Mukhil

by Ananthajothi K

AUTISIMSPECTRUM DISORDER ANALYSISAND 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 COLLEGEANNA

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

Dr. P. Kumar, Ph.D

HEAD OF THE DEPARTMENT

PROFESSOR

Department of Computer Science andEngineering

Rajalakshmi Engineering College

Chennai – 602 105

Signature

Dr. K. Ananthajothi, Ph.D

SUPERVISOR

PROFESSOR

Department of Computer Science and Engineering

Rajalakshmi Engineering College

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 Vice Chairman **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, 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 of Computer 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 210701164MUKHILAN 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 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.

TABLE OF CONTENTS

CHAPTER NO.	TOPIC	PAGE NO.
	ABSTRACT	**iv**
	LIST OF FIGURES	**vii**
1	**INTRODUCTION**	**1**
	1.1 GENERAL	1
	1.2 OBJECTIVE	2
	1.3 EXIXTING SYSTEM	2
	1.4 PROPOSED SYSTEM	2
2	**LITERATURE SURVEY**	**3**
3	**SYSTEM DESIGN**	**11**
	3.1 GENERAL	11
	3.2 SYSTEM ARCHITECTURE	11
	3.3 SYSTEM REQUIREMENTS	12
	3.3.1 HARDWARE REQUIREMENTS	12
	3.3.2 SOFTWARE REQUIREMENTS	12
	3.4 DESIGN OF THE ENTIRE SYSTEM	13
	3.4.1 USE CASE DIAGRAM	13
	3.4.2 DATA FLOW DIAGRAM	14

	3.4.3 ACTIVITY DIAGRAM	15
	3.4.4 SEQUENCE DIAGRAM	16
	3.4.5 CLASS DIAGRAM	17
4	**PROJECT DESCRIPTION**	**18**
	4.1 DATA COLLECTION AND PREPROCESSING	18
	4.1.1 DATA COLLECTION	18
	4.1.2 DATA PREPROCESSING	19
	4.1.3 MACHINE LEARNING MODEL	19
	4.1.4 MODEL TESTING AND EVALUATION	20
	4.1.5 DEPLOYMENT OF DIAGNOSTIC TOOL	20
5	**IMPLEMENTATION AND RESULTS**	**21**
	5.1 EXPLORATORY DATA ANALYSIS	21
	5.2 MACHINE LEARNING MODEL	21
	5.3 CONTINUOUS IMPROVEMENT	21
	5.4 OUTPUT	22
6	**CONCLUSIONS AND FUTURE WORK**	**25**
	6.1 CONCLUSIONS	25
	6.2 FUTURE WORK	25
7	**REFERNCES**	26
APPENDIX		29

LIST OF FIGURES

FIGURE NO.	TOPIC	PAGE NO.
3.1	SYSTEM ARCHITECTURE	10
3.2	ACTIVITY DIAGRAM	13
3.3	DATA FLOW DIAGRAM	14
3.4	USE CASE DIAGRAM	15
3.5	SEQUENCE DIAGRAM	16

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 to address 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 desktop application 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 actively participate 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 employing few-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 118 participants 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 using home 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, focusing on 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 potential to 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 a novel 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

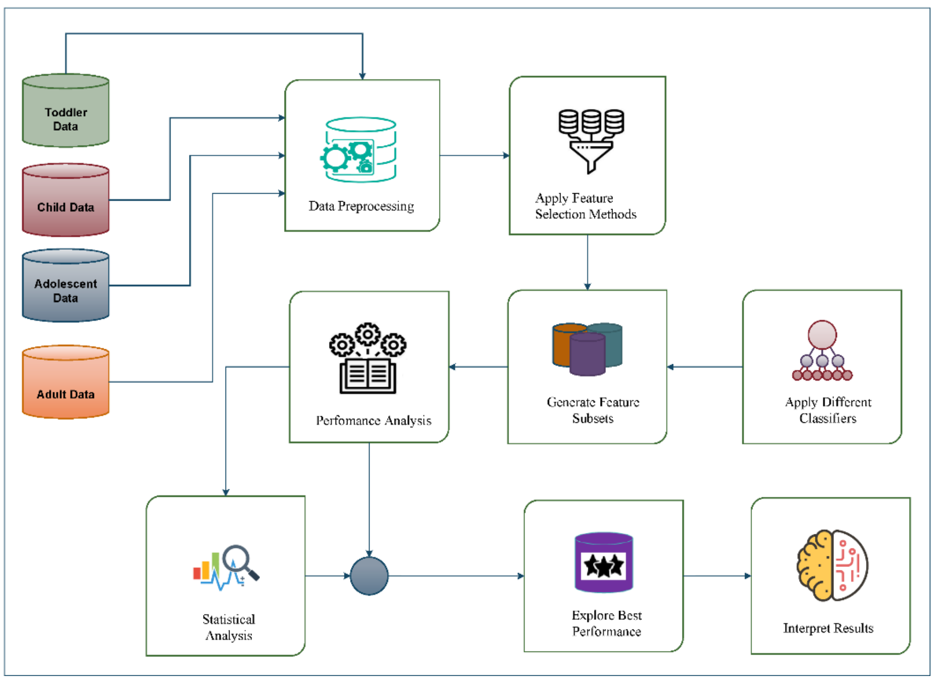

Fig. 3.1 System Architecture

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 the Model Training package, the preprocessed data is used to train various machine learning models. 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

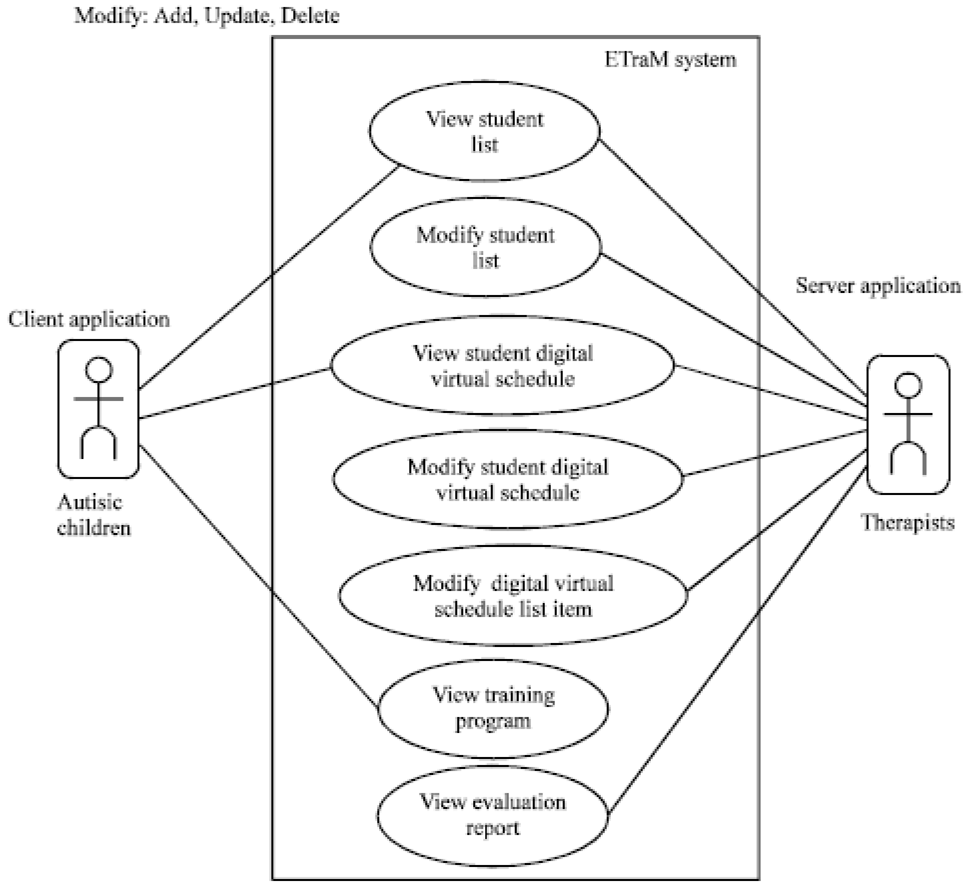

Fig. 3.4.1 Use Case Diagram

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 data input 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

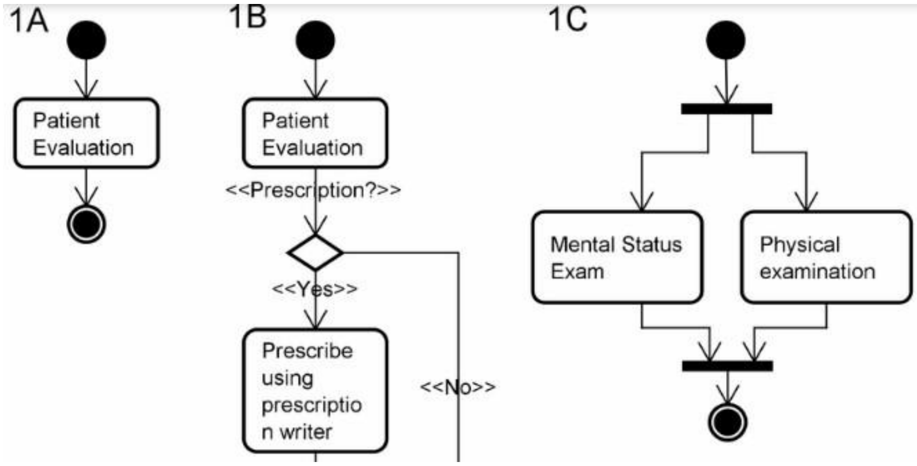

Fig. 3.4.3 Activity Diagram

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

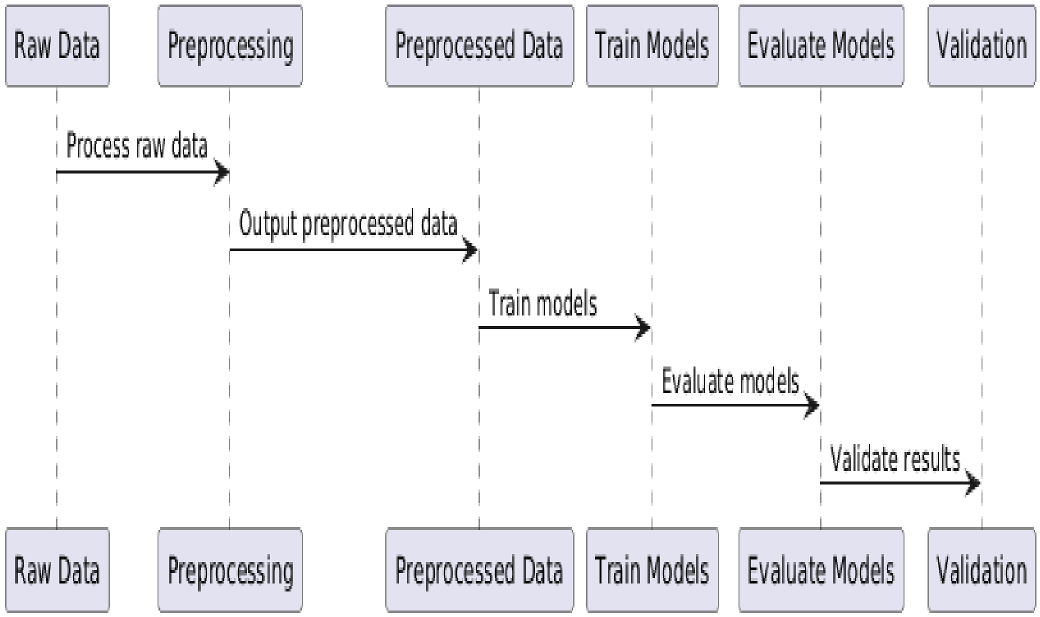

Fig. 3.4.4 Sequence Diagram

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

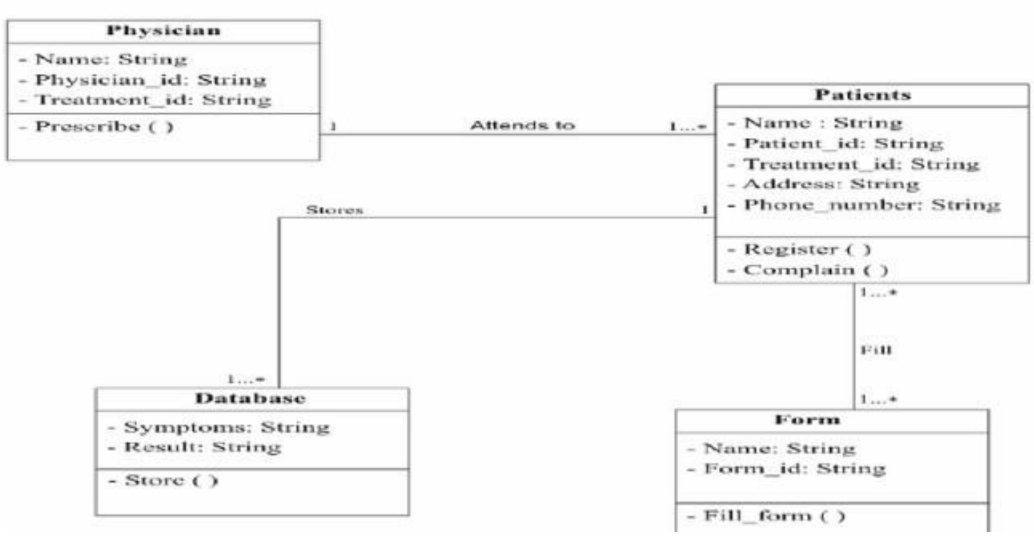

Fig. 3.4.5 Class Diagram

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 handles missing 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, support vector 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.

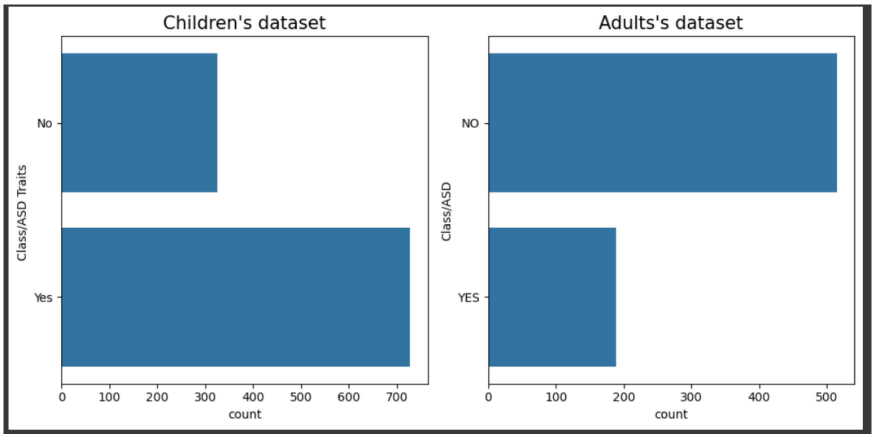

Figure 5.4.1 ASD Diagnosis CHILD vs ADULT

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.

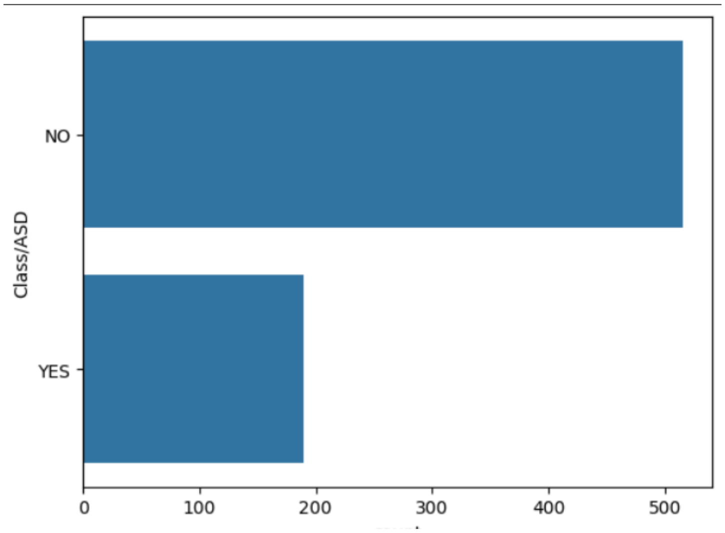

Figure 5.4.2 Class/ASD Combined

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

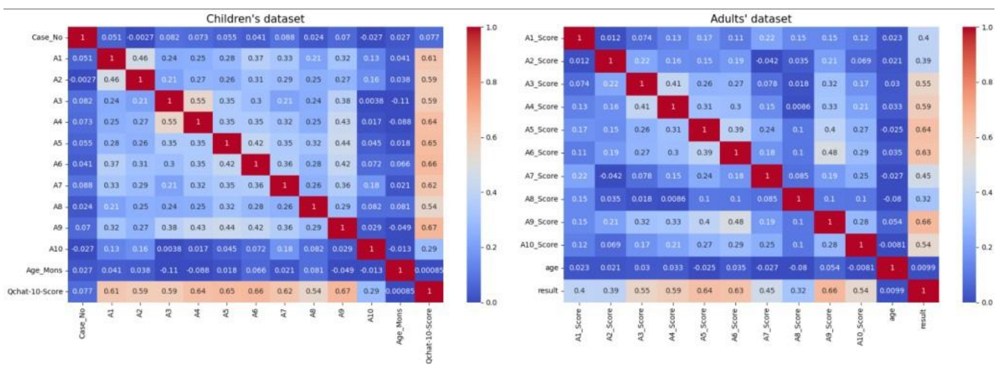

Figure 5.4.3, Correlation Graph of all features

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

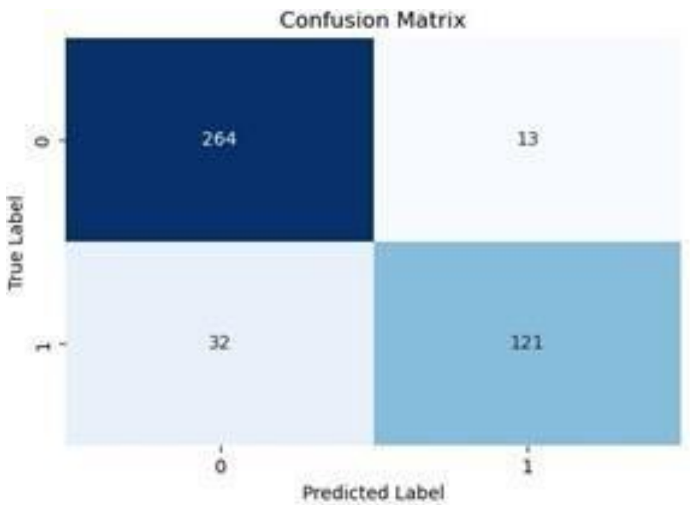

Figure 5.4.4 confusion matrix

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

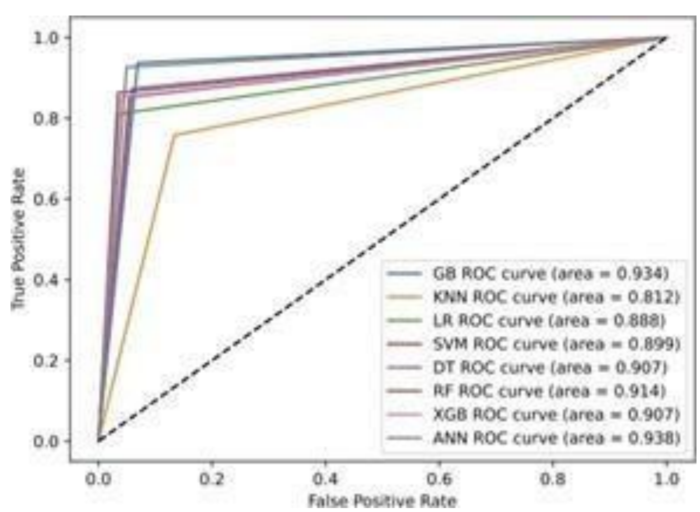

Figure 5.4.5. Training loss over epochs

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

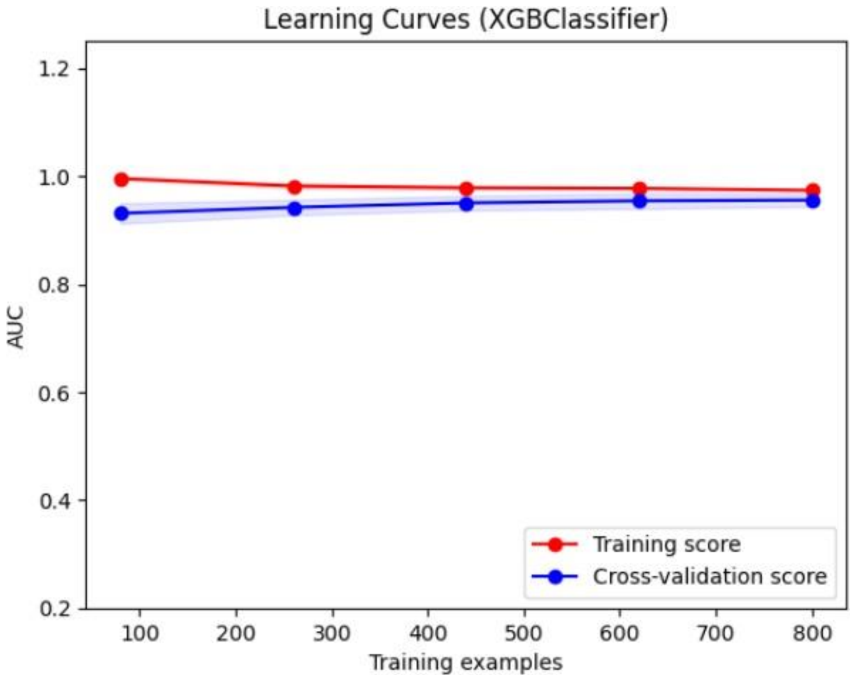

Figure 5.4.6 Performance metrics for XGB

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 of game 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 user a 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 Cancer Using 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- inant analysis 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 on user 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 andDetection 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

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 pltimport

plotly.express as px

from plotly.subplots import make_subplotsimport

plotly.graph_objects as go

from sklearn.impute import SimpleImputerimport warnings

warnings.filterwarnings('ignore')

from sklearn.metrics import confusion_matrixcm =

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_k
```

Lin re

ORIGINALITY REPORT

14%	16%	12%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Illinois at Urbana-Champaign Student Paper	6%
2	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	1%
3	www.springerprofessional.de Internet Source	1%
4	cibgp.com Internet Source	1%
5	krishi.icar.gov.in Internet Source	1%
6	Submitted to Ace Acumen Academy Student Paper	1%
7	link.springer.com Internet Source	1%
8	Submitted to University of Canberra Student Paper	1%

9	www.iieta.org Internet Source	<1%
10	"Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)", Springer Science and Business Media LLC, 2019 Publication	<1%
11	ijarcsms.com Internet Source	<1%
12	mdpi-res.com Internet Source	<1%
13	www.coursehero.com Internet Source	<1%
14	Submitted to Yeditepe University Student Paper	<1%
15	Submitted to Berlin School of Business and Innovation Student Paper	<1%
16	www.frontiersin.org Internet Source	<1%
17	Gonzalez, Emmanuel M.. "Interface Between Plant Science and Data Science: Extracting Insights From Phenomics Data", The University of Arizona, 2024 Publication	<1%

18	Zhanlin Ji, Ziheng Zhao, Xinyi Zeng, Jingkun Wang, Li Zhao, Xueji Zhang, Ivan Ganchev. "ResDSda_U-Net: A novel U-Net based residual network for segmentation of pulmonary nodules in lung CT images", IEEE Access, 2023 Publication	<1%
19	Submitted to La Trobe University Student Paper	<1%
20	francisxavier.ac.in Internet Source	<1%
21	kuey.net Internet Source	<1%
22	www.mdpi.com Internet Source	<1%
23	Submitted to SAMRAT ASHOK TECHNOLOGICAL INSTITUTE VIDISHA M.P Student Paper	<1%
24	Submitted to Visvesvaraya Technological University Student Paper	<1%
25	jestec.taylors.edu.my Internet Source	<1%
26	www.123seminarsonly.com Internet Source	<1%

27	www.nature.com Internet Source	<1%
28	ai4telcos.blogspot.com Internet Source	<1%
29	www.idr.iitkgp.ac.in Internet Source	<1%
30	www.researchsquare.com Internet Source	<1%
31	dokumen.pub Internet Source	<1%
32	iieta.org Internet Source	<1%
33	A Divya, D Sungeetha, Sundar Ramesh. "Efficient protection of golden nutri cereal implementing enhanced disease identification injecting hybrid split-attention mechanism through Novel SANDNES Mechanism", Multimedia Tools and Applications, 2024 Publication	<1%
34	Albert C. Cruz, Andrea Luvisi, Luigi De Bellis, Yiannis Ampatzidis. "X-FIDO: An Effective Application for Detecting Olive Quick Decline Syndrome with Deep Learning and Data Fusion", Frontiers in Plant Science, 2017 Publication	<1%

35	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security", CRC Press, 2023 Publication	<1%
36	Mehmet Metin Ozguven, Kemal Adem. "Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms", Physica A: Statistical Mechanics and its Applications, 2019 Publication	<1%
37	Submitted to University of Wollongong Student Paper	<1%
38	ebin.pub Internet Source	<1%
39	ejmcm.com Internet Source	<1%
40	iarjset.com Internet Source	<1%
41	ijritcc.org Internet Source	<1%
42	publikationen.bibliothek.kit.edu Internet Source	<1%
43	Morales, Mario Andres. "High-Throughput Phenotyping of Seed Quality Traits Using Imaging, Computer Vision, and Deep	<1%

Learning in Pulse Crop", North Dakota State University, 2024

Publication