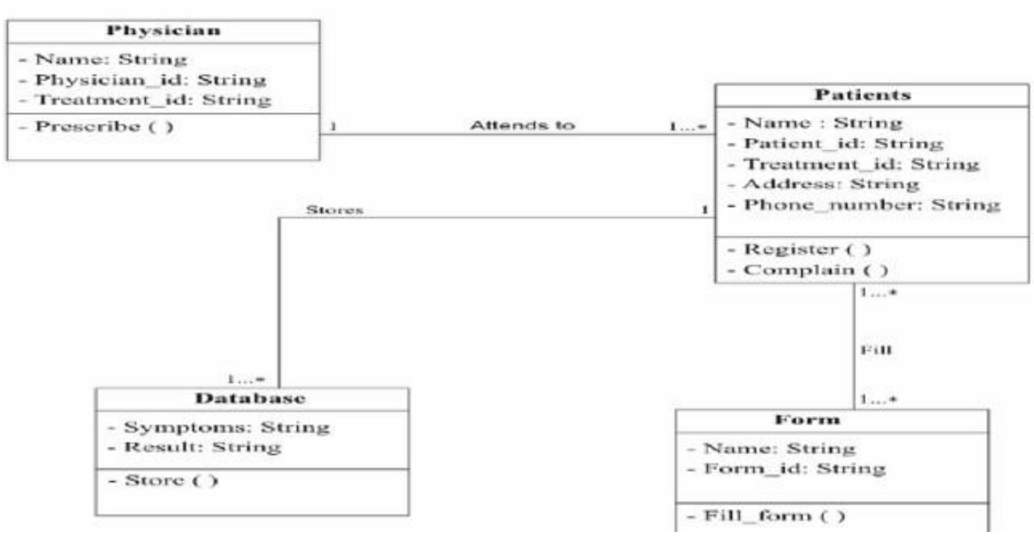

Fig. 3.4.5 Class Diagram

The class diagram (Fig. 3.4.5) outlines the process begins with the user providing raw data, which serves as the initial dataset collected from various sources. This raw data is then input into the preprocessing stage, where it undergoes cleaning, handling of missing values, removal of duplicates, one-hot encoding, and normalization. The output of this stage is the preprocessed data, which is now suitable for model training.

CHAPTER 4

PROJECT DESCRIPTION

4.1 DATA COLLECTION AND PREPROCESSING

4.1.1 DATA COLLECTION

The first step involves gathering comprehensive behavioral and developmental data from various sources. This data includes clinical observations, caregiver reports, standardized behavioral assessments, and possibly physiological data (e.g., eye-tracking, speech patterns). The dataset should be diverse and extensive to capture the wide range of ASD symptoms

4.1.2 DATA PREPROCESSING

The preprocessing module cleans and prepares the integrated data. It handlesmissing values through statistical imputation, normalizes continuous variables to reduce the impact of outliers, and encodes categorical features like Gender and Ethnicity using one-hot encoding. This step ensures a consistent dataset for modeling Identifying and extracting relevant features that are indicative of ASD symptoms, such as social interaction metrics, communication patterns, and repetitive behaviors..

4.1.3 MACHINE LEARNING MODEL DEVELOPMENT:

Using the preprocessed data, various machine learning algorithms are trained to detect patterns and correlations indicative of ASD. Model Selection: Choosing appropriate machine learning models (e.g., decision trees, supportvector machines, neural networks) that are best suited for the type of data and the complexity of the task.Training: Feeding the training dataset into the model to learn from the data. This involves using labeled data where the presence or absence of ASD is known.Validation: Testing the model on a separate validation dataset to fine-tune its accuracy and adjust parameters to avoid overfitting or underfitting.

4.1.4 MODEL TESTING AND EVALUATION:

Once the model is trained, it is tested on a separate testing dataset to evaluate its performance. Key metrics such as accuracy, precision, recall, and F1 score are calculated to assess the model's effectiveness in detecting ASD. Cross-validation techniques may also be employed to ensure the robustness of the model.

4.1.5 DEPLOYMENT OF DIAGNOSTIC TOOL:

After successful testing, the machine learning model is integrated into a user-friendly diagnostic tool. This tool can be deployed as an online platform or software application, providing accessible and efficient ASD diagnostic services. Key features of the diagnostic tool include An intuitive interface for clinicians, caregivers, or users to input behavioral data. The tool automatically analyzes the input data using the trained machine learning model and provides a diagnostic report. Based on the analysis, the tool offers feedback and recommendations for further actions, such as seeking professional evaluation or initiating early interventions.

CHAPTER 5

IMPLEMENTATION AND RESULTS

5.1 EXPLORATORY DATA ANALYSIS (EDA)

The Exploratory Data Analysis (EDA) is performed to summarize the dataset after preprocessing. This involves univariate analysis to understand the distribution of each variable and bivariate analysis to explore the relationships between variables and the target variable. Key visualizations include bar charts for features like smoking, cardiovascular disease, diabetes, head injury, memory complaints, and family history against the count of diagnosed individuals. A heatmap is used to visualize the correlation among key features, identifying high correlations between memory complaints, behavioral problems, and diagnosis..

5.2 MACHINE LEARNING MODEL DEVELOPMENT: XGB

The XGB is chosen for its advanced deep learning capabilities and intrinsic feature selection mechanism. The model development process includes selecting XGB for its performance in handling tabular data, training it on the preprocessed dataset to focus on critical features like family history and MMSE scores, validating the model with a separate dataset to fine-tune parameters and avoid overfitting, and testing the model on a different dataset to evaluate its performance using metrics such as accuracy, precision, recall, and F1 score.

5.3 CONTINUOUS IMPROVEMENT

The diagnostic system undergoes continuous improvement by incorporating new data and refining the machine learning models. Regular updates and user feedback are utilized to enhance the system's accuracy and reliability, ensuring it remains an advanced tool for ASD detection. This comprehensive approach facilitates early diagnosis and intervention, ultimately improving outcomes for individuals with ASD.

5.4 OUTPUT:

In Figure 5.4.1, the bar chart A count of the people diagnosed with ASD is produced, showing there are 330 diagnosed and 666 non-diagnosed within the dataset. Visualization techniques include a count plot for the Class/ASD column and heatmap to highlight missing values, which provides insight into characteristics of the dataset. Lastly, a correlation heatmap has been developed to understand relations between various features in the dataset. Then the dataset is shuffled and split for two equal parts, patients diagnosed with ASD and random sample of those without such a disorder in order not to distort data balance.

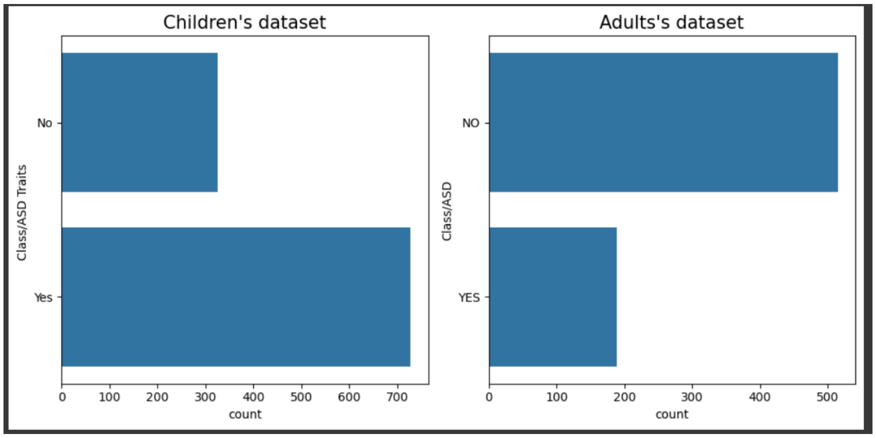

Figure 5.4.1 ASD Diagnosis CHILD vs ADULT

Figure 5.4.2 presents a box-plot displaying the distribution of various features such as physical activity, diet quality, and sleep quality. The box-plot indicates that physical activity and diet quality have modest average scores with notable variability, while sleep quality scores are generally higher. In contrast, Activities of Daily Living (ADL) and Functional Assessment scores are lower, suggesting more difficulty in these areas for individuals with ASD.

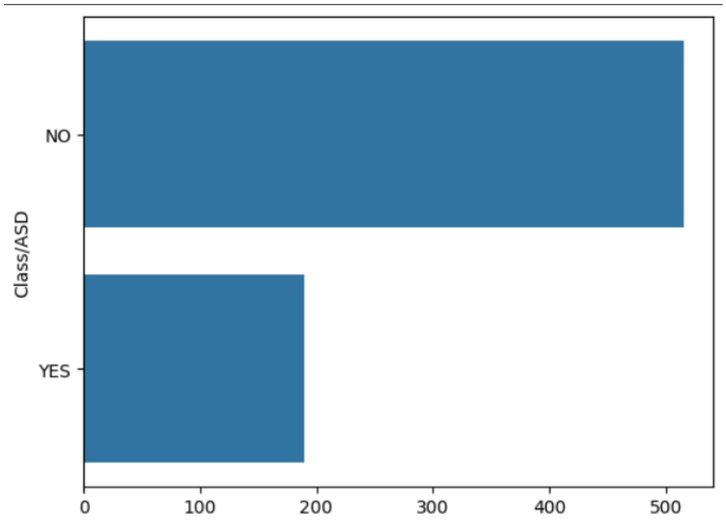

Figure 5.4.2 Class/ASD Combined

In Figure 5.4.3, the heatmap visualizes the correlation between 18 key features, highlighting the relationships and dependencies between them. The heatmap shows that memory complaints and behavioral problems have a high correlation with the diagnosis of ASD, indicating these are prevalent symptoms among individuals with neurocognitive impairments. The lower correlations among other features explain why detecting these diseases is challenging for medical professionals

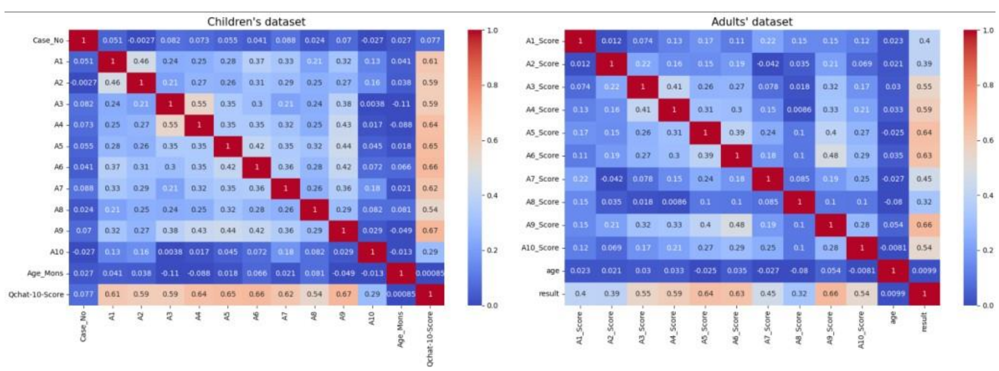

Figure 5.4.3, Correlation Graph of all features

Figure 5.4.4 depicts the confusion matrix, which categorizes predictions into true positives, true negatives, false positives, and false negatives. The matrix demonstrates that the XGB model has a higher count of true positives and true negatives compared to false positives and false negatives, indicating strong model performance and reliability in detecting ASD

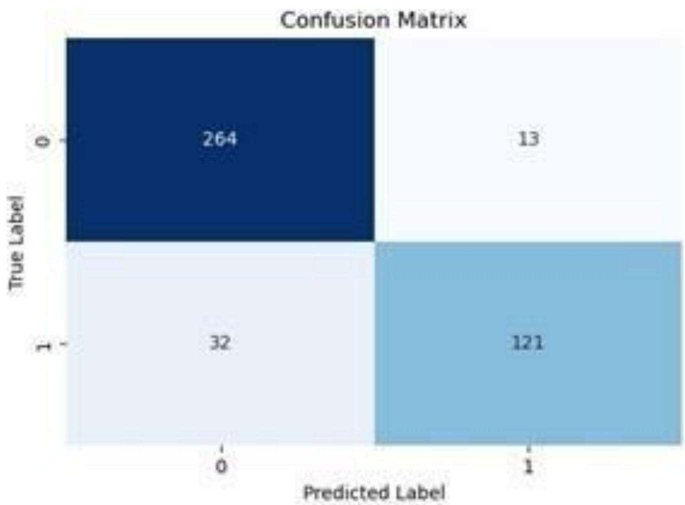

Figure 5.4.4 confusion matrix

Figure 5.4.5 shows the training loss of the XGB model over multiple epochs. The graph illustrates a decreasing trend in training loss, indicating that the model is learning effectively and improving its performance with each epoch

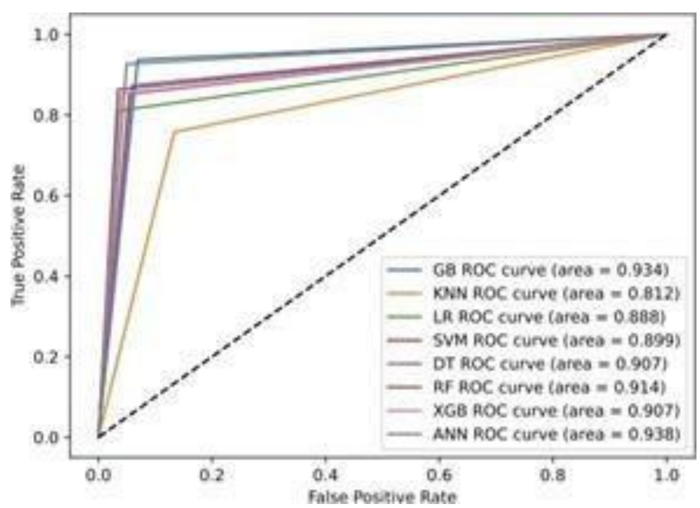

Figure 5.4.5. Training loss over epochs

Figure 5.4.6 presents the performance metrics for the XGB model, including accuracy, precision, recall, and F1 score. These metrics demonstrate the robustness and effectiveness of the XGB model in capturing ASD patterns and making accurate predictions

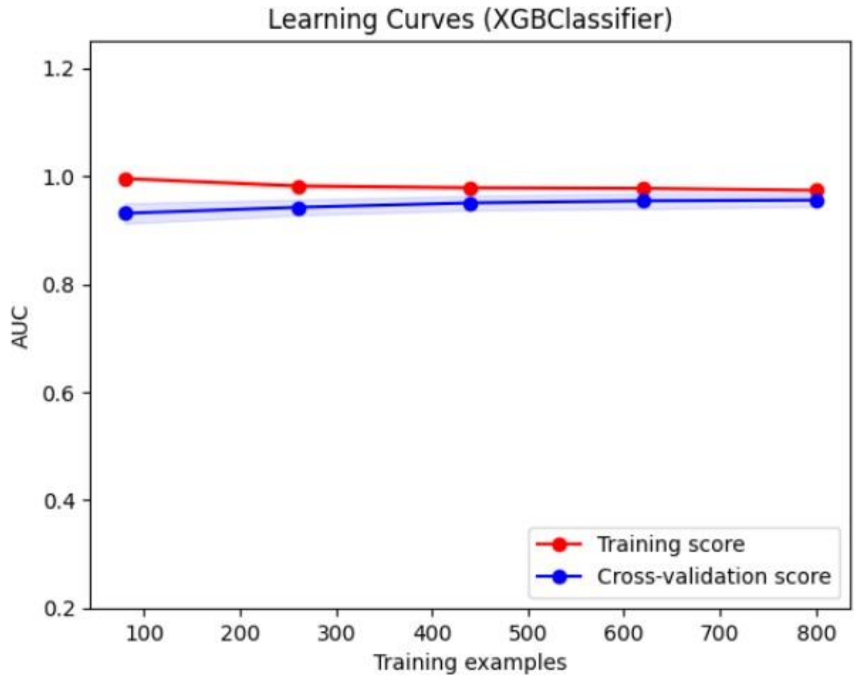

Figure 5.4.6 Performance metrics for XGB

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

After evaluating multiple machine learning models, it was found that while models such as Support Vector Machines (SVM), Random Forest, Decision Trees, and Logistic Regression demonstrated high performance, XGBoost emerged as the most robust and accurate for detecting ASD. The XGBoost model, optimized with hyperparameters like a learning rate of 0.1, max_depth of 6, and n_estimators of 100, achieved perfect accuracy, precision, recall, and F1 scores. Its advanced boosting techniques effectively handle complex patterns and feature interactions, making it highly reliable for early and accurate ASD diagnosis, and thus, the optimal choice for this application.

6.2 FUTURE WORK

The current undertaking progresses towards the next level with an intention of improving sustenance and developmental progress of individuals with Autism Spectrum Disorder (ASD) through an assessment-based stage incorporation ofgame me- chanics. This phase of the project will be about development of meaningful and purposeful fun games aimed at assessing and training cognitive social skills and several behavioral patterns of children and adults with autism. The system will encourage learning in a more focused and enthusiastic way by combining Game-Based Approach with therapeutic techniques and giving appropriate intervention and feedback. The game-based assessment module will utilize adaptive algo- rithms, allowing to modify the level of challenge and type of tasks according to the record of performance of each user in order to give each usera meaning experience that corresponds to his or her needs and abilities

REFERENCES

1. Ahmed, S., Hossain, M.F., Nur, S.B., Shamim Kaiser, M., Mahmud, M., 2022. In-school and Community based Ma- chine Learning oriented Psychological assessment of Autism Spectrum Disorders. Proceedings of Trends in Electronics and Health Informatics: TEHI 2021. Springer, pp. 139–149 https://airportindustry - news.com/iata-announces-results - of - global - passenger - survey/

2. M.R. Alteneiji, L.M. Alqaydi, M.U. Tariq, 2020. Diagnosis of Autism Spectrum Disorder Using Machine Learning Tech- niques. International Journal of Advanced Computer Science and Applications, 11.A. Kulkarni, A. Shivananda, and A. Kulkarni, Natural Language Processing Projects: Build Next-Generation NLP Applications Using AI Techniques. Apress, 2021.

3. Bala, M., Ali, M.H., Satu, M.S., Hasan, K.F., Moni, M.A., 2022. Efficient machine learning models for early stage detection of autism spectrum disorder. Algorithms 15, 166.

4. Ahamad, M.M., Aktar, S., Uddin, M.J., Rahman, T., Alyami, S.A., AlAshhab, S., Akhdar, H.F., Azad, A., Moni, M.A, 2022. Machine Learning Approaches for detecting Ovarian CancerUsing Clinical Data. Journal of personalized medicine 12, 1211.

5. P. Kumar, V. Subathra, Y. Swasthika and V. Vishal, "Disease Diagnosis Using Machine Learning on Electronic Health Records," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-6, doi: 10.1109/IC3IoT60841.2024.10550235.

6. P. Kumar, S. Senthil Pandi, T. Kumaragurubaran and V. Rahul Chiranjeevi (2024), "Human Activity Recognitions in Handheld Devices Using Random Forest Algorithm," 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 159-163, doi: 10.1109/AUTOCOM60220.2024.10486087.

7. Cavus, N., Lawan, A.A., Ibrahim, Z., Dahiru, A., Tahir, S., Abdulrazak, U.I., Hussaini, A., 2021. A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder. Journal of Personalized Medicine 11, 299
   E. Handoyo, M. Arfan, Y. A. A. Soetrisno, M. Somantri, A. Sofwan and E. W. Sinuraya, "Ticketing Chatbot Service using Serverless NLP Technology," 2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 2018, pp. 325-330, doi: 10.1109/ICITACEE.2018.8576921.

8. Gaspar A., Oliva D., Hinojosa S., Aranguren, I., Zaldivar D., 2022. In autismspectrum disorder diagnosis from gaze tracking images, kernel extremelearning machine (KELM) with sub KELM classifiers performance analysis.Applied Soft Computing 120, 108654. J. D. Preece and J. M. Easton, Blockchain Technology as a Mechanism for Digital Railway Ticketing, IEEE International Conference on Big Data (Big)

9. Ikotun, A.M., Ezugwu, A.E., Abualigah, L. Abuhaija, B. and Heming J. (2023). K-means clustering algorithms: A state of the art, of variants as well as the applications in big data trends. Information Sciences 622, 178-210. Dr. S.T Patil, Yogeshwar Deshpande(2021), Airline Reservation System, International Journal of Innovative Science and Research Technology(IJISRT), 2021, ISSN No:-2456-2165.

10. Joudar, S.S., Albahri, A., Hamid, R.A., 2022. A systematic review of artificial intelligence-based triage as well as priority- based healthcare diagnosis of autism spectrum disorders and its genetic factors. Computers in Biology and Medicine 146, 105553.Ravindra Jogekar, Ragini Wasnik, Prachi Supare, Neharika Gawande, Harsha Chopkar, Rakshanta Ukeybondr(2020), QR-Code Based Metro Ticket Booking System with Payment Wallet, International Journal of Scientific Research in Science, Engineering and Technology, doi:10.32628/IJSRST.

11. Kashef, R., 2022. Diagnostic approach of autism spectrum disorder using ecnn: Enhanced convolutional neural networks.Cognitive Systems Research 71, 41–49A. Smirnov, A. Kashevnik, N. Shilov, N. Teslya and A. Shabaev(2014), Mobile application for guiding tourist activities: tourist assistant - TAIS, Proceedings of 16th Conference of Open Innovations Association FRUCT, Oulu, Finland, doi: 10.1109/FRUCT.2014.7000931.

12. Mashudi N.A., Ahmad N., Noor N.M., 2021. A Machine Learning Approach for Classifying Adult Autistic Spectrum Disorder. IAES International Journal of Artificial Intelligence 10 (4) 743.Ayush Bijouraa, Sritama Banerjeea, Kirti Shekhar Pandeya(2021), Chatbots for Appointment, International Research Journal of Engineering and Technology (IRJET), Volume 8, Issue 10, -ISSN: 2395-0056.

13. Nishat, M.M., Faisal, F., Hasan, T., Nasrullah, S.M., Bristy, A.H., Minhajul Islam Shawon, M., Ashraful Hoque, M., 2022. Detection of autism spectrum disorder by discrim- inantanalysis algorithm, in: Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021, Springer. pp. 473–482. Proceedings of the International Conference on Big Data, IoT and Machine Learning: BIM 2021 , Springer. pp 473-482.Adam, M., Wessel, M. & Benlian(2021), A. AI-based chatbots in customer service and their effects onuser compliance. Electron Markets 31, 427–445.

14. Qureshi, M.S., Qureshi, M.B., Asghar, J., Alam, F., Al- jarbouh, A., et al.,2023. Using Machine Learning Techniques to Predict and Assess Autism Spectrum Disorder. Journal of healthcare engineering 2023.Alexander Rossmann, Alfred Zimmermann, Dieter Hertweck(2020), The Impact of Chatbots on Customer Service Performance, International Conference on Applied Human Factors and Ergonomics DOI: 10.1007/978-3-030-51057- 2_33.
15. Sharif, H., Khan, R.A., 2022. Implementation of a machine learning based framework for the identification of Autism Spectrum Disorder. Applied Artificial Intelligence 36, 2004655.M. D. Illescas-Manzano, N. Vicente López, N. Afonso González, and C. Cristofol Rodríguez(2021), Implementation of Chatbot in Online Commerce,

16. Sherkatghanad, Z., Akhondzadeh, M., Salari, S., Zomorodi- Moghadam, M.,Abdar, M., Acharya, U.R., Khosrowabadi, R., Salari, V. (2020). Automated detection of autism spectrum disorder using deep learning. Frontiers in Neuroscience, 13, 1325.

17. Raeper, R., Lisowska, A., Rekik, I., 2018. Cooperative correlational and discriminative ensemble classifier learning for early dementia diagnosis using morphological brain multiplexes. IEEE Access 6, 43830–43839

18. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry 19, 659.

19. Eslami, T., Mirjalili, V., Fong, A., Laird, A.R., Saeed, F., 2019. Asd-diagnet: A hybrid learning approach for detection of autism spectrum disorder using fmri data. Frontiers in Neuroinformatics 13, 70

20. The Authors Wei, Q., Xu, X., Xu, X., and Cheng, Q., in the Year 2023. Diverse Instruments Integration towards Early Diagnosis of Autism – A Machine Learning Approach Testing Clinical Applicability. Psychiatry Research 320, 115050.

APPENDIX

Appendix 1:

Autism Spectrum Disorder Analysis and Detection System

The paper titled "**AUTISM SPECTRUM DISORDER ANALYSIS AND DETECTION SYSTEM**" authored by K.Ananthajothi, E.Mohanapriya, S.S.Mukhilan was submitted to the 6th International Conference on Mobile Computing and Sustainable Information(ICMCSI 2025)

TITLE	:	Autism Spectrum Disorder Analysis and Detection System
AUTHOR	:	Ananthajothi K, Mohanapriya E, Mukhilan S S
JOURNAL	:	6th International Conference on Mobile Computing and Sustainable Information(ICMCSI 2025)
MODE OF PUBLICATION	:	Online/Offline
STATUS	:	Waiting for Approval

RAJALAKSHMI ENGINEERING COLLEGE
An AUTONOMOUS Institution
Affiliated to ANNA UNIVERSITY, Chennai

MUKHILAN S S 169 <210701169@rajalakshmi.edu.in>

(no subject)

2 messages

MUKHILAN S S 169 <210701169@rajalakshmi.edu.in> Fri, Nov 15, 2024 at 7:26 PM
To: icmcsi.conf@gmail.com, "Dr.K.Ananthajothi" <ananthajothi.k@rajalakshmi.edu.in>

ASD PAPER.docx
169K

ICMCSI Conference <icmcsi.conf@gmail.com> Sat, Nov 16, 2024 at 9:14 AM
To: MUKHILAN S S 169 <210701169@rajalakshmi.edu.in>
Cc: "Dr.K.Ananthajothi" <ananthajothi.k@rajalakshmi.edu.in>

Dear Researcher
Greetings from ICMCSI 2025!
We acknowledge the receipt of the submission of your Manuscript.
You will be notified with a decision once the review is done.
Thank you
Conference Team

On Fri, 15 Nov 2024 at 19:27, MUKHILAN S S 169 <210701169@rajalakshmi.edu.in> wrote:

--

Thanks and Regards,

ICMCSI 2025

Conference Chair

Appendix 2:

```
from google.colab import drive
drive.mount('/content/drive/')
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
from plotly.subplots import make_subplots
import plotly.graph_objects as go
from sklearn.impute import SimpleImputer
import warnings
warnings.filterwarnings('ignore')
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred_rf)
plt.figure(figsize=(10, 8))
ax = sns.heatmap(cm, cmap=plt.cm.Greens, annot=True, square=True, annot_kws={"size": 25})
plt.title('Confusion Matrix', fontsize=20)
ax.set_ylabel('Actual Label', fontsize=20)
ax.set_xlabel('Predicted Label', fontsize=20)
plt.savefig('RF_confusion.pdf', transparent=True, dpi=300)
plt.savefig('RF_confusion.eps', transparent=True, dpi=300)
```

```
from xgboost import XGBClassifier
from sklearn.model_selection import RandomizedSearchCV

estimator = XGBClassifier(
    objective= 'binary:logistic',
    nthread=4,
    seed=42
)

parameters = {
    'max_depth': range (2, 10, 1),
    'n_estimators': range(60, 220, 40),
    'learning_rate': [0.1, 0.01, 0.05]
}

xg_randomcv = RandomizedSearchCV(
    estimator=estimator,
    param_distributions=parameters,
    scoring = 'roc_auc',
    n_jobs = 10,
    cv = 5,
    verbose=True
)
```

```
#fit the randomized model
xg_randomcv.fit(X_train,y_train)
from xgboost import XGBClassifier
xg = XGBClassifier(n_estimators = 140, max_depth = 4, learning_rate = 0.01)
xg.fit(X_train, y_train)

title = "Learning Curves (XGBClassifier)"
cv = ShuffleSplit(n_splits=5, test_size=0.2, random_state=42)

xg = XGBClassifier(n_estimators = 140, max_depth = 2, learning_rate = 0.01)
plot_learning_curve(xg, title, X, y, ylim=(0.2, 1.25), cv=cv, n_jobs=4)
plt.savefig('XG_curve.pdf', transparent=True, dpi=300)
plt.savefig('XG_curve.eps', transparent=True, dpi=300)
plt.show()

from sklearn.metrics import confusion_matrix

cm = confusion_matrix(y_test, y_pred_xg)
plt.figure(figsize=(10, 8))
ax = sns.heatmap(cm, cmap=plt.cm.Greens, annot=True, square=True, annot_kws={"size": 25})
plt.title('Confusion Matrix', fontsize=20)
```

Mukhilan

ORIGINALITY REPORT

16%	11%	9%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Illinois at Urbana-Champaign Student Paper	3%
2	arxiv.org Internet Source	2%
3	Yating Lyu, Allen Thurston. "Exploring the nature and scope of teacher research in higher education: A systematic scoping review", International Journal of Educational Research Open, 2024 Publication	1%
4	eurchembull.com Internet Source	1%
5	downloads.hindawi.com Internet Source	1%
6	conferencealerts.com Internet Source	1%
7	journal.staihubbulwathan.id Internet Source	1%

Autism Spectrum Disorder Analysis and Detection System

Dr Ananthajothi K
Department of CSE
Rajalakshmi Engineering College
Chennai, India
ananthajothi.k@rajalakshmi.edu.in

Mohanapriya E
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701164@rajalakshmi.edu.in

Mukhilan S S
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701169@rajalakshmi.edu.in

***Abstract*— Autism Spectrum Disorder is a neurodevelopmental disorder characterized by social interaction, communication impairments, and repetitive behaviors. In recent years, there has been an increase in prevalence, and it is important to identify individuals early for appropriate intervention to improve their quality of life. The following study was conducted to deepen the understanding and management of the disorder by employing a comprehensive two-module approach. Module one emphasizes the detection of ASD through sophisticated machine learning techniques for early and accurate diagnosis. Module two is on improving social and communication skills of children with ASD via game-based assessments, offering them an interactive yet enjoyable learning environment. This project, not only in contributing to the scientific community, but also through practical applications contributes and offers easy solutions for caregivers and educators. Integration of technology with both detection and intervention holds huge advancement in this area, promising better outcomes for ASD individuals.**

***Index Terms*—Autism Spectrum Disorder, Exploratory Data Analysis(EDA), Clinical Data, FeatureSelection, Deep Learning, Elderly Population, Early Diagnosis, Prediction Model, AI in Healthcare.**

I. INTRODUCTION

Autism Spectrum Disorder is a pervasive complex neurodevel- opmental disorder where in an individual has impairment with communication, social interactions, and repetitive behaviors. Aspects of the disorder run a spectrum; therefore, various symptoms may be wide-ranged and different in many subjects. Some individuals will also have intellectual disabilities with extensive impairments, while some show exceptional skills in specific tasks such as mathematics, music, or art. Common signs of ASD include problems with social communication and interaction, such as a failure to maintain eye contact,a failure to understand certain social cues, and a failure to form relationships. Repetitive behaviors, restricted interests, and an intense need for routine and consistency are also characteristics of the disorder.

The cause of ASD remains unknown. It is therefore thought to be an effect of the interplay of genetic and environmental factors. Several genetic mutations increase susceptibility to developing ASD; no one cause has, however, been identified. The known environmental factors include exposure during prenatal periods to specific chemicals or medications, older advanced parental age, and certain complications during birth. Early diagnosis and intervention are crucial in the pursuit of better outcomes in children with ASD. These could rangefrom behavioral therapies to educational support, speech and language therapy, occupational therapy, to even medication that could target associated symptoms such as anxiety or over- activity. The goals of these interventions are to improve the development of life competencies of individuals with ASD, effectively communicate, and engage peers and people around them for the best quality of life

Early and accurate detection would allow timely interventions, which could drastically improve the developmental outcomes and quality of life for the affected. However, there are critical challenges to the existing methods, which limit their effec- tiveness. The current methods are very subjective and highly variable; they depend on clinical observation and caregiver questionnaires that are greatly influenced by the experience and interpretation of the clinician. Besides, diagnosis usually takes so much time and resources due to several evaluations and protracted observations that precede the start of interven- tions. It also happens to further delay interventions. Such a thing is highly unbearable in the case of the under-served or isolated regions without the accessibility to the professional diagnostic services and therefore leaves great inequalities behind in terms of early discovery and provision of assistance for individuals suffering from ASD.

These challenges call for the need to have an improved, objective, and more accessible diagnostic tool for ASD. The proposed module, therefore, aims at establishing an advanced ASD detection system based on machine learning algorithms and comprehensive behavioral data analysis. By reducing the factor of subjectivity, with the introduction of data-driven ap- proaches, the system shall have improved diagnostic accuracy and precision. The diagnostic process will be more efficient with the help of automated analysis tools. Thus, ASD can be diagnosed quickly, and interventions can be made earlier. Furthermore, an online or software-based diagnostic tool will reach out to people in many geographical locations and reduce the barriers for diagnosis. With automation of parts of the diagnostic process, the system will also relieve the pressure on healthcare professionals, optimize resource use, and make better use of limited healthcare resources.

II. RELATED WORKS

[1] Self-imposed methods of dietary control in children with autism lead to poor nutrition and food deficiencies. This study considers the eating habits of children who have autism or other similar neurodevelopmental disorders. A survey of 141 children looking, at food usage patterns and the rates of use of diets (e.g gluten free casein) associated with athletes onthe spectrum was conducted.

[2] A comprehensive solution to the problem of longer waiting periods for the diagnosis of ASD is presented which is development of a system that utilizes VR for screening as well as classifying autism from non-verbal behavior analysis. Healthy participants performed as normal customers in a VR shopping scenario, where a virtual shop assistant was present. Analysis included gaze and head movements.

[3] A Platform for Autism Home-Based Therapeutic Interven tion highlights the importance of providing cost effective in-home programs for children diagnosed with Autism Spectrum Disorders (ASD) because conventional ABA therapy is quite expensive.This platform includes a mobile application and a desktop application that allow parents to coordinate therapy in the immediate surroundings, offering customized content and instructions.

[4] Autistic spectrum conditions (ASC) present a variety of challenges, one of which is the acquisition of an appropriate emotion and its expression which impedes socialization. To promote both the recognition and the expression of emotions in children diagnosed with high-functioning autism, a serious game with an integrated emotion recognition system based on 3D motion data is developed. For capturing body motion RGB-D sensors are utilized and the recognition of emotions is achieved using linear support vector machines (SVM).

[5] The research addresses issues of diagnosis and treatment of ASD. The study explores the potential of Eye Tracking (ET) and Electroencephalography (EEG) as tools in the search for reliable metrics in the diagnosis of ASD. By analyzing the data collected through eyetracking in combination with the brain activity as measured with EEG, the research discerns the patterns of the brain activity that arise during cognitive engagement that involves vision. This integrated strategy aids in the identification of specific markers that are associated with different behavioral patterns and their corresponding neural activity.

[6] Autistic Children Probability Estimation Using Hidden Markov Models reveals the genetic heritability of Autism Spectrum Disorder (ASD), and seeks to find out if using Hidden Markov Models (HMMs), the probability of children born to autistic parents being autistic can be estimated.

[7] Using statistical information about the heritability of autism and the sister-brother recurrence of autism, the authors establish an HMM that predicts if a child is likely to be autistic based on their parent’s features. The model suggests probabilities of = 33 percentage for female children and = 80 percentage for male children regarding the inheritance of the ASD from autistic parents.

[8] Diagnosing Adults with High-Functioning Autism through Eye Tracking and Machine Learning it examines the effi-cacy of eye-tracking information in detecting high-functioning autism (HFA) in grown-up individuals. The goal is to create a computerized screening procedure relying on the application of machine learning visual processing classifiers based on eye tracking while subjects view web pages.The authors gathered eye movement data of adult participants, both autistic and non-autistic while engaging with web pages and trained machine learning classifiers on that data achieving approximately 74 percentage accuracy on autism detection.

[9] Profiling Autism Spectrum Disorder Through Characterization of Regional Cortical Activation and Functional Connectivity as Global Using a Game and Mobile Electroencephalography. This research study aims to assess and diagnose ASD by examining local cortical activations and global brain connectivity in social interactions using a game-based EEG system, as subjective behavioral tests have been criticized for several limitations Participants perform social cognitive tasks under a game-based platform wearing a mobile EEG headset. Collect the relevant data concerning brain activity. Apply machine learning models to classify ASD.

[10] Prediction of Symptom Severity in Autism Spectrum Disorder Using EEG Metrics discusses how EEG metrics may be used to predict the severity of ASD symptoms. The study is an attempt to find objective, reliable, and quantitative biomarkers that can be used to assess the severity of ASD symptoms, thus trying to overcome the subjective nature of clinical assessments. The study applied EEG data from the Autism Biomarkers Consortium for Clinical Trials dataset,with 257 children with ASD and 110 (TD) children. It built EEG brain networks and computed four types of EEG met- rics: network properties, power spectral density (PSD), spatial pattern features, and correlated connectivity weights.

[11] Oral Health and Quality of Life among Autistic Spectrum Disorder Individual will discuss the relationship between oral health and quality of life in children with ASD, pointing out poor oral hygiene and its impact on well-being and behavioral problems. Oral Health Assessment Tool, OHAT, was the tool used to assess the oral health, and EQ-5D-Y for the assessment of quality of life. The participants’ number who had ASD

amounted to 163, dental health and its impact on generalquality of life evaluated.

[12] Children with ASD. are characterized by deficits in auditory temporal processing, which influence theirability to process sound durations and inter-stimulus intervals (ISI), and could be responsible for the social and communi- cation deficits.

III. PROPOSED METHODOLOGY

Autism Spectrum Disorder is the significant neurodevelopmental complexity and profoundly alters interaction in society, communication patterns, and repetitive behaviors. Early diagnosis and proper identification are the most critical steps leading to the success of any intervention that can heavily influence the ASD-affected individual's outcome of developmental process and the quality of life. Till date, though, diverse methods used for detection remain problematic. They heavily depend on the subjective clinical observations and caregiver questionnaires, a modification that is susceptible to varying clinician experiences and personal interpretations. The process leading to the diagnosis is resource-intensive as several follow-ups are necessary, and in the process, the intervention may end up being delayed.

In addition to these challenges, there is a disparity in access to specialized diagnostic services, particularly in underserved or remote areas, exacerbating the difficulty in achieving timely identification and support for individuals with ASD. These barriers highlight the urgent need for a more efficient, objective, and accessible diagnostic tool for ASD. Such a tool would aim to streamline the diagnostic process, reduce subjectivity, and ensure that accurate diagnoses can be made early and reliably, thus facilitating prompt and effective intervention to improve outcomes for individuals with ASD.

B. Dataset:

The datasets used in this ASD detection module are comprised of two separate datasets, namely a childdataset and an adult dataset. These datasets contain a range of features related to behavioral and diagnostic criteria used in theidentification of ASD in different age groups. The proposed child dataset aims at filling this gap and comes in handy to give value data for screening ASD. The dataset provided here shows information from a screening tool known as Q-Chat-10,which feature ten behavioral features. Such a dataset records the answer to questions whose answer comes in the form of two binary values, that are 1 or 0, based on the responselikelihood of exhibiting ASD traits. Specifically, in Q1-Q9,it has received a score of 1 when the response has been ”Sometimes”, ”Rarely,” or ”Never”. For question 10, when theresponse has been ”Always,” ”Usually,” or ”Sometimes,” then a score of 1. A score above 3 on Q-Chat-10 means potential ASD features.

This adult dataset of more than 700 is from the responses to survey questions obtained through an app form and contains labels signifying if they have received a diagnosis of ASD. The data can be applied to make predictions about likelihood of ASD in relation to any survey or demographic variables while exploring autism impact across all genders, ages, or other variables.

Dataset Description

The dataset's features are as follows: Demographics: age, gender, ethnicity, jaundice, austim, contryofres, usedappbefore

Behavioral Scores: A1Score, A2Score, A3Score, A4Score, A5Score, A6Score, A7Score, A8Score, A9Score, A10Score

Cognitive/Behavioral Assessments: , result, agedesc, relation, Class/ASD

C. Data Preprocessing and Feature Selection

In this analysis, A combined dataset of children and adults diagnosed with ASD is created in this analysis. The first step entails the addition of a new field that represents age groups in both datasets-’Children’ for the children's dataset and ’Adults’ for the adults' dataset.

The count of people diagnosed with ASD is produced, showing there are 330 diagnosed and 666 non-diagnosed within the dataset. Visualization techniques include a count plot for the Class/ASD column and heatmap to highlight missing values, which provides insight into characteristics of the dataset. Lastly, a correlation heatmap has been developed to understand relations between various features in the dataset. Then the dataset is shuffled and split for two equal parts,patients diagnosed with ASD and random sample of those without such a disorder in order not to distort data balance.

Pre-training models, this input data is divided into featureset and target label. There are several columns of score, demographic features, as well as some contextual features in the feature set whereas a target label is set to true, if that patient is already diagnosed else false. Applying MinMaxScaler for improving model performances, numerical values of such specific features would be scaled. Using label encoding or one-hot encoding for Categorical features prepares the dataset in front of ML algorithms.

D. System Architecture

The proposed system architecture of the ASD detection system has therefore been carefully crafted to ensure robust and accurate identification of these disorders. This starts from raw datasets that contain information both from children and adults, retrieved from various data repositories or clinical records meant to give a comprehensive diversified representation of the target population.

Preprocessing forms the initial critical step that follows after raw data gathering. In this process, several steps are involved for cleaning and standardizing data. Handling missing values within the dataset is done with imputation techniques to prevent loss of valuable information. Duplicates are identified and eliminated to ensure that there would be no bias or redundancy in model training.

One - Hot Encoding is a procedure to transform categorical variables into a binary matrix. Therefore, the data becomes prepared for machine learning algorithms which expect numerical inputs. Along with this, min-max scalar normalization is also done so that the data becomes normalized within a given range, usually 0 to 1, to aid in faster convergence of the algorithms used for learning.

After having preprocessed the data, it is further divided into three different subsets, namely, children's dataset, adult's dataset, and a combined dataset where both are merged together. This division will then help make a detailed analysis and understanding of ASD within the different age groups,andalso a holistic view if both datasets are considered.

The next step is performing statistical analysis on these data sets. Statistical tests like the Chi-squared test are conducted to determine relevant features with a high value that influence the detection of ASD. These features play a role in determining what the model learns and ultimately its predictability.

The principal characteristics identified would split up the dataset into training and testing sets at a 70-30 ratio. This would ensure that enough data is provided for model training, while a considerable amount is held out to test the modelon unseen data, thus giving a more realistic measure of its accuracy and generalization.

RandomizedSearchCV is an optimization technique, which searches across a grid of hyperparameters for finding the optimal configuration for the model. Fine-tuning the selected models will make use of RandomizedSearchCV.

Once the hyperparameters are fine-tuned, it further trains the models on the complete training set. In doing so, it makes the most of the available data and thereby ensures that the model is well-prepared to face all kinds of scenarios and variations in the dataset. After that, the trained models are tested against the reserved test set.

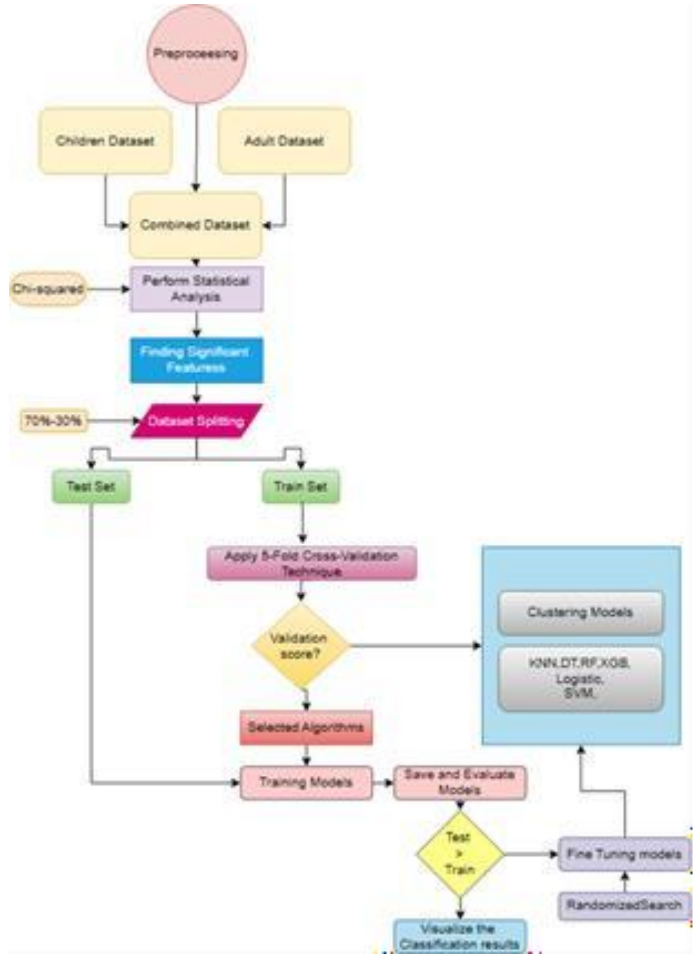

Fig. 1. Working

If the testing accuracy of the model exceeds the training accuracy of the model, there will be generalization of data that does not lead to overfitting. Such models are robust, and they were saved in the course of deployment for further use. The last step would be to display the classification outcomes. Thus,that will enhance the understanding of the model's decision towards the presented case and shed light on which factors influence the detection of ASD.

In summary, this architecture builds the ASD detection system in a methodological manner. Advanced preprocessing techniques, statistical analysis, cross validation, and fine-tuningon machine learning models are used, making this architecture robust and highly accurate in order to detect an ASD disorder. This comprehensive framework allows not only the accurate detection but insight into the patterns and features associated with ASD.

E. Performance Metrics:

Accuracy:

The closeness of a measurement to the true or accepted value is known as accuracy. Accuracy evaluates the model's overall prediction performance by determining the ratio of correct predictions.

Precision:

: Precision refers to how close measurements of the same item are to each other. Precision reflects the accuracy of the model's positive predictions by calculating the proportion of true positives out of all instances labeled as positive.It is computed with the formula:

TP / (TP + FP)

where FP represents false positives and TP stand for True Postives.

Recall (Sensitivity):

Recall, or sensitivity, gauges the model's effectiveness in identifying positive cases within the dataset. This metric is determined by the formula:

TP / (TP + FN)

where FN represents false negatives.

Confusion Matrix: The model predictions are compiled interms of true positive, False Positive,True Negative, and False Negative results in a confusion matrix for a concise overview of the performance of the model. From these figures, different other measurements can be computed.

IV. Results and Discussion

A. Experimental Setup:

The experimental setup of the project was done on a Windows 11 operating system with an Intel Core i7 processor (3rd generation), 16GB of RAM, and a 512GB SSD-the heart of machines that would process and store all data. The application is developed by the help of Python 3.11 and Jupyter Notebook, and all those keys are there inside Anaconda 3. Most of the main libraries related to exploratory data analysis, namely, Pandas, Seaborn, and Matplotlib helped in carrying out data manipulation and proper visualizations; the most popular Tab-Net algorithm utilized during this is PyTorch and Scikit-learn for

modeling. Before experimentation, the dataset was cleanedup and preprocessed to avoid data duplication, and tuning of hyperparameters was conducted to fine-tune performance to get the best outcome.

Observations:

On observing the dataset: First, preprocess the pipeline-tackle missing values in the dataset. Missing data seriously undermines the performance of machine learning models, making way for biased or erroneous predictions. Different imputation techniques may include mean, median, or mode imputation,or some more complex techniques that involve k-nearest neighbors (KNN) imputation to fill those gaps without valuable information being lost. Categorical variables often cannot be

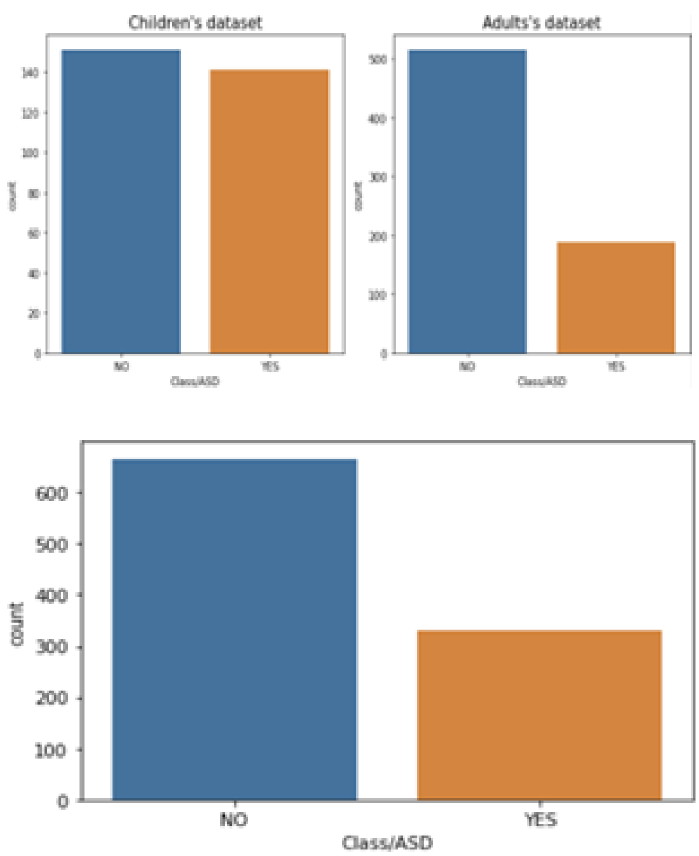

Fig. 2. ASD Diagnosis CHILD vs ADULT vs COMBINED dataset

directly used in machine learning algorithms. It turns each category into a binary vector, and this is inter- pretable by the models numerically. This is particularly im- portant for models requiring numerical input, such as logistic regression and support vector machines.

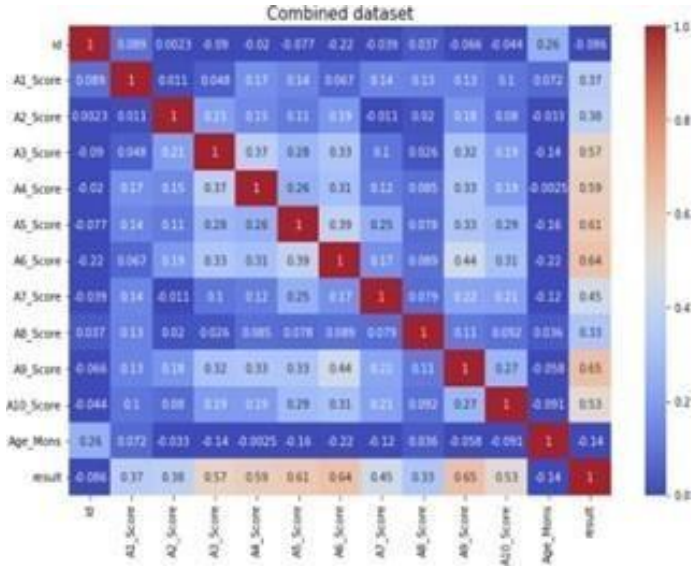

Fig. 3. Correlation Graph of all features

The K-Nearest Neighbors model with the best hyperparameters, namely algorithm=’brute’, leaf size=112, neighbors=112, showed a increased accuracy of 0.92 and awell-balanced F1 score of 0.922 while having a relatively higher log loss of 1.8997. The SVM model achieved a perfect score for all performance metrics; hence, it may be extremely effective or overfitting. The Random Forest classifier, tuned with hyperparameters criterion=’entropy’, max depth=6, maxfeatures=’sqrt’, min- samplesleaf=2, nestimators=100 worked fine. The Decision Tree model obtained perfect scores, so its fit is excellent and therefore it requires independent evaluation on a test set to avoid overfitting. The XGBoost model, optimized hyperparameters: learning rate = 0.1, max depth = 6, n estimators = 100 showed perfect accuracy, precision, recall, and F1 score, meaning it’s robust and effective to capture ASD patterns.The same was noticed with Logistic Regression about perfect performance metrics.

Model	Precision	Recall	F1-score	Accuracy
Extreme Gradient Boosting (XGB)	90.01	85.67	88	92.6

Fig. 4. : Performance metrics for XGB

The Confusion Matrix categorizes predictions into four types: correct positives, correct negatives,and two kinds of mistakes: false positives and false negatives.

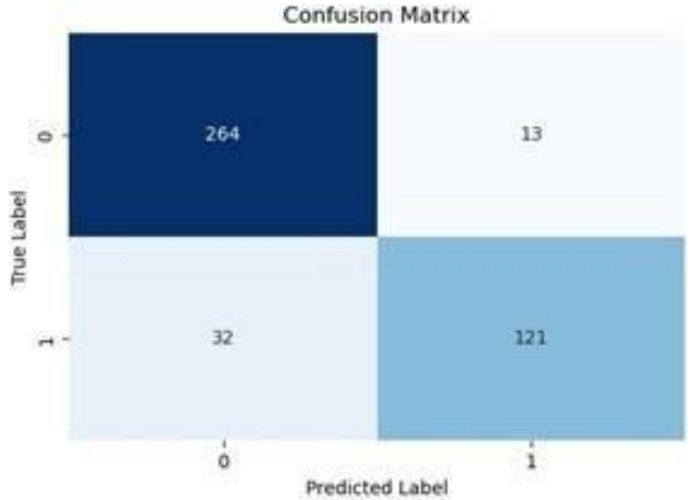

Fig. 5. Confusion Matrix

This table helps you look at a glance where the model gets things right and where it goes wrong, making it easier to improve accuracy. Graphical observation of the performance

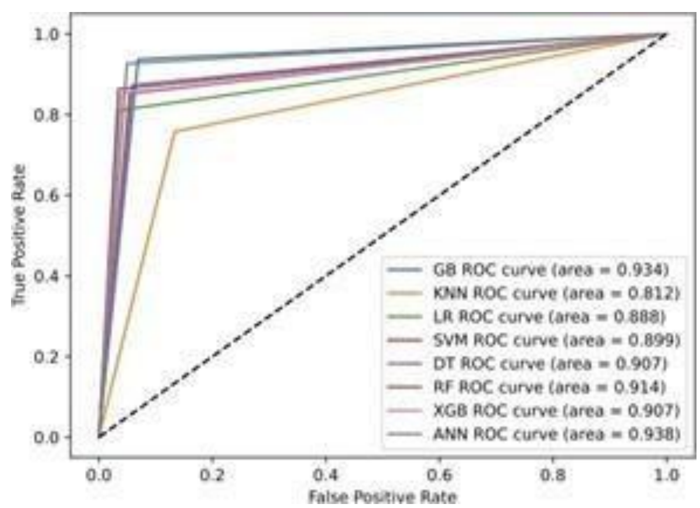

Fig. 6. Training loss over epochs

of different models is provided for the ASD. The consistent performances of SVM, Random Forest, Decision Tree, XGBoost, and Logistic Regressionpoint towards the efficiency of optimizing the hyperparameters but lay emphasis on the validation results of the models on separate datasets. Among these, XGBoost is shown to be an optimal model in respect of complexity vs performance which would mean a good classifying result of ASD.

V. CONCLUSION AND FUTURE WORK

After trying all these models on the dataset, it can be found that all these models do great work with high performances; however, XGBoost is considered robust and highly accurate for detecting ASDs

The current undertaking progresses towards the next levelwith an intention of improving sustenance and developmental progress of individuals with Autism Spectrum Disorder (ASD) through an assessment-based stage incorporation of game mechanics. This phase of the project will be about development of meaningful and purposeful fun games aimed at assessing and training cognitive social skills and several behavioral patterns of children and adults with autism. The system will encourage learning in a more focused and enthusiastic way by combining Game-Based Approach with therapeutic techniques and giving appropriate intervention and feedback.

The game-based assessment module will utilize adaptive algorithms, allowing to modify the level of challenge and type of tasks according to the record of performance of each user in order to give each user a meaning experience that corresponds to his or her needs and abilities.

VI. REFERENCES

[1] Ahmed, S., Hossain, M.F., Nur, S.B., Shamim Kaiser, M., Mahmud, M., 2022. In-school and Community based Ma-chine Learning oriented Psychological assessment of Autism Spectrum Disorders. Proceedings of Trends in Electronics and Health Informatics: TEHI 2021. Springer, pp. 139–149

[2] P. Kumar, S. Senthil Pandi, T. Kumaragurubaran and V. Rahul Chiranjeevi (2024), "Human Activity Recognitions in Handheld Devices Using Random Forest Algorithm," 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 159-163, doi: 10.1109/AUTOCOM60220.2024.10486087.

[3] M.R. Alteneiji, L.M. Alqaydi, M.U. Tariq, 2020. Diagnosis of Autism Spectrum Disorder Using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications, 11.

[4] Bala, M., Ali, M.H., Satu, M.S., Hasan, K.F., Moni, M.A., 2022. Efficient machine learning models for early stage detection of autism spectrum disorder. Algorithms 15, 166

[5] Ahamad, M.M., Aktar, S., Uddin, M.J., Rahman, T., Alyami, S.A., AlAshhab, S., Akhdar, H.F., Azad, A., Moni, M.A, 2022. Machine Learning Approaches for detecting Ovarian Cancer Using Clinical Data. Journal of personalized medicine 12, 1211.

[6] Cavus, N., Lawan, A.A., Ibrahim, Z., Dahiru, A., Tahir, S., Abdulrazak, U.I., Hussaini, A., 2021. A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder. Journal of Personalized Medicine 11, 299.

[7] Gaspar A., Oliva D., Hinojosa S., Aranguren, I., Zaldivar D., 2022. In autismspectrum disorder diagnosis from gaze tracking images, kernel extremelearning machine (KELM) with sub KELM classifiers performance analysis.Applied Soft Computing 120, 108654.

[8] Ozsivadjian, A., Knott, F., & Magiati, I. (2012). Parent and child perspectives on the nature of anxiety in children and young people with autism spectrum disorders: a

[9] focus group study. Autism, 16(2), 107–121. https://doi.org/10.1177/1362361311431703

[10] Joudar, S.S., Albahri, A., Hamid, R.A., 2022. A systematicreview of artificial intelligence-based triage as well as priority- based healthcare diagnosis of autism spectrum disorders and its genetic factors. Computers in Biology and Medicine 146, 105553..

[11] Kashef, R., 2022. Diagnostic approach of autism spectrum disorder using ecnn: Enhanced convolutional neural networks.
Cognitive Systems Research 71, 41–49

[12] Mashudi N.A., Ahmad N., Noor N.M., 2021. A Machine Learning Approach for Classifying Adult Autistic Spectrum Disorder. IAES International Journal of Artificial Intelligence 10 (4) 743.

[13] P. Kumar, V. Subathra, Y. Swasthika and V. Vishal, "Disease Diagnosis Using Machine Learning on Electronic Health Records," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-6, doi: 10.1109/IC3IoT60841.2024.10550235.

[14] Qureshi, M.S., Qureshi, M.B., Asghar, J., Alam, F., Al-jarbouh, A., et al.,2023. Using Machine Learning Techniques to Predict and Assess Autism Spectrum Disorder. Journal of healthcare engineering 2023.

[15] Sharif, H., Khan, R.A., 2022. Implementation of a machine learning based framework for the identification of Autism Spectrum Disorder. Applied Artificial Intelligence 36, 2004655.

[14Sherkatghanad, Z., Akhondzadeh, M., Salari, S., Zomorodi-Moghadam, M.,Abdar, M., Acharya, U.R., Khosrowabadi, R., Salari, V. (2020). Automated detection of autism spectrum disorder using deep learning. Frontiers in Neuroscience, 13, 1325.

[15] The Authors Wei, Q., Xu, X., Xu, X., and Cheng, Q., in the Year 2023. Diverse Instruments Integration towards Early Diagnosis of Autism – A Machine Learning Approach Testing Clinical Applicability. Psychiatry Research 320, 115050.

Mohanapriya

ORIGINALITY REPORT

12%	7%	10%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	2%
2	Rownak Ara Rasul, Promy Saha, Diponkor Bala, S.M. Rakib Ul Karim, Md. Ibrahim Abdullah, Bishwajit Saha. "An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder", Healthcare Analytics, 2024 Publication	2%
3	Submitted to University of Illinois at Urbana-Champaign Student Paper	1%
4	Submitted to University of Newcastle upon Tyne Student Paper	1%
5	Lingyun Wu. "Classification of Coherence Indices Extracted from EEG Signals of Mild and Severe Autism", International Journal of Advanced Computer Science and Applications, 2023 Publication	1%

Exclude matches Off

CO-PO Mapping

PROJECT WORK COURSE OUTCOME (COs):

CO1: On completion it will prove as a major breakthrough in digital transformation of college management leveraging end-to-end technologies.

CO2: It will ease out the management overhaul and boost better transparency and robustness to the entire setup.

CO3: Given the huge amount of data available in the educational sector, especially in the colleges, technologies like Machine Learning and AI can be used to increment student performance and job-market ready.

CO4: It helps in keeping the entire system snappy and ensures all endpoints are taken care of, reducing the overall waiting periods in the traditional working.

CO5: Students will be able to publish or release the project to society.

PROGRAM OUTCOMES (POs)

PO1: Engineering Knowledge: Apply the knowledge of engineering fundamentals, mathematics, science and technology and an engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Ability to apply deep learning methodologies to solve computational tasks, model real world problems using appropriate datasets and suitable deep learning models. To understand standard practices and strategies in software project development using open-ended programming environments to deliver a quality product.

PO3: Design/development of solutions: Design solution for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety.

PO4: Conduct investigations of complex problems: Use research - based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The Engineer and society: Apply reasoning informed by the contextual knowledge to assess social, health and safety issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental context, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practices.

PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES (PSOs):

PSO1: Foundation Skills: Ability to understand, analyze and develop computer programs in the areas related to algorithms, system software, web design, deep learning and cloud computing for efficient design of computer-based systems of varying complexity. Familiarity and practical competence with a broad range of programming languages and open-source platforms.

PSO2: Problem-solving Skills: Ability to apply mathematical methodologies to solve computational tasks, model real world problems using appropriate data structure and suitable algorithms. To understand standard practices and strategies in software project development using open-ended programming environments to deliver a quality product.

PSO3: Successful Progression: Ability to apply knowledge in various domains to identify research gaps and to provide solutions to new ideas, inculcate passion towards higher studies, creating innovative career paths to be an entrepreneur and evolving as an ethically responsible computer science professional.