Malnad College of Engineering

(An Autonomous Institution under Visvesvaraya Technological University, Belagavi)

Hassan–573202

"NeuroSpectra: Revolutionizing Autism Detection Through AI"

A Dissertation submitted to Malnad College of Engineering, Hassan, during the academic year 2024-25 in partial fulfillment for the award of the degree of

Bachelor of Engineering

in

Information Science and Engineering

by

Pavan B 4MC21IS077	**S M Yathin 4MC21IS090**
Rohithnaik S 4MC21IS089	**Sinchana SV 4MC21IS098**

Under the Guidance of

Mrs. Priyanka H L

(Assistant Professor)

Department of ISE

Department of Information Science & Engineering

Malnad College of Engineering

Hassan-573202

Tel.:08172-245093 Fax:08172-245683 URL:www.mcehassan.ac.in

2024-25

Malnad College of Engineering

(An Autonomous Institution under Visvesvaraya Technological University, Belagavi)

Hassan – 573 202

Department of Information Science & Engineering

CERTIFICATE

Certified that the Project Work (20IS702) titled

"NeuroSpectra: Revolutionizing Autism Detection Through AI"

is a bonafide work carried out by

Pavan B	(4MC21IS077)	S M Yathin	(4MC21IS090)
Rohithnaik S	(4MC21IS089)	Sinchana SV	(4MC21IS098)

in partial fulfillment for the award of

Bachelor Degree in Information Science and Engineering

of

Malnad College of Engineering

affiliated to

Visvesvaraya Technological University, Belagavi

during the year 2024-25. It is certified that all corrections/ suggestions indicated for Internal Assessment have been incorporated in the Project report deposited in the Department Library. The Project Report has been approved, as it satisfies the academic requirements in respect of Project Work prescribed for the Bachelor of Engineering Degree.

(Mrs. Priyank H L)	(Dr. Ananda Babu J)	(Dr. A.J Krishnaiah)
Guide	**Head of the Department**	**Principal**

External Viva

Name of the Examiners	**Signature with Date**
1.	
2.	

ACKNOWLEDGEMENT

We have made efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them.

We would like to express our gratitude to our respected principal **Dr. A.J. Krishnaiah** for providing a congenial environment and surroundings to work in. We would like to express our sincere gratitude to **Dr. Ananda Babu J**, Head of the Department of **Information Science and Engineering**, for his continuous support and encouragement.

We are highly indebted to **Mrs. Priyanka H L** for her guidance and constant supervision as well as for providing necessary information regarding the project & also for her support in completing the project.

We would like to express our gratitude to my parents & members of Malnad College of Engineering for their kind co-operation and encouragement which helped us in the completion of this project.

Our thanks and appreciation also go to our colleagues in developing the project and the people who have willingly helped us out with their abilities.

Pavan B - 4MC21IS077

Rohithnaik S - 4MC21IS089

S M Yathin - 4MC21IS090

Sinchana SV- 4MC21IS098

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by difficulties in social interaction, communication, and repetitive behaviors. Early and accurate detection is essential for timely intervention and improved developmental outcomes. This project aims to develop an intelligent system for ASD detection using various supervised machine learning algorithms, including Support Vector Machine (SVM), Random Forest, Decision Tree, Logistic Regression, Naive Bayes, and K-Nearest Neighbors (KNN). The system processes input features derived from behavioral, neurological, and visual indicators collected through standard screening datasets. After preprocessing and feature selection, each algorithm is trained and evaluated, with SVM achieving the highest classification accuracy. The system allows for user-friendly input through a web interface and provides immediate diagnostic feedback. Evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess performance. The proposed system serves as a reliable, scalable, and cost-effective tool that can assist medical professionals in early ASD screening and support decision-making in clinical environments.

CONTENTS

List of Figures

List of Tables

List of Snapshots

Chapter-1

INTRODUCTION

1.1 INTODUCTION TO THE AREA

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, and repetitive behaviors. Early and accurate diagnosis of ASD is crucial for effective intervention and support, yet traditional diagnostic procedures are often time-consuming, subjective, and require specialized clinical expertise. In response to these challenges, NeuroSpectra aims to harness the power of Artificial Intelligence (AI) to transform the way ASD is detected.

This project leverages various machine learning algorithms—including Random Forest, Support Vector Machine (SVM), Logistic Regression, Linear Regression, Decision Tree, K-Nearest Neighbors (KNN), and Naive Bayes—to analyze behavioral and neurological data and predict the presence of ASD. By training these models on relevant datasets, NeuroSpectra seeks to build a robust, accurate, and scalable diagnostic tool that can assist healthcare professionals in early ASD identification.

Through comparative analysis and performance evaluation of these algorithms, the project not only demonstrates the feasibility of AI-driven diagnosis but also identifies the most effective techniques for real-world implementation. Ultimately, NeuroSpectra represents a step forward in utilizing machine learning to support mental health diagnostics, aiming to make early autism detection more accessible, efficient, and objective.

Ultimately, NeuroSpectra represents a significant step toward democratizing ASD diagnostics by reducing reliance on specialist assessments and enabling timely intervention. By combining the analytical power of machine learning with clinically relevant data, the project aspires to make autism detection more accessible, efficient, and objective, particularly in under-resourced settings.

1.2 PROBLEM DEFINITION

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental condition that manifests through a wide range of symptoms affecting communication, social interaction, and behaviour. Despite the increasing prevalence of ASD globally, early and accurate diagnosis remains a major challenge. Traditional diagnostic methods are often subjective, reliant on behavioural observations and expert evaluations, and may result in delayed identification—particularly in regions with limited access to specialized healthcare providers. This delay can significantly hinder early intervention, which is crucial for improving long-term outcomes for individuals with ASD.

By automating the diagnostic process and providing data-driven insights, *NeuroSpectra* aims to reduce the dependency on specialized clinicians, improve diagnostic consistency, and accelerate the early detection of autism. The ultimate goal is to develop a reliable AI-based system that supports healthcare professionals in making timely and informed decisions, thereby enhancing the quality of care and life for individuals affected by ASD.

1.3 OBJECTIVE OF THE PROJECT

The primary objective of the NeuroSpectra project is to develop an AI-driven diagnostic system for early and accurate detection of Autism Spectrum Disorder (ASD). To achieve this, the project is guided by the following specific objectives:

- Develop and validate a machine learning model capable of detecting Autism Spectrum Disorder with high accuracy and reliability.
- Data Analysis: Identify, preprocess, and analyze datasets containing features relevant to ASD diagnosis, such as questionnaire responses, physiological data, or behavioural patterns.
- Model Development: Explore and implement suitable ML techniques (e.g., supervised learning, deep learning) to create a predictive model.
- Feature Identification: Determine the most significant factors contributing to ASD detection, providing insights into its characteristics.
- Early Diagnosis: Focus on improving the sensitivity of the system to detect ASD at an early age, ensuring timely intervention.
- Model Validation: Evaluate the model’s performance using robust validation techniques, ensuring generalizability across diverse datasets and populations

1.4 SCOPE OF THE PROJECT

The scope of the NeuroSpectra project encompasses the development and evaluation of an AI-driven system for the early detection of Autism Spectrum Disorder (ASD) using machine learning algorithms. The project utilizes behavioral and neurological datasets to train and test a variety of supervised learning models, including Random Forest, Support Vector Machine (SVM), Logistic Regression, Linear Regression, Decision Tree, K-Nearest Neighbors (KNN), and Naive Bayes. The primary focus is on building a predictive system that can assist healthcare professionals by providing fast, reliable, and objective assessments of ASD. Through comparative analysis of these algorithms, the project aims to determine the most effective model based on metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. A functional prototype is developed to demonstrate the feasibility of the system in practical applications. The project also emphasizes scalability and accessibility, with the long-term vision of deploying the tool in both clinical and remote environments, particularly in areas with limited access to specialized diagnostic services. Ethical considerations, data handling, and performance reliability are carefully addressed to ensure the model’s real-world applicability and impact.

Chapter 2

LITERATURE SURVEY

2.1 EXISTING SYSTEMS

1. Autism spectrum disorder Screening with Behavioural Features

One of the most widely used approaches for early Autism Spectrum Disorder (ASD) detection is based on analyzing behavioral characteristics through structured datasets. A prominent example is the Autism Screening Adult Dataset (ASD) available from the UCI Machine Learning Repository. This dataset includes essential features such as age, gender, ethnicity, family history of ASD, usage of screening apps, and results from standardized tests like the AQ-10 (Autism Spectrum Quotient). These features are valuable for machine learning models to recognize patterns indicative of ASD.

Multiple machine learning algorithms have been successfully applied to this dataset, particularly Support Vector Machines (SVM), Logistic Regression, Decision Trees, K-Nearest Neighbours (KNN), and Random Forest. These models are trained to classify individuals into two categories—ASD or non-ASD—based on their responses to behavioral questionnaires and personal information. Preprocessing techniques such as label encoding, normalization, and feature selection are typically employed to prepare the data for model training, ensuring better accuracy and computational efficiency.

One of the advantages of using behavioural features is the non-invasive and cost-effective nature of data collection, making such systems highly suitable for early screening and mass-level assessments, particularly in low-resource settings. However, the models' performance can be influenced by data imbalance, small sample size, or demographic bias, which must be addressed through proper sampling techniques and validation strategies such as cross-validation and stratified splitting.

2. NeuroImaging-Based systems

These neuroimaging-based systems aim to uncover the underlying structural and functional brain abnormalities commonly associated with Autism Spectrum Disorder. Functional MRI (fMRI) captures real-time brain activity by measuring changes in blood oxygen levels, while EEG records electrical activity across different regions of the brain. Such data offer a more objective and biologically grounded basis for diagnosis compared to behavioral features alone. However, they also present challenges due to their high dimensionality, noise, and subject variability, requiring advanced preprocessing techniques such as spatial normalization, signal filtering, and region-of-interest (ROI) extraction. Deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are often employed to handle these complex data structures. For example, CNNs have been used to analyze brain scans to detect atypical neural connectivity patterns that are statistically significant in individuals with ASD. The ABIDE dataset has emerged as a benchmark in many of these studies, providing access to multi-site fMRI and anatomical MRI data from both ASD and control subjects. Despite their promise, these neuroimaging-based systems are often constrained by cost, accessibility, and the need for clinical infrastructure, making them more suitable for research and high-end clinical environments rather than large-scale early screening.

3. Hybrid Models

Hybrid models represent a promising direction in ASD diagnosis as they capitalize on the complementary strengths of behavioral assessments and neurobiological indicators. By integrating features such as questionnaire responses, demographic information, and neuroimaging biomarkers, these systems provide a more holistic view of an individual's condition. Feature fusion techniques—such as concatenation, attention-based mechanisms, or dimensionality reduction through PCA (Principal Component Analysis)—are used to merge data from different modalities into a unified input for machine learning models. Advanced ensemble methods like XGBoost and AdaBoost are particularly effective in this context due to their ability to handle heterogeneous data types and complex feature interactions. Studies have shown that hybrid models can significantly outperform single-modality approaches, yielding improved sensitivity and specificity in ASD classification. However, the integration of multi-modal data also introduces challenges related to data alignment, missing values, and computational complexity, which must be addressed through careful model design and preprocessing. These systems are particularly valuable in clinical decision support, where maximizing diagnostic accuracy is critical, and they hold potential for future deployment in intelligent health monitoring platforms. These hybrid systems also pave the way for personalized diagnostic insights, as they can account for individual variability across both cognitive behavior and brain function. As research progresses, such integrative approaches are expected to become central in developing next-generation, AI-powered diagnostic tools for neurodevelopmental disorders like ASD.

2.2 RELATED WORKS

1. An early-stage automated forecasting system to detect children diagnosed with autism spectrum disorder through key sociodemographic and family-related attributes by A.S Albahri (2022)

Initial automated forecasting model for diagnosing and identifying children individuals with autism spectrum disorders utilizing significant socio-demographic and family-features characteristic traits. The ASD primarily impacts children. This ASD will impact individuals' social lives and their way of living. Numerous global health organizations and centers focusing on autism diagnosis and detection are encountering difficulties in delivering an accurate model for detecting and diagnosing Autism Spectrum Disorder. The data regarding ASD detection is influenced by several unidentified factors of the condition, and a prompt resolution is needed to address these factors for ASD. Thus, enhancing the chances to present proof that 'environmental and genetic factors are the primary indicators of ASD is a scientific challenge that must be addressed. This paper primary goal is to develop a predictive model to detect Autism spectrum condition in children as soon as possible, taking into account familial and social influences. This study adopts a three-phase method. To begin with, this research includes data gathering and processing, and the gaps in the data are addressed through the 1-NN model. The characteristics needed for ASD detection are obtained through the Chi-square and relief techniques. To guarantee equity in training, the dataset is balanced using the oversampling technique for minority class (SMOTE). Eight various machine learning algorithms—Decision Tree, Random Forest, Naïve Bayes, KNN, SVM, Logistic Regression, Adaboost, and MLP—were then tested and trained with the generated dataset. Evaluation metrics, such as Accuracy, Precision, and Recall, F1-score, and AUROC, are then employed to evaluate the model.

Initial automated forecasting model for diagnosing and identifying children individuals with autism spectrum disorders utilizing significant socio-demographic and family-features characteristic traits. The ASD primarily impacts children. This ASD will impact individuals' social lives and their way of living. Numerous global health organizations and centers focusing on autism diagnosis and detection are encountering difficulties in delivering an accurate model for detecting and diagnosing Autism Spectrum Disorder. The data regarding ASD detection is influenced by several unidentified factors of the condition, and a prompt resolution is needed to address these factors for ASD.

2. Recognizing Autism Spectrum Disorder with the help of a One-Dimensional convolutional deep learning model by Aythem Khairi Kareem (2023)

The neurological disorder known as autistic spectrum disorder, or ASD, has an impact on behaviour, social interaction, and communication. The field of artificial intelligence known as machine learning is devoted to creating algorithms that recognize patterns in input data and classify ASDs. Applying machine learning techniques to classify ASD has produced a variety of results. Further studies are required to enhance the performance of ASD classification. In order To tackle this, the following deep learning techniques, including 1D CNN have been introduced as an alternative to the categorization of ASD detection.

Three different publicly available ASD datasets (children, adults, and adolescents) are used to evaluate the proposed approaches. Because 1D CNNs are better suited to analyzing Time-series data frequently applied in diagnosing autism spectrum disorder, results indicate that they achieve superior to classical machine learning techniques in classifying ASD across all datasets, with higher accuracies of 99.45%, 98.66%, and 90% For the screening of the autism spectrum condition in Adults, Children, and Adolescents, respectively.

3. The application of artificial intelligence in identifying autism through DTI and fMRI by EmanHelmy (2023)

A Survey of ASD, or neurodevelopmental disorders disorder, includes a number of disorders characterized by difficulties with Speech interactions, repetitive actions, and social behaviours, and linguistic cues. As reported by the Centers for Disease Control(CDC), 1 out of every 44 American children currently suffers from ASD. Clinical behavioural observation tests, which are confidential, time-consuming, and only permit late detection (a kid must be at least two years old to be eligible for an observation report), are the leading method for diagnosing ASD. However, Magnetic resonance imaging, a key neuroimaging tool (MRI) has shown that it can support quick, objective, and early diagnosis and identification of ASD.

Recent advancements in machine literacy (ML) and artificial intelligence (AI) have contributed to the development of appropriate technologies for early detection and automated ASD opinion. Deep literacy (DL), a recent development in artificial intelligence (AI) that relies on artificial neural networks (ANNs), has made it easier to analyze brain MRI data and improved individual capacities for people with ASD. This research focuses on two main MRI types-functional MRI (fMRI) and prolixity tensor imaging (DTI)—to investigate the role of AI in the diagnosis and understanding of autism.

4. A novel strategy for identifying autism spectrum disorder (ASD) using an ensemble diagnostic method based on blood tests by Asmaa H. Rabie (2023)

A New Approach to Identifying Autism Spectrum Disorder through Race and Ensemble Styles Test Information The neurological disorder commonly called autism spectrum disorder (ASD) has an influence on a child's gestures and gregarious communication skills. Common or common signs in early age include repetitive behaviours, limited hobbies, and gregarious commerce. Despite these indications, many people deserve the awareness or comprehension required to recognize ASD early on. Thus, to ensure prompt response and operation, early and accurate discovery using Artificial Intelligence approaches are crucial.

The individual Autism Diapason Complaint (DASD) program is a novel individual path that is introduced in this study. It uses race test data and an ensemble-based AI methodology to swiftly and precisely identify ASD. The Data Preprocessing Layer (DPL) and the individual Subcaste (DL) constitute the two main components of the DASD frame.

Two optimization methods are assumed in the DFL. While the Binary Genetic Algorithm (BGA) is employed to remove erroneous or outlier training data, Binary Gray Wolf Algorithm (BGWA) is used for point election to determine which qualities in the dataset are most relevant. Only important and high-quality data will be applied in the individual phase thanks to this preparation procedure.

5. Utilizing machine learning methods to identify children with autism spectrum disorder

Early identification of autism spectrum disorder (ASD) is often beneficial to children's long-term health. Because discovery styles rely on the pricey and confidential evaluation of specialists. In order to describe children with ASD, we proposed an engine literacy approach in this work that combines behavioral data (such as eye preoccupation and facial expression) with physiological data (such as electroencephalography, or EEG). Its use can lower prices and improve the efficacy of discovery. First, we used a creative approach to identify the salient characteristics of the EEG data, facial expressions, and ocular preoccupation. Additionally, a mongrel emulsion path with a bracket delicacy of 87.50 was provided for multimodal data emulsion, based on a weighted naive Bayes algorithm.

It effects imply that the engine mastering bracket method in this investigation is effective for identifying ASD early. Distraction matrices and graphs show that EEG may be the most discriminating information, and that ocular obsession, facial expression, and EEG have different discriminatory dominions for the identification of ASD and typically developing children. There are significant reciprocal features between the behavioural and physiological data. Therefore, bracket delicacy can be much improved by the engine literacy approach suggested in this work, which incorporates the reciprocal information.

2.3 GAPS IDENTIFIED

1. **Limited Generalizability:** Many existing models are trained and validated on small or homogeneous datasets, reducing their ability to generalize to diverse populations with varying demographics, cultures, and behavioural patterns.
2. **Lack of Multi-Modal Integration:** Although hybrid models exist, most systems rely exclusively on either behavioural or neurological data. There is a significant gap in the seamless integration of multi-modal data to achieve comprehensive and accurate diagnosis.
3. **Insufficient Early Screening Tools:** Current systems often focus on clinical diagnosis rather than scalable early screening. There is a need for lightweight, accessible, and easy-to-use AI tools for mass-level preliminary assessments.
4. **Data Imbalance and Bias:** Many datasets used in ASD detection suffer from class imbalance and lack of diversity, leading to biased predictions and reduced reliability for underrepresented groups.
5. **Lack of Clinical Validation:** A majority of machine learning models are still in experimental or academic phases, with few having undergone real-world clinical trials or received validation from healthcare professionals.
6. **Interpretability and Trust:** While some systems achieve high accuracy, they lack interpretability, making it difficult for clinicians to trust and adopt these AI tools in their diagnostic workflows.

Chapter 3

SYSTEM AND REQUIREMENT ANALYSIS

3.1 SOFTWARE REQUIREMENTS AND SPECIFICATION

Software Requirements Specification (SRS) provides an overview of the entire SRS with purpose, scope, definitions, acronyms, abbreviations, references and overview of the SRS. A software requirements specification (SRS) is a comprehensive description of the intended purpose and environment for software under development. The SRS fully describes what the software will do and how it will be expected to perform the various gestures and determining its accuracy.

The SRS is a requirements specification for a software system, is a description of the behaviour of a system to be developed and may include a set of use cases that describe interactions the users will have with the software. In addition, it also contains nonfunctional requirements. Nonfunctional requirements impose constraints on the design or implementation.

An SRS minimizes the time and effort required by developers to achieve desired goals and minimizes the development cost. A good SRS defines how an application will interact with system hardware, other programs and human users in a wide variety of real-world situations. Parameters such as operating speed, response time, availability, portability, maintainability, footprint, security and speed of recovery from adverse events are evaluate.

3.2 EXISTING SYSTEM ANALYSIS

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges with social skills, repetitive behaviors, speech, and nonverbal communication. Early diagnosis is crucial for effective intervention, but traditional diagnostic methods are often time-consuming, subjective, and require extensive clinical evaluation.

In the current healthcare system, the diagnosis of ASD is primarily conducted through standardized behavioral assessments such as the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R). These assessments require trained professionals and multiple sessions, making the process lengthy and less accessible, especially in rural or underdeveloped areas. Furthermore, the subjective nature of observational diagnostics introduces a risk of inconsistency and misinterpretation.

➢ **Disadvantages of Existing System**

1. **Subjectivity in Traditional Diagnosis**

   Traditional diagnostic methods heavily rely on behavioral assessments conducted by clinicians, which are often subjective. Different professionals may interpret symptoms differently, leading to inconsistent or delayed diagnoses.

2. **Time-Consuming and Resource-Intensive**

   Clinical assessments like ADOS and ADI-R are lengthy and require multiple sessions with trained professionals, which can delay early intervention and are often not accessible in remote or under-resourced areas.

3. **Limited Accessibility**

   In many regions, particularly rural or low-income areas, access to specialized mental health professionals and diagnostic tools is limited, making early detection difficult.

4. **Dependence on Structured Clinical Data**

   Most machine learning systems depend on well-structured, pre-labeled datasets collected in controlled environments. Such data may not always be available in real-world scenarios, limiting model deployment.

5. **Insufficient Generalization**

   Many existing machine learning models are trained on limited or imbalanced datasets. As a result, they may not generalize well to unseen data or diverse populations, potentially reducing their effectiveness across different demographics.

3.3 BREIF DESCRIPTION ABOUT PROPOSED SYSTEM

The proposed system aims to enhance the early and accurate detection of Autism Spectrum Disorder (ASD) using supervised machine learning algorithms. The system is designed to overcome the limitations of traditional diagnostic methods by offering a faster, more objective, and scalable approach.

In this system, a dataset containing behavioral and psychological attributes of individuals is used to train and evaluate multiple machine learning classifiers, including Support Vector Machine (SVM), Random Forest, Decision Tree, Logistic Regression, and Naive Bayes. After preprocessing the data and selecting relevant features, each model is trained to classify individuals as autistic or non-autistic.

Among the models tested, Support Vector Machine (SVM) achieved the highest accuracy, demonstrating its ability to effectively handle high-dimensional data and classify complex patterns. The results highlight the potential of machine learning to assist healthcare professionals in making quicker and more accurate diagnostic decisions.

The system is scalable and can be integrated into online platforms or healthcare tools to facilitate ASD screening in clinical and non-clinical settings. It reduces dependency on lengthy assessments and makes preliminary diagnosis more accessible, especially in underserved areas.

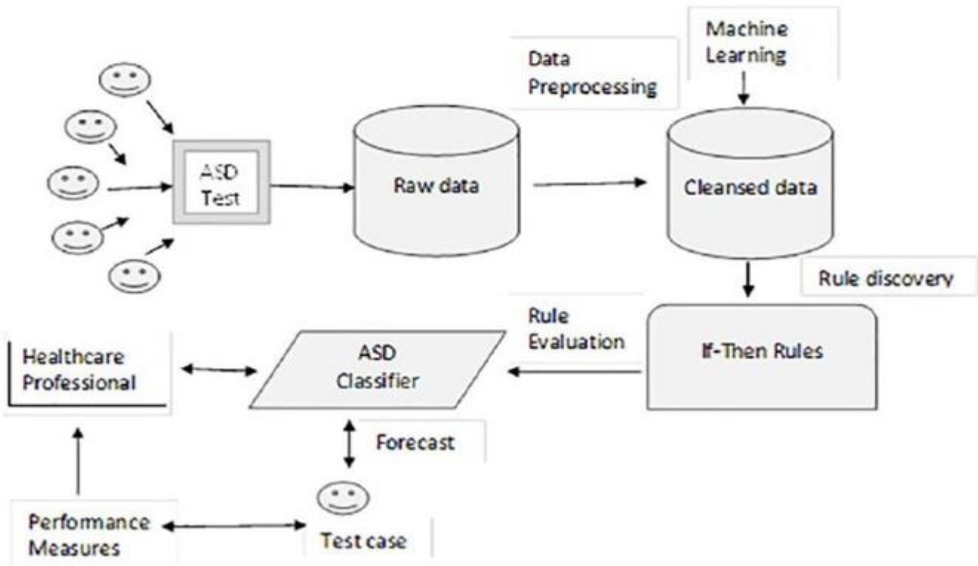


Figure 3.3: Architecture of the Proposed System

The architecture for Autism Spectrum Disorder (ASD) detection using machine learning follows a systematic approach that starts with collecting data from individuals through ASD screening tests. This raw data, which includes various behavioral and psychological attributes, is stored for further processing. Since raw data often contains noise, missing values, or inconsistencies, it undergoes data preprocessing to cleanse and standardize it. The cleansed data is then fed into machine learning algorithms such as SVM, Random Forest, Decision Tree, Logistic Regression, and Naive Bayes to train models capable of identifying patterns associated with ASD. During this training process, the system also performs rule discovery to generate interpretable "If-Then" rules that enhance decision-making transparency.

Once the model is trained, an ASD classifier is developed to predict the likelihood of ASD in new test cases. This classifier uses both the trained model and the discovered rules to generate forecasts. These predictions are then validated through rule evaluation to ensure logical consistency. The results are presented to healthcare professionals, who can use the forecast as a decision-support tool rather than a standalone diagnosis. Additionally, the system incorporates a performance measurement module that evaluates the accuracy, precision, recall, and F1-score of the classifier, ensuring the model's reliability and guiding further improvements. This architecture aims to provide a fast, accurate, and interpretable method for ASD detection while supporting clinical judgment.

3.4 FUNCTIONAL REQUIREMENTS

A functional requirement defines a function of a software system or its components. A function is described as a set of inputs, the behaviour, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish.

Behavioural requirements describing all the cases where the system uses the functional requirements are captured in use cases. Functional requirements are supported by non-functional requirements (also known as quality requirements), which impose constraints on the design or implementation (such as performance requirements, security, or reliability).

1. **User Data Input Module:** The system shall allow users or healthcare providers to input ASD test data (e.g., answers to screening questions, behavioral traits, demographic information). The system shall validate the input format and completeness before processing.

2. **Data Preprocessing Module:** The system shall clean the raw data by handling missing values, encoding categorical variables, and normalizing numerical values. The system shall store the preprocessed (cleansed) data for further analysis.

3. **Machine Learning Model Training:** The system shall allow training of multiple machine learning models such as SVM, Random Forest, Decision Tree, Logistic Regression, and Naive Bayes. The system shall split the dataset into training and testing subsets for model evaluation.

4. **Rule Discovery Engine:** The system shall extract interpretable "If-Then" rules based on the trained data for better transparency and decision-making support.

5. **ASD Classifier Module:** The system shall use trained models to classify whether a new test case indicates a potential ASD diagnosis. The classifier shall generate a binary output (ASD or Non-ASD) based on input features.

6. **Prediction and Forecasting:** The system shall predict the likelihood of ASD for individual test cases submitted by users. The system shall display prediction results to healthcare professionals in a comprehensible format.

7. **Rule Evaluation:** The system shall evaluate predictions against rule-based logic to ensure consistency and reliability.

8. **Performance Evaluation:** The system shall calculate performance metrics such as accuracy, precision, recall, and F1-score for each machine learning model. The system shall compare these metrics to identify the best-performing model (e.g., SVM).

3.5 NON-FUNCTIONAL REQUIREMENTS

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviours. They are contrasted with functional requirements that define specific behaviour or functions.

Non-functional requirements define how a system is supposed to be. These are in the form of "system shall be requirement", an overall property of the system as a whole or of a particular aspect and not a specific function. The system's overall properties commonly mark the difference between whether the development project has succeeded or failed. Non-functional requirements are often called “quality attributes” of a system.

1. **Performance:** The system shall provide predictions within a few seconds of data submission. The system shall efficiently handle datasets of varying sizes without significant degradation in performance.
2. **Scalability:** The system shall be scalable to accommodate larger datasets and more user inputs in the future. It shall support the integration of additional machine learning models or features as needed.
3. **Reliability:** The system shall consistently produce accurate and reliable results across multiple runs. It shall maintain functionality even under high usage or minor system failures.
4. **Availability:** The system shall be available for use 24/7 with minimal downtime. Regular backups shall be implemented to ensure availability of data and models.
5. **Maintainability:** The system shall be modular and well-documented to support easy maintenance and updates. The codebase shall follow standard development practices to facilitate future enhancements.
6. **Security:** The system shall ensure that all user and patient data is stored and transmitted securely. It shall implement proper authentication and authorization mechanisms to prevent unauthorized access.
7. **Usability:** The user interface shall be intuitive and easy to use for healthcare professionals and non-technical users. Clear instructions, tooltips, and labels shall be provided to guide the user through the process.
8. **Portability:** The system shall be deployable on various platforms, including web-based systems or local desktop environments. It shall be compatible with major operating systems like Windows and Linux.

3.6 HARDWARE REQUIREMENTS

Processor	:	intel i5
RAM	:	6GB
Hard Disk	:	256GB
Speed	:	1.2 GHz

- **Processor:** The processor, often referred to as the CPU (Central Processing Unit), serves as the brain of the computer, executing instructions and processing data. An intel i5 processor or higher is recommended for the system's CPU, ensuring adequate processing power to handle the computational requirements of the software project. Higher-grade processors typically offer better performance, efficiency, and support for advanced features, which can contribute to faster execution of tasks and improved overall system responsiveness.

- **RAM (Random Access Memory):** 6 GB RAM plays a crucial role in the system's performance by providing temporary storage for running applications and data. With a minimum of 6 GB of RAM recommended for the system, it ensures that the computer can effectively handle the execution of the software project and store necessary data in memory. More RAM can further improve system performance, particularly when working with large datasets or running resource-intensive tasks such as data processing, analysis, or simulations.

- **Hard Disk:** 500 GB Adequate storage space is essential for storing the operating system, installed applications, and project-related data, including files, datasets, and databases. With a minimum of 500 GB of hard disk space recommended for the system, it ensures that there is sufficient room to accommodate the software project's storage requirements without running out of disk space. Having ample storage space also allows for future expansion and the addition of more data or applications as needed.

- **Processor Speed**: 2 GHz: Processor speed, measured in gigahertz (GHz), indicates how quickly the CPU can execute instructions. With a recommended processor speed of 2 GHz or higher, it ensures that the CPU can efficiently process instructions and handle computational tasks at a satisfactory rate. Higher processor speeds generally result in faster computing tasks, reducing processing time and improving system performance, particularly for tasks that require significant computational resources.

3.7 SOFTWARE REQUIREMENTS

Operating System	:	Windows 11
Language	:	Python 3.8
Database	:	MySQL

- **Operating System:** An operating system (OS) is system software that manages computer hardware and software resources and provides common services for computer programs. The operating system is a component of the system software in a computer system. Application programs usually require an operating system to function. Windows operating system is used in this project of versions Windows XP or higher.

- **Python:** Python is a versatile and widely-used programming language known for its simplicity, readability, and rich set of libraries, making it ideal for rapid development and applications like AI, data science, and IoT. It is extensible, embeddable, and portable, and its object-oriented and open-source nature supports productivity across platforms. However, it has some limitations. It is generally slower than compiled languages, less suitable for mobile development, and has limited browser support.
- **MySQL:** As for the backend database management system, MySQL is selected. SQL Server is a powerful and reliable relational database management system (RDBMS) that offers advanced features such as high availability, scalability, and security. It provides seamless integration, allowing for efficient data storage, retrieval, and management.

Chapter 4

SYSTEM MODELLING AND DESIGN

4.1 SYSTEM MODELS

Systems modelling is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering.

4.1.1 USE CASE DIAGRAM

A use case diagram is a type of diagram used in Unified Modeling Language (UML) to visually represent the interactions between users (or actors) and a system. It outlines the different ways users can interact with the system to achieve their goals, providing a high-level view of the system's functionality.

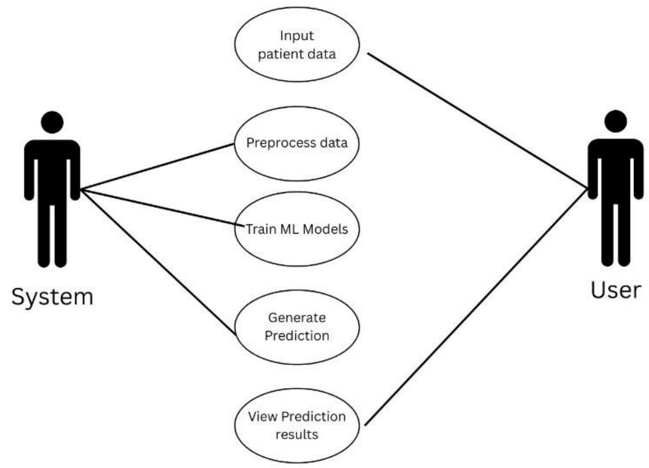


Figure 4.1.1: Use-case Diagram

The use case diagram for the Autism Spectrum Disorder (ASD) Detection System represents the interaction between the user (typically a healthcare professional) and the system's core functionalities. The user begins by inputting patient data obtained from ASD screening tests. This data is then preprocessed by the system to remove inconsistencies and prepare it for analysis. The system allows training of various machine learning models such as

SVM, Random Forest, Decision Tree, Logistic Regression, and Naive Bayes. Once trained, these models are used to generate predictions for new patient cases, identifying whether the individual is likely to have ASD. The user can view the prediction results, along with performance metrics like accuracy and precision for each model. Additionally, the system supports downloading of detailed reports for clinical or research use. This structured workflow enhances the speed, accuracy, and reliability of ASD detection, providing valuable decision support to healthcare professionals.

4.1.2 FLOW CHART DIAGRAM

A flowchart is a diagram widely used across various fields to depict processes or algorithms in clear, easy-to-understand diagrams. Using shapes like rectangles, ovals, and diamonds connected by arrows, flowcharts range from simple hand-drawn charts to complex computer-generated diagrams.

They are among the most common diagrams globally, utilized by technical and non-technical individuals. Flowcharts, also known as Process Flowcharts or Business Process Mapping, play a key role in documenting, planning, and communicating processes.

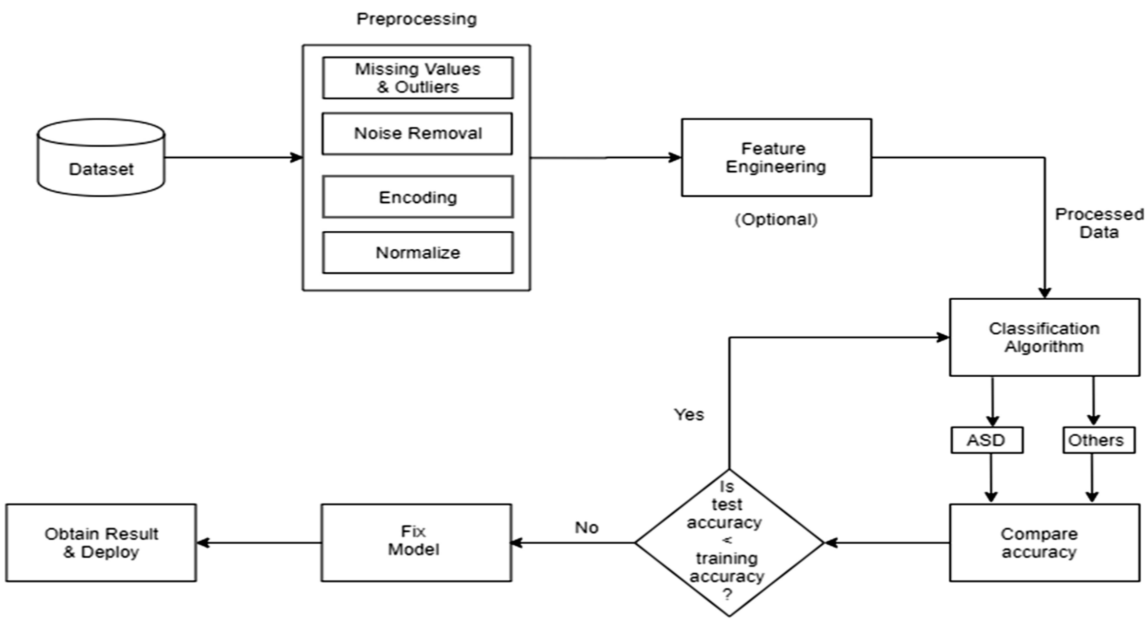


Figure 4.1.2: Flow chart Diagram

This flowchart illustrates the step-by-step process for detecting Autism Spectrum Disorder (ASD) using a machine learning approach. The process begins with a dataset that undergoes preprocessing, where missing values, outliers, and noise are handled, and the data is encoded and normalized. The processed data is then fed into a classification algorithm, which predicts whether a case falls under ASD or other categories. The accuracy of the model is evaluated by comparing test accuracy with training accuracy. If the test accuracy is lower, it indicates possible overfitting, and the model is refined. Once the model achieves acceptable performance, the final results are obtained and deployed for practical use. This ensures a reliable and optimized model for early ASD detection.

4.2 DESIGN / WORKING DOCUMENT

4.2.1 SYSTEM DESIGN

The system is designed to automate the detection of Autism Spectrum Disorder (ASD) using various machine learning algorithms. The core components of the system include a data preprocessing module, an optional feature engineering unit, classification models, performance evaluation logic, and result deployment. The preprocessing module is responsible for handling missing values, removing outliers and noise, encoding categorical values, and normalizing numerical data. This ensures the dataset is clean and consistent before feeding it into the models. The classification module supports algorithms like Support Vector Machine (SVM), Random Forest, Decision Tree, Logistic Regression, and Naive Bayes, with SVM providing the highest accuracy. The model's performance is monitored using accuracy comparison between training and test data to avoid overfitting.

#	Attribute Name	Description
1	ID	ID of the patient
2	Age	Age of the patient in years
3	Gender	Gender of the patient
4	Ethnicity	Ethnicity of the patient
5	Jaundice	Whether the patient had jaundice at the time of birth
6	autism	Whether an immediate family member has been diagnosed with autism
7	contry_of_res	Country of residence of the patient
8	used_app_before	Whether the patient has undergone a screening test before
9	result	Score for AQ1-10 screening test

10	relation	Relation of patient who completed the test
11	Class/ASD	Classified result as 0 or 1. Here 0 represents No and 1 represents Yes. This is the target column, and during submission submit the values as 0 or 1 only.

Table 4.2.1: Dataset attributes and description

4.2.2 WORKFLOW DESCRIPTION

The system starts by importing a dataset consisting of ASD screening data. It is passed through several preprocessing steps such as missing value treatment, noise filtering, encoding, and normalization. Optionally, feature engineering may be performed to improve model performance. The processed data is then passed to the classification model for training and testing. If the test accuracy is significantly lower than the training accuracy, the model is diagnosed for overfitting and retrained or adjusted. Once the accuracy is acceptable and consistent, the model is finalized. The final predictions classify subjects into ASD or Non-ASD categories. The results, along with performance metrics, are deployed and made accessible to healthcare professionals for real-time analysis and support in clinical decision-making.

Chapter 5

IMPLEMENTATION

Implementation is the stage in the project where the theoretical design is turned into the working system and is giving confidence to the new system for the users i.e., will work efficiently and effectively. It involves careful planning, investigation of the current system and its constraints on implementation, design of method to achieve the changeover, an evaluation, of change over methods. Apart from planning major task of preparing the implementation is education of users. The more complex system is implemented, the more involved will be the system analysis and design effort required just for implementation. An implementation coordinating committee based on policies of individual organization has been appointed. The major elements of implementation plan are test plan, training plan, equipment installation plan, and a conversion plan. The implementation state involves the following tasks:

➢ Careful planning

➢ Investigation of system and constraints.

➢ Design of methods to achieve the changeover.

➢ Training of the staff in the changeover phase

5.1 DESCRIPTION OF MODULES

1. Data Collection Module

This module is responsible for gathering ASD-related data from reliable sources such as screening test results, clinical records, or publicly available datasets. The data typically includes attributes related to behavioral patterns, communication skills, and developmental history of individuals.

2. Data Preprocessing Module

In this module, raw data is cleaned and transformed to prepare it for model training. Tasks include:

- Handling missing values and outliers
- Noise removal
- Encoding categorical features (e.g., gender, yes/no responses)
- Normalizing numerical features

3. Model Training Module

This module applies various machine learning algorithms: SVM, Random Forest, Decision Tree, Logistic Regression, and Naive Bayes on the processed data. The data is split into training and testing sets, and each model is trained to learn patterns associated with ASD.

4. Model Evaluation Module

After training, this module evaluates the performance of each algorithm using metrics such as:

- Accuracy: Accuracy is the ratio of correctly predicted observations to the total observations.
- Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.
- Recall: Recall is the ratio of correctly predicted positive observations to all actual positives.
- F1-Score: F1 Score is the **harmonic mean of precision and recall**. It balances the two when you want to ensure both false positives and false negatives are minimized.

It also compares test accuracy with training accuracy to check for overfitting or underfitting.

5. Prediction Module

This module uses the trained model to predict whether a new input (test case) is ASD or Non-ASD. It provides real-time classification results and can be integrated into a user-facing application or API.

6. Result Visualization & Reporting Module

This module presents the prediction results and model performance metrics in an understandable format. It may include graphs, tables, or downloadable reports that healthcare professionals can use for decision-making and documentation.

5.2 ALGORITHM / LOGIC USED

1. Support Vector Machine (SVM)

SVM is a powerful supervised learning algorithm used for classification. It works by finding the optimal hyperplane that best separates the data into different classes. In this project, SVM effectively distinguishes between ASD and non-ASD cases by maximizing the margin between classes. It is particularly effective in high-dimensional spaces and delivered the highest accuracy in your study.

2. Random Forest

Random Forest is an ensemble learning algorithm that builds multiple decision trees and merges their results to improve classification accuracy. It reduces overfitting and handles both categorical and numerical data well. In this project, it helps in identifying complex patterns within the ASD dataset by averaging the results of several trees for robust predictions.

3. Decision Tree

Decision Trees classify data by asking a series of questions based on feature values, forming a tree-like structure. Each internal node represents a test on an attribute, and each leaf node represents a class label. Although simple and easy to interpret, decision trees can overfit on small or noisy datasets, but they are very useful for understanding decision rules in ASD detection.

4. Logistic Regression

Logistic Regression is a statistical model used for binary classification tasks. It calculates the probability that a data point belongs to a specific class (ASD or not) using a logistic (sigmoid) function. It’s fast, interpretable, and performs well when the data has a linear relationship between input features and the target.

5. Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' Theorem. It assumes that all features are independent of each other (naive assumption). Despite this simplification, it works surprisingly well for many real-world datasets, especially when the features are not highly correlated. It is effective for initial predictions in ASD screening tests.

6. K-Nearest Neighbors (KNN)

KNN is an instance-based algorithm that classifies a new data point based on the majority class among its 'K' nearest neighbors in the feature space. It does not require training and is highly effective when the data distribution is simple. In ASD detection, KNN provides good accuracy but can be sensitive to irrelevant features and data scaling.

Out of these 6 algorithms SVM gives highest accuracy.

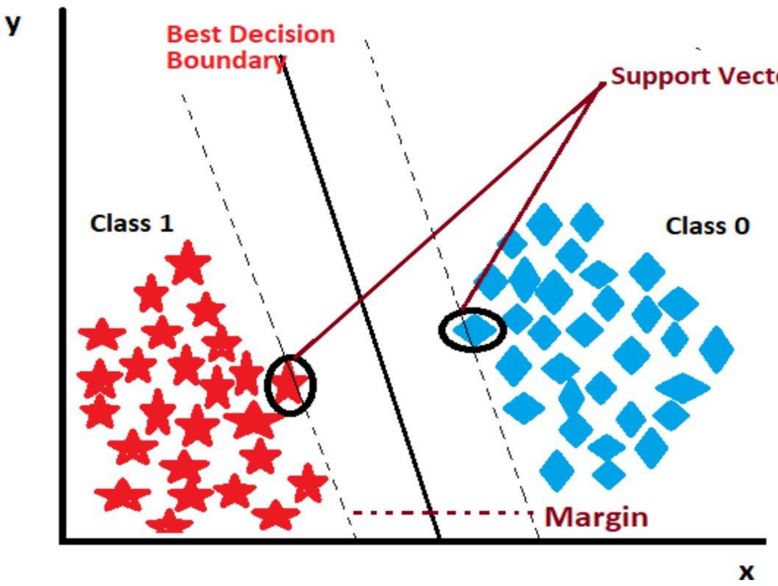


Figure 5.2.1: Working of Support vector Machines

5.3 METHODOLOGY

The methodology adopted for this project involves a structured machine learning pipeline aimed at detecting Autism Spectrum Disorder (ASD) based on behavioral and psychological data. The process is divided into several systematic phases, as described below:

1. Data Collection

The first step involves collecting relevant ASD screening data from publicly available sources such as UCI Machine Learning Repository or healthcare datasets. These datasets typically include personal, medical, and behavioral attributes related to individuals, such as age, gender, social responsiveness, and questionnaire responses.

2. Data Preprocessing

The collected data often contains missing values, noise, or inconsistent entries. Preprocessing involves:

- Handling missing or null values
- Removing duplicates and outliers
- Encoding categorical features into numerical form (e.g., gender, yes/no answers)
- Normalizing or scaling features to ensure uniformity

This step ensures that the input is clean and structured for machine learning models.

3. Model Implementation

Multiple supervised learning algorithms are applied to the dataset:

- Support Vector Machine (SVM)
- Random Forest
- Decision Tree
- Logistic Regression
- Naive Bayes
- K-Nearest Neighbors (KNN)

Each model is trained on a portion of the dataset and evaluated on a separate test set.

4. Model Evaluation

The trained models are evaluated using standard classification metrics:

- Accuracy: Accuracy is the ratio of correctly predicted observations to the total observations.
- Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.
- Recall: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.
- F1-Score

A comparison is made among all algorithms, and the one with the highest performance (SVM in this case) is selected as the final model.

5. Prediction and Result Generation

The selected model is used to predict ASD for new or unseen patient data. The prediction is binary (ASD or Non-ASD), and the output can also include a probability/confidence score.

6. Result Visualization & Reporting

The system presents the results in a user-friendly format, including visualizations such as bar charts, confusion matrices, or tables. Reports can be generated for documentation or further medical review.

5.3 TECHNOLOGIES USED IN THE PROJECT

PYTHON

Python is a versatile and widely-used programming language known for its simplicity, readability, and rich set of libraries, making it ideal for rapid development and applications like AI, data science, and IoT. It is extensible, embeddable, and portable, and its object-oriented and open-source nature supports productivity across platforms. However, Python has some limitations. It is generally slower than compiled languages, less suitable for mobile development, and has limited browser support. Its database access layers are underdeveloped, and dynamic typing can lead to runtime errors. Despite these drawbacks, Python remains a top choice for academic and research projects.

Its active developer community ensures constant improvement and support. Additionally, Python's compatibility with frameworks like TensorFlow and Stream lit makes it an excellent choice for machine learning and interactive web applications.

MYSQL

MySQL is a popular open-source relational database management system (RDBMS) known for its speed, reliability, and ease of integration with web applications. It supports structured data, multi-user access, and ACID compliance for transaction safety. It is a widely used with platforms like Python, PHP, and Java, and provides strong security features, scalability, and replication support. It is ideal for small to medium-sized applications, especially for web and mobile projects. However, it has some limitations, such as basic support for large-scale data analytics, limited stored procedure capabilities, and fewer advanced features compared to enterprise solutions like PostgreSQL or Oracle. It may also require manual optimization for complex queries and lacks native support for full-text indexing across all storage engines.

VSCODE

Visual Studio Code (VS Code) is a lightweight, open-source code editor developed by Microsoft. It is highly customizable, with a vast library of extensions for different languages, frameworks, and tools. It supports intelligent code completion, debugging, Git integration, and more, all in a user-friendly interface. Despite its strengths, VS Code can sometimes consume more system resources, and users may face a steeper learning curve when managing large projects or configuring extensions for the first time.

Chapter 6

RESULTS AND DISCUSSION

6.1 TESTING AND OUTPUT ANALYSIS

6.1.1 SOFTWARE TESTING

Testing is a process of executing a program with the aim of finding errors. To make our software perform well, it should be error-free. If testing is done successfully, it will remove all the errors from the software.

Software testing is a crucial phase in the software development life cycle (SDLC) as it ensures the reliability, security, and overall quality of the product. It involves various methods such as unit testing, integration testing, system testing, and user acceptance testing, each focusing on different aspects of the application.

Testing can be manual or automated, and it helps verify that the software meets the specified requirements and behaves as expected in different scenarios. Moreover, early detection of bugs through testing reduces the cost and time required for fixing them later. By ensuring proper testing, developers can deliver robust and user-friendly software, improving user satisfaction and reducing the chances of failure after deployment

6.1.2 TEST CASES

Testcases for Positive Scenario

Testcase ID	Positive Scenario	Required Input	Expected Output	Actual output	Test Pass/Fail
TC001	User submits valid ASD questionnaire data	Age: 25, Gender: Male, Q1-Q10: Mostly “Yes”, Jaundice: Yes, Family History: Yes	Output: ASD Positive	Output: ASD Positive	Pass
TC002	User submits valid Non-ASD data	Age: 30, Gender: Female, Q1-Q10: Mostly “No”, Jaundice: No, Family History: No	Output: Non-ASD	Output: Non-ASD	Pass

TC003	System processes a valid CSV dataset	Uploaded dataset with 100 clean records	Dataset processed and split into train/test sets	Dataset successfully processed	Pass
TC004	Evaluate model with test dataset	Valid pre processed test data (50 samples)	Accuracy > 85%, SVM performs best	Accuracy = 100%, SVM is best	Pass
TC005	Train SVM with balanced dataset	Clean, labeled training data (ASD & Non-ASD)	Model trains successfully, no errors	Model training completed with 98% accuracy	Pass
TC006	Generate report after prediction	Valid prediction results	Report generated with prediction, accuracy, F1-score	Report generated successfully	Pass

Table 6.1.2 (A) : Testcases for Positive Scenario

Testcases for Negative Scenario

Testcase ID	Positive Scenario	Required Input	Expected Output	Actual output	Test Pass/Fail
TC001	Missing mandatory input fields	Age: –, Gender: Male, Q1-Q10: Valid	Error message: “Age is required”	Error displayed: “Age is required”	Pass
TC002	Incomplete questionnaire submission	Only 5 out of 10 questions answered	Error: “All questions must be answered”	Error displayed	Pass
TC003	Unsupported file format uploaded	Uploaded file: image.png instead of .csv	Error: “Invalid file type”	Error displayed	Pass
TC004	Incorrect labels in dataset	Labels: “Maybe” instead of “Yes/No”	Error: “Unrecognized label values”	Label validation error	Pass
TC005	Model prediction without preprocessing	Raw data input without encoding	Error: “Input must be preprocessed”	System shows preprocessing required	Pass

TC006	Overlarge input file (>10MB)	Uploads 50MB dataset	Error: “File size exceeds limit”	Error message displayed	Pass
TC007	Attempt to train model with all same class	Dataset: 100 ASD cases, 0 Non-ASD	Warning: “Unbalanced dataset” or model training fails	Model flags imbalance	Pass

Table 6.1.2 (B) : Testcases for Negative Scenario

Result

When Users enters all their Medical Data the System using Support Vector Machines predicts that if that person has Autism Spectrum disorder or not.

6.2 Snapshots of User Interface

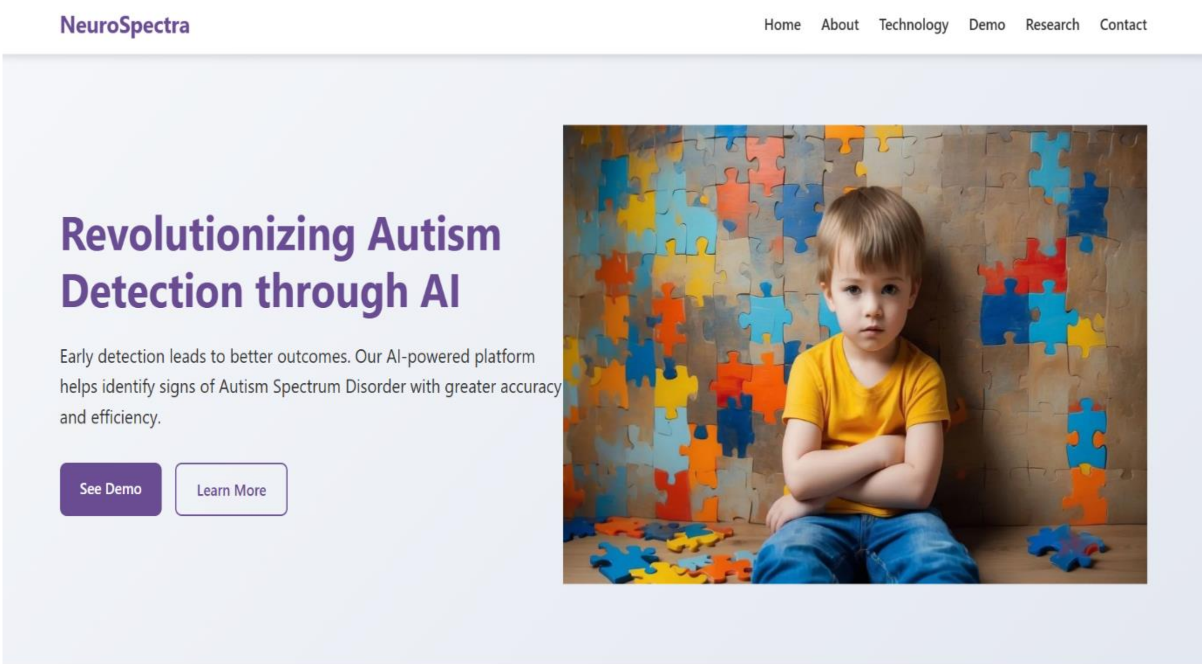


Figure 6.2.1 : Home Page

Home Page : The Home Page of the Autism Spectrum Disorder Detection system provides an intuitive and user-friendly interface for users to begin the ASD screening process. It offers clear navigation to upload data, view model results, and access information about ASD. The design is simple and informative, ensuring ease of use for both professionals and general users.

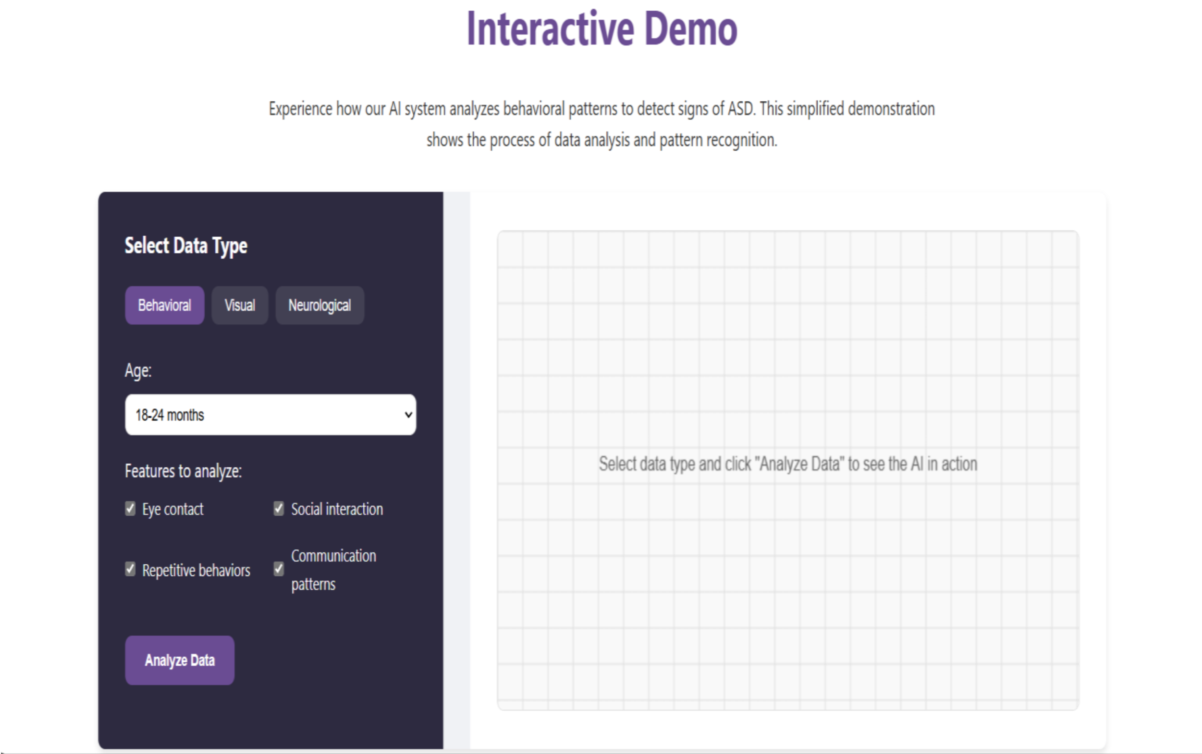


Figure 6.2.3 : Features Input page

Features Input page : The Feature Input Page collects critical information categorized into Behavioral, Neurological, and Visual features relevant to Autism Spectrum Disorder detection. Behavioral features include responses to social interaction questions and repetitive behaviors, while Neurological features cover aspects like family history and developmental delays.

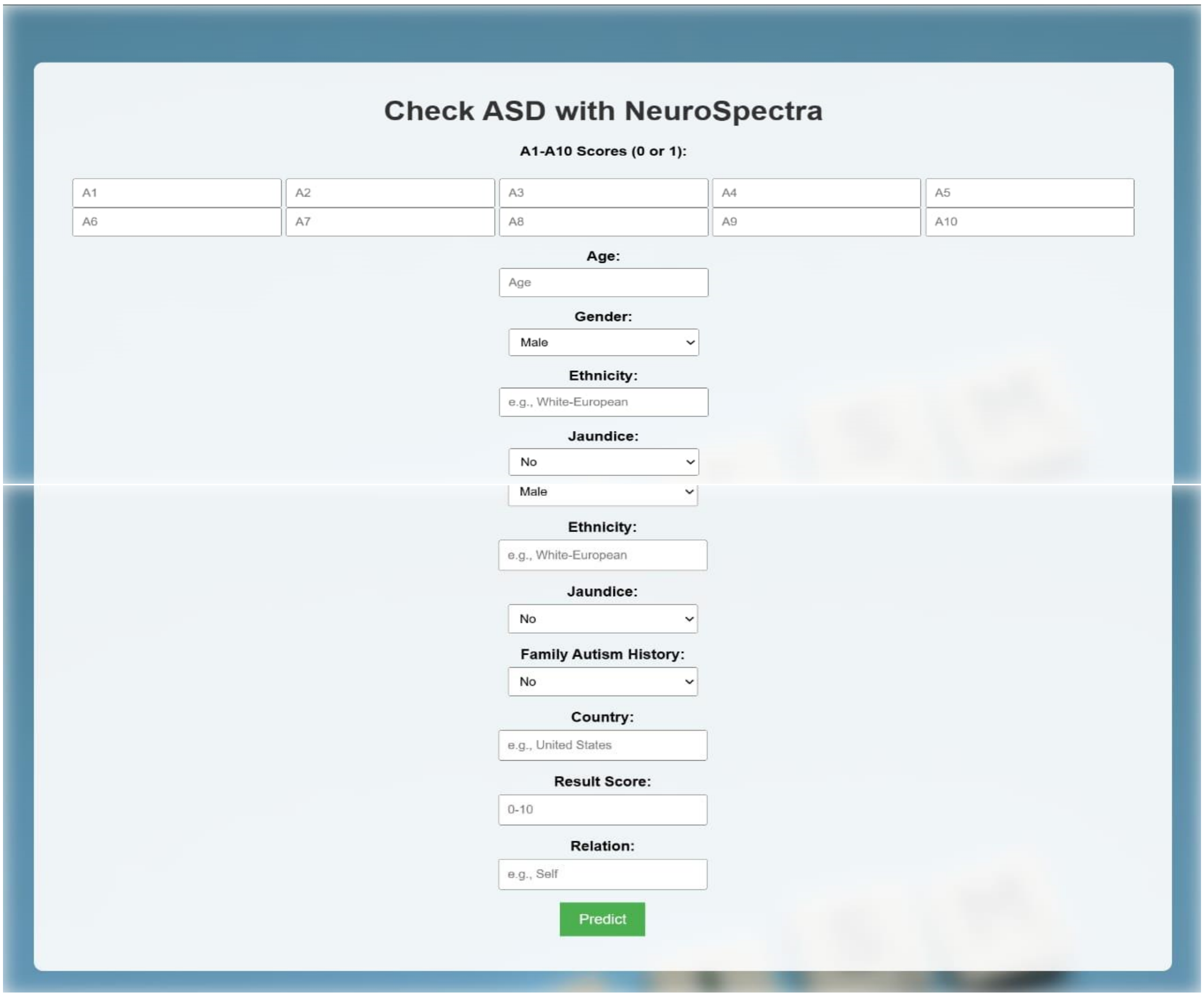


Figure 6.2.4 : Data Input page

Data Input Page : The Data Input Page allows users to enter relevant personal, behavioral, and medical information required for ASD prediction. It includes fields such as age, gender, questionnaire responses, family history, and other diagnostic indicators. The page ensures data validation to maintain accuracy before submitting the input for model evaluation.

Chapter 7

CONCLUSION AND FUTURE WORK

7.1 SUMMARY OF THE WORK

This project focuses on developing a machine learning-based system to detect Autism Spectrum Disorder (ASD) using a combination of behavioral, neurological, and visual features. A structured dataset comprising user responses and medical indicators is preprocessed and fed into multiple supervised learning algorithms, including Support Vector Machine (SVM), Random Forest, Decision Tree, Logistic Regression, Naive Bayes, and K-Nearest Neighbors (KNN). Among these, SVM achieved the highest accuracy and proved most effective for ASD classification. The system allows users to input features via a web interface and receive predictions instantly. It also supports batch processing of datasets for healthcare professionals. The goal is to assist in early screening and support clinical diagnosis by offering a fast, reliable, and scalable AI-driven solution.

This project aims to detect Autism Spectrum Disorder (ASD) using machine learning techniques applied to behavioral, neurological, and visual features. Various algorithms like SVM, Random Forest, Decision Tree, Logistic Regression, Naive Bayes, and KNN were trained and tested, with SVM delivering the best accuracy. The system accepts user inputs through a feature input interface or dataset upload, processes the data, and predicts the likelihood of ASD. It ensures early and cost-effective screening, which can assist healthcare professionals in timely intervention.

7.2 LIMITATIONS OF THE STUDY

- **Limited Dataset Availability:** The model is trained on publicly available datasets, which may not fully represent the diversity of real-world ASD cases across different age groups, cultures, and regions.
- **Binary Classification Only:** The system classifies individuals as ASD or non-ASD, without considering the severity levels or spectrum range of autism, which limits clinical applicability.

- **Lack of Real-Time Clinical Validation:** The model has not yet been validated or tested in real-world clinical settings with live patient data, reducing its immediate use in medical diagnosis.
- **Feature Dependency:** The accuracy of the model heavily depends on the quality and completeness of the input features. Missing or inaccurate data can significantly affect predictions.
- **No Integration of Deep Learning for Visual Cues:** Although some visual features are considered, advanced techniques like CNN for image/video analysis of facial expressions or eye movement are not integrated in this version.
- **Not a Replacement for Medical Diagnosis:** The system serves as a supportive tool and cannot replace professional medical evaluation and diagnosis by certified clinicians.

7.3 SCOPE OF THE FUTURE WORK

In the future, this project can be significantly enhanced by integrating deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to analyze visual and speech-based data for improved ASD detection. The system can be developed into a real-time diagnostic tool, aiding healthcare professionals during consultations. Additionally, the model can be upgraded from binary classification to multi-class classification to identify the severity levels of autism, offering more personalized intervention strategies. Creating a mobile application can further increase accessibility, especially in rural or underserved regions. Expanding the dataset to include more diverse demographic and clinical records will enhance model reliability and generalizability. Furthermore, integration with electronic health records (EHRs) can support automated monitoring and assist in long-term tracking of developmental patterns in individuals at risk of ASD.4

REFERENCES

1. An early-stage automated model for diagnosing and detecting autism spectrum disorders in children using key sociodemographic and family characteristics (2021), Authors : A. S. Albahri, Rula A. Hamid, A. A. Zaidan
2. Detection of autism spectrum disorder (ASD) in children and adults using machine learning (2023), Authors : Cao, X, Cao, J.
3. Autism spectrum disorder detection using machine learning techniques (2024), Authors: Abdelhakim Ridouh, Fayçal Imedjdouben, and Sarra Mahi
4. Early Detection of Autism Spectrum Disorder in Children Using Machine Learning (2024), Authors : Anitha , Deepthi
5. Combining Radiomics and Machine Learning Approaches for Objective ASD Diagnosis: Verifying White Matter Associations with ASD (2024), Authors : Junlin Song, Yuzhuo Chen, Yuan Yao, Zetong Chen, Renhao Guo, Lida Yang, Xinyi Sui, Qihang Wang, Xijiao Li, Aihua Cao, and Wei Li
6. Detection of Autism Spectrum Disorder Using A 1-Dimensional Convolutional Neural Network (2023) Authors: Mohammed M Al-Ani, Ahmed Adil Nafea
7. A Survey on the use of Artificial Intelligence for Diagnosing Autism Through DTI and fMRI (2023) Authors: Eman Helmy, Yaser El Nakieb
8. An early-stage automated model for Diagnosing and detecting Autism Spectrum disorder Strategy using Ensemble diagnosis Methodology based on blood test (2023) Authors : Asmaa H Rabi, Ahmed I Salehn
9. An Update on psychopharmological treatment of Autism Spectrum disorder (2022) Authors: Ramkumar Aishwarya, Tatiana Valica
10. The cost of Autism Spectrum disorders (2014) Authors : Chiara Horlin

Department of Information Science and Engineering

Malnad College of Engineering, Hassan

Publication Details, AY: 2024-25

1	Paper Title	**NEUROSPECTRA: REVOLUTIONIZING AUTISM DETECTION THROUGH AI**	
2	Authors	1	Priyanka H L
		2	Pavan B
		3	Rohithnaik S
		4	S M Yathin
		5	Sinchana S V
3	Journal Name	IJSREM (International Journal of Scientific Research in Engineering & Management)	
4	Status of the Technical Paper	Reviewed	
5	Volume	Volume 09	
6	Year of Publication	2025	
7	ISSN/LSBN	2582-3930	
8	DOI	10.55041/IJSREM46677	

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

NeuroSpectra : Revolutionizing Autism detection through AI

Priyanka H. L , Pavan. B , Rohithnaik. S , S. M Yathin , Sinchana S. V

Malnad College of Engineering, Hassan-573202, Karnataka, India

phl@mcehassan.ac.in, pavanbasavaraj25@gmail.com, rohithnaiks2003@gmail.com, sm.yathin4@gmail.com , janapada147@gmail.com

Abstract: Identifying and Assessing Autism Spectrum Disorder is usually dependent on the behavioral Developments in the human life and the condition known as Autism Spectrum Disorder is an neurological and developmental disorder which occurs during the initial stages of the child life i.e the initial two years of child birth. As per the recent census this a neurodevelopmental condition characterized by challenges in social interaction and communication affects 1% across the entire world's population. While its primary origin lies in genetics, early detection is crucial, and leveraging machine learning offers a promising avenue for a faster and more cost-effective diagnosis. This disorder is mainly caused due to environmental changes it is very essential for the world to detect and diagnose this disorder in the early stages of the life as it brings behavioral changes in the human beings so in order to detect this disorder and to increase the accuracy in the detection of this disorder we apply the AI technologies specifically the predictive models powered by machine learning and Neural networks. In the Machine learning models our approach involves supervised learning algorithms such as the Support Vector-based classifiers Random Forest algorithms to detect the disorder in Adults and We apply convolutional neural networks (CNN) in conjunction with a Recurrent Neural Network (RNN) to detect ASD in Children. Where the machine learning algorithm works on feature selection and the symptoms which are there in the Autism spectrum disorder which includes the behavioural changes which will be present in the individual if they are having autism spectrum disorder.

Keywords : Autism spectrum disorder, Neural network, Machine learning, Feature selection, Supervised learning

I. INTRODUCTION

Over the past few years, the condition known as Autism Spectrum Disorder is very common in the world. As number of individuals who are getting affected by ASD is growing so the demand for proper detection and diagnosis is also increasing rapidly. Detection is the first phase in Autism spectrum disorder usage of accurate machine learning model plays a vital role. Autism spectrum disorder affects the individual to connect in the world. This ASD involves various characteristics including challenges with social engagement and they face difficulty in perceiving or learning skills due to lack of communication this makes an individual to face difficulty in their lives and they face difficulties in tolerating the changes in the environment and changes in their routine. So it is very essential for us to detect this ASD in the early stages of the life thus preventing them to get affected by Autism. The advancements in the technology specially in the field of Artificial intelligence has enabled to detect this ASD by using various algorithms by analysing the patterns in the behaviours of the person who can get affected by this disorder by taking the patterns of the already affected people. AI has excelled in identifying the behavioural changes, Speech recognition as well as the genetic markers that may overcome the difficulties in the traditional approach. AI has the capability of providing more accurate and reliable results in detecting the Autism spectrum disorder.

This paper involves the inclusion of various Artificial technology techniques in order to detect the ASD by reviewing the current methodologies in detecting this disorder and the challenges associated with the current methodologies. As AI has the capacity of analysing interpreting organizing the data which gives more accurate results. By analysing the behavioural changes environmental changes and the genetic markers AI will yield more consistent outcomes by using speech analysis in the adults and the characteristics in human beings the Neural networks and machine learning models are capable of detecting ASD. By integrating AI in detection of this disorder we can get user friendly applications where the parent can also get to know whether the child is suffering from ASD. So This paper revolves under the addition of various emerging AI technologies to detect the neurodevelopmental disorder (ASD). Thus providing more accurate and reliable results.

II. LITERATURE REVIEW

[1] An early-stage automated forecasting system to detect children diagnosed with autism spectrum disorder through key sociodemographic and family-related attributes

Initial automated forecasting model for diagnosing and identifying children individuals with autism spectrum disorders utilizing significant socio-demographic and family-features characteristic traits. The ASD primarily impacts children. This ASD will impact individuals' social lives and their way of living. Numerous global health organizations and centers focusing on autism diagnosis and detection are encountering difficulties in delivering an accurate model for detecting and diagnosing Autism Spectrum Disorder. The data regarding ASD detection is influenced by several unidentified factors of the condition, and a prompt resolution is needed to address these factors for ASD. Thus, enhancing the chances to present proof that 'environmental and genetic factors are the primary indicators of ASD is a scientific challenge that must be addressed. This paper primary goal is to develop a predictive model to detect Autism spectrum condition in children as soon as possible, taking into account familial and social influences. This study adopts a three-phase method. To begin with, this research includes data gathering and processing, and the gaps in the data are addressed through the 1-NN model. The characteristics needed for ASD detection are obtained through the Chi-square and relief techniques. To guarantee equity in training, the dataset is balanced using the oversampling technique for minority class (SMOTE). Eight various machine learning algorithms—Decision Tree, Random Forest, Naïve Bayes, KNN, SVM, Logistic Regression, Adaboost, and MLP—were then tested and trained with the generated dataset. Evaluation metrics, such as Accuracy, Precision, and Recall, F1-score, and AUROC, are then employed to evaluate the model.

The following was shown by the model's results: (1) Out of ten important family and sociodemographic characteristics, seven have been linked to autism cases. (2) A strong positive connection (correlation 0.751) between the father's and mother's ages at childbirth is found by correlation sensitivity analysis. (3) Higher accuracy outcomes of 0.995, 0.9925, 0.9834, and 0.9876 are shown employing machine learning techniques like AdaBoost, K-nearest neighbor, neural networks, and decision tree methods, respectively. Alternative machine learning methods such as logistic regression, Alternatively, models including Random Forest and Naive Baye, provide lesser accuracies of 0.8002, 0.8199, and 0.8297, respectively.

Conversely, with a performance rate of 0.7105, the margin-based classification algorithm (SVM) method shows the lowest accuracy. Based on four evaluation metrics—area under the curve (AUC), F1 score, precision, and recall—the AdaBoost achieves the highest performance, accompanied by values of 0.9999, 0.9995, 0.9995, and 0.9995. The recently balanced and pre processed ASD dataset acts as a resource for autism studies. When compared with the original ASD dataset, the preprocessing methods can be considered accurate and produce better results. The classification precision was much improved by feature-selection techniques with comparable results from Chi2 and Relief. When compared to previous models at different comparing points, the study confirms the performance of the proposed prediction model. This suggested paradigm makes it possible to anticipate autism in its early stages.

[2Recognizing Autism Spectrum Disorder with the help of a One-Dimensional convolutional deep learning model

The neurological disorder known as autistic spectrum disorder, or ASD, has an impact on behaviour, social interaction, and communication. The field of artificial intelligence known as machine learning is devoted to creating algorithms that recognize patterns in input data and classify ASDs. Applying machine learning techniques to classify ASD has produced a variety of results. Further studies are required to enhance the performance of ASD classification. In order To tackle this, the following deep learning techniques, including 1D CNN have been introduced as an alternative to the categorization of ASD detection.

Three different publicly available ASD datasets (children, adults, and adolescents) are used to evaluate the proposed approaches. Because 1D CNNs are better suited to analyzing Time-series data frequently applied in diagnosing autism spectrum disorder, results indicate that they achieve superior to classical machine learning techniques in classifying ASD across all datasets, with higher accuracies of 99.45%, 98.66%, and 90% For the screening of the autism spectrum condition in Adults, Children, and Adolescents, respectively.

[3] The application of artificial intelligence in identifying autism through DTI and fMRI

A Survey of ASD, or neurodevelopmental disorders disorder, includes a number of disorders characterized by difficulties with Speech interactions, repetitive actions, and social behaviours, and linguistic cues. As reported by the Centers for Disease Control(CDC), 1 out of every 44 American children currently suffers from ASD. Clinical behavioural observation tests, which are confidential, time-consuming, and only permit late detection (a kid must be at least two years old to be eligible for an observation report), are the leading method for diagnosing ASD. However, Magnetic resonance imaging, a key neuroimaging tool (MRI) has shown that it can support quick, objective, and early diagnosis and identification of ASD.

Recent advancements in machine literacy (ML) and artificial intelligence (AI) have contributed to the development of appropriate technologies for early detection and automated ASD opinion. Deep literacy (DL), a recent development in artificial intelligence (AI) that relies on artificial neural networks (ANNs), has made it easier to analyze brain MRI data and improved individual capacities for people with ASD. This research focuses on two main MRI types-functional MRI (fMRI) and prolixity tensor imaging (DTI)—to investigate the role of AI in the diagnosis and understanding of autism.

Similarly, the check displays the abnormal DTI and fMRI findings associated with autism. Additionally, new methods for using fMRI and DTI to identify ASD are described and discussed. Future tendencies are eventually described in depth. The results of this investigation show how useful AI is for the early, confidential identification and assessment of ASD. In the future, new AI findings that may be used in healthcare settings may be revealed.

[4] A novel strategy for identifying autism spectrum disorder (ASD) using an ensemble diagnostic method based on blood tests

A New Approach to Identifying Autism Spectrum Disorder through Race and Ensemble Styles Test Information The neurological disorder commonly called autism spectrum disorder (ASD) has an influence on a child's gestures and gregarious communication skills. Common or common signs in early age include repetitive behaviours, limited hobbies, and gregarious commerce. Despite these indications, many people deserve the awareness or comprehension required to recognize ASD early on. Thus, to ensure prompt response and operation, early and accurate discovery using Artificial Intelligence approaches are crucial.
The individual Autism Diapason Complaint (DASD) program is a novel individual path that is introduced in this study. It uses race test data and an ensemble-based AI methodology to swiftly and precisely identify ASD. The Data Preprocessing Layer (DPL) and the individual Subcaste (DL) constitute the two main components of the DASD frame.

Two optimization methods are assumed in the DFL. While the Binary Genetic Algorithm (BGA) is employed to remove erroneous or outlier training data, Binary Gray Wolf Algorithm (BGWA) is used for point election to determine which qualities in the dataset are most relevant. Only important and high-quality data will be applied in the individual phase thanks to this preparation procedure.

To effectively and directly diagnose ASD, the DL employs a novel Ensemble Opinion Methodology (EDM). The Meliorated K-Nearest Neighbors (EKNN) model is an essential components of EDM. It combines three methods: Chimp Optimization Algorithm (COA) to induce synthetic data and decrease the amount of training dataset, Naïve Bayes to transform data from a point room to a weighted room, and For the final opinion, the optimized data is classified using K-Nearest Neighbors (KNN). A racial test dataset for ASD was utilized to determine the DASD program, which was subsequently combined with other current individual techniques. Delicacy (0.93), inaccuracy rate (0.07), recall (0.83), perfection (0.82), micro-average perfection (0.80), macro-average perfection (0.83), micro-average recall (0.79), macro-average recall (0.81), F1-grievance (0.79), and perpetration time (1.53 seconds) were among the numerous interpretation criteria that showed the superiority of the DASD system.

[5] AI-driven classification for recognizing autism spectrum disorder using video analysis

The neurobehavioral disorder known as an autism spectrum condition (ASD) impairs a person's capacity to communicate and engage with others. Additionally, repetitive behaviors and narrow interests are examples of it. Although there isn't a one-size-fits-all approach to autism, early identification and treatment can significantly improve a person's quality of life. Two promising fields of research that could help us better understand autism and undertake better therapies are engine literacy and deep literacy. Artificial intelligence techniques such as engine literacy and deep literacy enable machines to access data without explicit programming**.**

It may be possible to use these models to improve our understanding and capacity to interact with those who have autism. In order to diagnose autism early on, colorful engine literacy techniques are used. Among the engine literacy techniques used in this research area are Brace Vector Machine (SVM), resolution tree, Naïve Bayes, Random Forest, Logistic Regression, and K-Nearest Neighbour. The development of ASD Discovery, which uses engine literacy and deep literacy, has benefited from the vibrant advancements in the fields of engine literacy and artificial intelligence (AI).

The vaticination of autism diapason complaint has been carried out on a videotape dataset in this investigation. The videotape collection includes footage of children with autism and those without the condition engaging in four distinct behaviours. Convolutional Neural Network (CNN) models, such as Inception V3 and Resnet50, have been used to extract the videotape characteristics. These CNN models are trained using long short tenure mind (LSTM) grounded models, and by using this, we obtain 91 accuracy.

[6] Utilizing machine learning methods to identify children with autism spectrum disorder

Early identification of autism spectrum disorder (ASD) is often beneficial to children's long-term health. Because discovery styles rely on the pricey and confidential evaluation of specialists. In order to describe children with ASD, we proposed an engine literacy approach in this work that combines behavioral data (such as eye preoccupation and facial expression) with physiological data (such as electroencephalography, or EEG). Its use can lower prices and improve the efficacy of discovery. First, we used a creative approach to identify the salient characteristics of the EEG data, facial expressions, and ocular preoccupation. Additionally, a mongrel emulsion path with a bracket delicacy of 87.50 was provided for multimodal data emulsion, based on a weighted naive Bayes algorithm.

It effects imply that the engine mastering bracket method in this investigation is effective for identifying ASD early. Distraction matrices and graphs show that EEG may be the most discriminating information, and that ocular obsession, facial expression, and EEG have different discriminatory dominions for the identification of ASD and typically developing children. There are significant reciprocal features between the behavioral and physiological data. Therefore, bracket delicacy can be much improved by the engine literacy approach suggested in this work, which incorporates the reciprocal information.

III. METHODOLOGY

I. Data collection

Data collection is the initial and most important step in developing any Machine learning models. For data collection we use specific age related data by using various tools such as (Kaggle) and We have visited the hospitals to collect the Visual data and Speech data. For children we have collected the Visual data such as the behaviour of a child and the facial expressions, body movements and the Audio data such as speech recordings in order to collect the rhythm, pitch and tone of the specific affected child and the body movements. In Adults we have collected the Clinical data such as responses to the standardized questions related to Autism spectrum disorder and the results of the cognitive tests such as memory and executive functions.

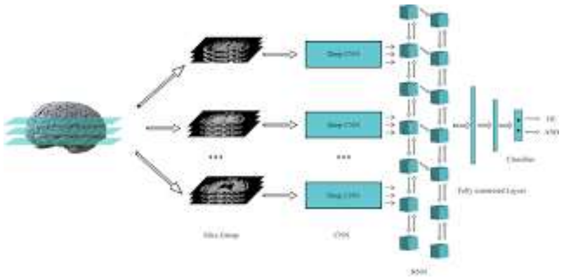
After the data collection it is necessary to preprocess the data which involves various steps such as Cleaning the data in order to remove the noise and next step involves normalizing the data which will sort the redundant data and by normalizing we can achieve standardized scores for the questionnaires.

II. Detection for children : CNN and RNN models

In children the detection is using the Convolutional neural network (CNN) and Recurrent neural netwok

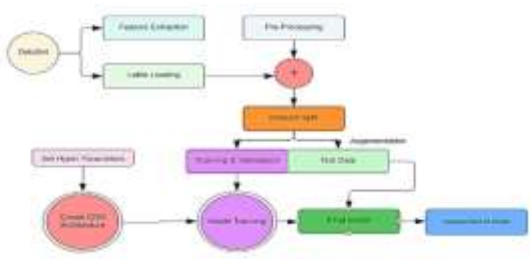
(RNN) where in CNN it includes the pre-processed data such as (Images & Visuals) as input data by which the feature extraction takes place where the features such as Eye deviations, Repetitive movements and the intensity of facial expressions will be extracted then the extracted features will undergo classification under various layers.

RNN will analyse the sequential data such as speech sequences or behavioural sequences. The LSTM (Long short term memory) will helps RNN to identify the irregularities in speech thus feature analysis will take place through Recurrent neural networks.



III. Detection in Adults : Machine learning algorithms.

In Adults as the data is way more structured compared to the data of children so the use of Machine learning algorithms is necessary the supervised learning algorithms such as Support vector machines (SVM) and the Random forest (RF) are used to detect the Autism spectrum disorder in Adults. By using ML models the feature engineering takes place where extracting features such as scores indicating difficulty in social interaction and identifying metrics from cognitive test results such as response time, accuracy etc. As the data is structured the data will be divided into training, Validation and testing sets. Performing the cross validation will help to improve the model to perform in various subsets of the data. Hyperplane tuning helps in model optimization. The final output is model classifies the (ASD-Positive or Negative) based on the input data.



IV. EVALUATION METRICS

The Evaluation metrics involves various aspects such as

- Accuracy: The overall correct predictions
- Precision: Capacity to avoid false positives
- Recall: Capacity to identify true positives
- F1 score: It is the harmonic mean of precision and recall
- AUC-ROC: It is a metric used to evaluate how well a machine learning model can distinguish between positive and negative classes

V. Deployment and Implementation

The implemented model will be deployed as user friendly applications to enhance the adaptability in detecting Autism Spectrum disorder.

For children the implemented model is deployed to a user friendly application that allows Parents to upload the video and Audio thus enhances scalability thus the app provides risk assesments.

For Adults the implemented model is deployed to a web based platform where the user inputs clinical data.

IV. MATHEMATICAL MODEL

To ensure precise operation and control the system relies on mathematical model thus the implement AI technologies will perform with more accuracy. Thus the model involves various equations

1. Convolutional neural network
Given input vector x and a weight matrix W the output z is calculated as:

$$\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b}$$

Where,
W is a weight matrix
b is a bias vector
z is a output vector

2. Support vector machine
Given binary classification of two classes
The equation of the linear hyperplane can be written as:

$$w^T x + b = 0$$

Where W represents normal vector to hyperplane
b represents offset distance of the hyperplane from the origin

The distance between a data point xi and the decision boundary can be calculated as:

$$d_i = \frac{w^T x_i + b}{||w||}$$

where ||w|| represents the Euclidean norm of the weight vector w.

V. DISCUSSION

The inclusion of various emerging AI techniques will help to increase accuracy in Autism spectrum disorder detection. The Various techniques like CNN, RNN and Machine learning models will improve the accuracy of detection of this disorder as it helps to detect Autism in different age groups. Let's discuss the strengths and implications and future work related to the proposed methodology of Autism spectrum disorder detection

Strengths :

The use of CNN & RNN in detecting autism in children will improve the detection of unstructured data dynamically and use of the Machine learning algorithms in Adults will help to increase the accuracy in prediction where the input is structured data as it involves clinical data and results of the cognitive tests. By integrating the visual and audio data in children it provides clear features for detecting this disorder. The deployment of the model into a application it improves accessibility.

Future Work:

Including various other datasets which involves more diverse population to increase generalizability. Creating awareness among the parents about the disorder and involvement of Multimodal input such as combination of video and audio inputs or the combination of visual , Audio and the questionnaire data will increase the efficiency and robustness of the model. Development of detecting capacity which allows parents to constantly notice the behavioural changes in the children.

Implications :

The proposed methodology has the potential ability to provide more efficient, cost effective and accurate detection capabilities than the existing detections. AI based tools which help to increase the efficiency in detecting the Autism spectrum disorder.

VI. CONCLUSION

In conclusion, The AI powered techniques for Autism spectrum disorder detection improves the accessibility, efficiency, and accuracy of the detecting model use of neural networks in detecting Autism in children will help to classify the unstructured data.

The proposed approach will provides a framework which will enhances the autism detection and provides more accessibility and scalable solutions.

In conclusion this research underscores that inclusion of AI plays a pivotal role in detection of Autism spectrum disorder.

VII. REFERENCES

[1] Aythem Khairi Kareem 1 , Mohammed M. AL-Ani 2 and Ahmed Adil Nafea 3 (2023) Detection of Autism Spectrum Disorder Using A 1-Dimensional Convolutional Neural Network

[2] EmanHelmy1, AhmedElnakib2 ,YaserElNakieb (2023) A Survey on the Use of Artificial Intelligence for Diagnosing Autism through DTI and fMRI

[3] A. S. Albahri1 • Rula A. Hamid1,3 • A. A. Zaidan (2021) An early-stage automated model for diagnosing and detecting autism spectrum disorders in children using key sociodemographic and family characteristics

[4] Asmaa H. Rabie and Ahmed I. Salehn (2023) A new diagnostic autism spectrum disorder (DASD) strategy using ensemble diagnosis methodology based on blood test.

[5] Ramkumar Aishworiya1,2,3 · Tatiana Valica1,4 (2022) An Update on Psychopharmacological Treatment of Autism Spectrum Disorder

[6] ChiaraHorlin1,MaritaFalkmer1,2,RichardParsons1 (2014) The Cost of Autism Spectrum Disorders

[7] EmanuelDiCicco-Bloom,1 Lord, C., Zwaigenbaum, L., Courchesne, E., & Dager, S. R. (2006). The Neurobiological Development of Autism Spectrum Disorder

[8] Patricia Howlin, Philippa Moss (2011) Adults with Autism Spectrum disorders

[9] Scott M Myers, Chris Plauche Johnson Management of children with Autism spectrum disorders

[10] Uta Frith, Francesca Happe Autism spectrum disorder

[11] Muhammed Shoaib Farooq, Rabia Tehseen, Maidah Sabir, Zabihulla hatal (2023) Detection of autism spectrum disorder(ASD) in children and adults using machine learning

[12] Vakadkar, K., Purkayastha, D. & Krishnan, D. (2021) Detection of autism spectrum disorder in children using machine learning techniques

[13] Al Banna MH, Ghosh T, Taher KA, Kaiser MS, Mahmud M. A monitoring system for patients of autism spectrum disorder using artifcial intelligence. In: International conference on brain informatics. Cham: Springer; 2020. pp. 251–62