Detection of Autism Spectrum Disorder of Children Using Machine Learning Techniques

Submitted By
Student Name: Kaniz Fatema
Student ID: 221-31-005

This Report in Fulfillment of the Requirements for the Degree of Masters of Science in Electronics and Telecommunication Engineering

Supervised By
Supervisor Name: Mr. MD. Taslim Arefin
Designation: Associate Professor and Head
Department of ETE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECLARATION

I hereby declare that this project has been done by me under supervision of **Md. Taslim Arefin, Associate Professor Department of ETE** Daffodil International University. I also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for the award of any degree or master.

Supervised By:

Md. Taslim Arefin

Associate Professor

Department of ETE

Daffodil International University

Submitted By:

Kaniz Fatema

ID:221-31-005

Department of ETE

Daffodil International University

DECLARATION

This thesis report has been submitted in partial fulfillment of the requirements for a master's degree in the Department of Electronics and Telecommunication Engineering, at Daffodil International University. Therefore, I hereby declare that this report is based on the survey found by me. Materials found in works by other researchers are cited by reference. A thesis report that, in whole, has not previously been submitted for any degree.

Supervised By:

Prof.Mr. MD. Taslim Arefin

Associate Professor and Head

Department of ETE

Daffodil International University

Submitted By:

Kaniz Fatema

ID:221-31-005

Email Address: fatema31-005@diu.edu.bd

Department of ETE

Daffodil International University

DEDICATION

I dedicate this research to Almighty Allah and my beloved father Eng. AB. Kader, who resides in the corners of the universe.

ACKNOWLEDGEMENT CERTIFICATE

First and foremost, I would like to express and convey my gratitude to the almighty God, for his blessing approval, protection, mental power, and wisdom in all aspects of my life, and all applause to Allah of Complete this thesis. The real spirit of achieving a goal is through the way of excellence and austere discipline. I would have never succeeded in completing my task without the cooperation, encouragement, and help provided by various personalities.

This work would not have been possible without the support and guidance of Prof. Mr. MD. Taslim Arefin,Associate Professor and Head, Department of Electronics and Telecommunication Engineering, Daffodil International University, Dhaka,who happens to be my research supervisor for his endless patience, scholarly guidance continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading inferior drafts and correcting them at all stages have made it possible to complete this research. And I choose this topic and developed the research. Deep Knowledge and keep interest of my supervisor in the field of Detection of Autism Spectrum Disorder of Children Using Machine Learning Techniques and influenced me to carry out of this research.

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by a wide range of symptoms that affect an individual's social interaction, communication, and behavior. Early diagnosis and intervention play a crucial role in improving the quality of life for individuals with ASD. This thesis presents a comprehensive investigation into the analysis and detection of ASD using advanced machine learning techniques. The primary objective of this research is to develop a robust and accurate ASD screening tool that can assist in early identification. To achieve this goal, we leverage machine learning algorithms and a diverse range of data sources, including behavioral assessments, clinical records, and demographic information. The research explores the application of supervised learning, feature selection, and data preprocessing techniques to enhance the performance of ASD detection models. This thesis also delves into the development of a prototype application that combines sophisticated machine learning models with an intuitive user interface. The application enables caregivers, educators, and healthcare professionals to conduct preliminary ASD assessments efficiently and receive real-time feedback. Furthermore, the study examines the significance of feature selection and engineering in improving the interpretability of ASD detection models. It explores the potential of neural networks, support vector machines, and other state-of-the-art algorithms in achieving high diagnostic accuracy. The findings presented in this thesis contribute to the growing body of research in the field of autism spectrum disorder detection. The research outcomes have practical implications for early intervention and support for individuals with ASD, ultimately promoting better outcomes and enhanced quality of life. In conclusion, this thesis serves as a valuable resource for researchers, healthcare professionals, and stakeholders in the field of ASD. It underscores the potential of machine learning techniques in advancing the accuracy and efficiency of ASD detection and sets the stage for further research and innovation in this critical area of healthcare.

Contents

Chapter One : Introduction

Autism Spectrum Disorder (ASD) poses a significant challenge in healthcare due to its diverse and complex nature. Early and accurate diagnosis is crucial for timely intervention, yet existing methods exhibit limitations. This research addresses this gap by leveraging the potential of machine learning techniques for ASD analysis. By exploring the intersection of ASD, machine learning, and datasets like ABIDE, we aim to enhance diagnostic accuracy and contribute to the evolving landscape of ASD detection methods. This study not only responds to the pressing need for efficient ASD identification but also aligns with the broader paradigm shift toward personalized and data-driven healthcare solutions.

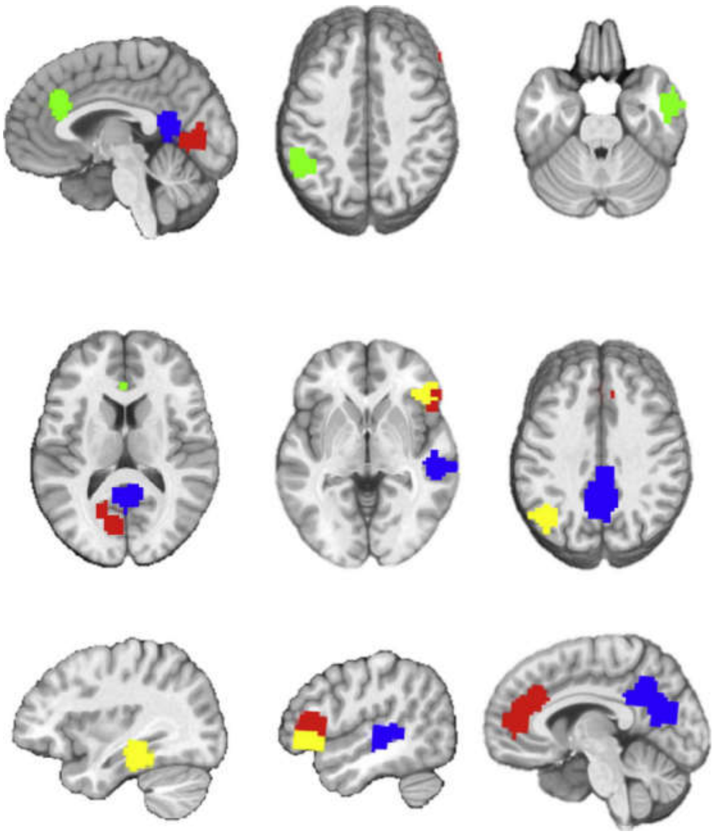


Figure 1: Anti-correlated (underconnected) areas for ASD subjects.

1.1 Background :

1.1.1 concept of Autism Spectrum Disorder (ASD) : The complex neurodevelopmental illness known as autism spectrum disorder (ASD) is characterised by enduring difficulties with speech, social interaction, and repetitive behavioural patterns. People with ASD are affected throughout their lives, beginning in early childhood and extending until adolescence and maturity. Because each person experiences it differently, with varied levels of disability and individual strengths, it is referred to as a "spectrum" disorder.

Due to the observable behavioural patterns and challenges ASD sufferers have in social interactions, ASD is frequently referred to as a "behavioural disorder". A person with ASD often experiences symptoms within the first two years of life, though the degree and presentation may differ. Repetitive motions or behaviours, decreased social engagement, delayed language development, and sensitivity to sensory inputs are among early warning indicators.

Although the precise causes of ASD are not entirely understood, it is thought that a combination of genetic and environmental factors is to blame. Although no single gene has been isolated as the primary cause of ASD, research suggests that some genes may contribute to the disorder's onset. ASD development may also be influenced by environmental factors during the early stages of brain development, such as prenatal exposure to specific chemicals or difficulties during pregnancy or delivery.

Significant difficulties are frequently faced by people with ASD in their daily lives. Lack of eye contact, trouble understanding and expressing emotions, and problems forming and keeping relationships are just a few examples of social interaction difficulties. Language delays and issues understanding and utilising nonverbal communication, such as gestures and facial expressions, are only a few examples of communication difficulties.

People with ASD may engage in repetitive behaviours, stick to routines, and have intense interests in particular subjects or things. Distress and anxiety can be brought on by unexpected events or changes in habit. Individuals who have sensory sensitivities are more sensitive or less sensitive to stimuli like light, sound, taste, or touch.

Supporting people with ASD requires early detection and intervention. Although there is no known treatment for ASD, early intervention programmes and therapies can help to enhance socialisation, communication, and adaptive behaviour, allowing people with ASD to live happy and self-sufficient lives.

Because ASD is so diverse and complicated, identifying autism can be difficult, thus researchers have resorted to machine learning approaches to help identify and analyse ASD. Researchers are working to develop precise and effective ways for predicting and diagnosing ASD, permitting early intervention and better outcomes for people with ASD. They do this by using machine learning algorithms and analysing diverse datasets.

The use of machine learning approaches to detect ASDs is covered in the parts that follow. We'll also examine the methodologies that have been developed so far and compare various models using ASD datasets that are available to the public. The findings of these studies shed light on the potential of machine learning for bettering ASD diagnosis and comprehending the illness in its entirety.

1.1.2prevalence and significance of ASD in today's society : A neurodevelopmental condition known as autism spectrum disorder (ASD) affects people all over the world. For the purpose of recognising its effects and meeting the needs of people with ASD and their families, it is essential to comprehend the prevalence and importance of ASD in contemporary society. An explanation of the prevalence and importance of ASD is provided below:

1.1.2.1Prevalence of ASD: ASD is a reasonably common developmental disease, and during the past few decades, its prevalence has been continuously rising. The Centres for Disease Control and Prevention (CDC) estimate that one in every 54 children in the United States has an ASD. Similar patterns have been seen in other nations, demonstrating the global reach of ASD. The frequency varies by population, and men are more likely than women to receive an ASD diagnosis.

1.1.2.2 Significance of ASD in Today's Society:

- ASD has a substantial impact on both the lives of the affected persons and those of their family. It may result in difficulties with behaviour, communication, social engagement, and sensory processing. In order to reach their full potential and enhance their quality of life, people with ASD frequently need specialised assistance, therapies, and interventions. For those who have the disorder, ASD can affect social interactions, employment chances, and educational success.
- Healthcare and Education Systems: ASD has a huge negative impact on these two sectors of society. Paediatricians, psychologists, and therapists all play an important part in diagnosing ASD, offering early intervention, and managing related physical and mental health issues. To ensure the success and inclusion of people with ASD in the educational system, the education system must take into account their specific learning preferences and support requirements.
- Social Acceptance and Awareness: The prevalence of ASD highlights the need for greater social acceptance and awareness. Increasing knowledge about ASD aids in eradicating stigma, fostering acceptance, and fostering inclusive environments that support those who have the condition. Better chances for people with ASD to engage fully in social, educational, and employment settings are brought about by increased societal acceptance.
- Research and Innovation: Ongoing research and innovation in the subject are motivated by the importance of ASD in contemporary society. Researchers, scientists, and clinicians strive to comprehend the root causes of ASD, create efficient therapies, and enhance diagnostic equipment. Technology advancements like artificial intelligence and machine learning are being investigated to improve ASD analysis, detection, and personalised therapies.
- Economic Impact: ASD has a significant financial impact on people, families, and society as large. The expenses related to ASD include those for carers as well as for medical and therapeutic treatments, educational support, employment assistance, and financial aid. Inclusion, recognising the needs of those with ASD, and offering the right therapies all

help to lessen the financial burden while enabling those with ASD to live productive, happy lives.

For the purpose of creating efficient policies, interventions, and support systems, it is crucial to comprehend the incidence and importance of ASD in contemporary society. We may work to build a more accepting culture that values the talents and qualities of people with ASD by acknowledging the difficulties faced by those with ASD and their families.

1.1.3some of the challenges faced by individuals with Autism spectrum disorder (ASD) and their families:

- Communication difficulties: Both verbal and nonverbal communication are frequently problematic for people with ASD. They could struggle to successfully express their needs and emotions, grasp and use language, read social cues, maintain eye contact, and interpret social cues. This may cause annoyance and make it challenging to develop relationships.

- Social and interactions relationships: People with ASD frequently struggle in social situations. They could struggle to comprehend social conventions, make friends, start and maintain discussions, and take part in group activities. Feelings of exclusion and isolation may result from this.

- Sensory sensitivities: Many people with ASD experience heightened sensitivity to sensory stimuli including loud noises, bright lights, specific textures, or potent odours. Navigating typical situations might become difficult as a result of the sensory overload, discomfort, and anxiety that can follow from this.
- Repetitive actions and constrained interests: People with ASD frequently take part in monotonous actions or have intense, specific interests. These habits may involve hand flapping or other repetitive gestures, following rigorous routines, fixating on particular subjects, or obsessing over particular items. These actions may make it difficult for them to carry out everyday tasks and to participate in a variety of activities.
- Executive functioning issues: People with ASD may struggle with executive functions like organisation, planning, problem-solving, and impulse control. Their capacity to manage time, accomplish work, adjust to changes, and make independent judgements may be impacted by these issues.

- possibilities for education and work: Many people with ASD encounter obstacles in their quest for suitable educational and job possibilities. To suit their particular demands, they could need specialised educational help and accommodations. Due to social and communicative barriers, transitioning to higher education or the workforce can be extremely tough.

- Emotional and mental health issues: ASD patients are more likely to have co-occurring mental health diseases such anxiety, depression, obsessive-compulsive disorder (OCD), and attention deficit hyperactivity disorder (ADHD). Emotional and mental health problems. For family members and carers, the difficulties posed by ASD may also be a factor in their elevated stress levels and mental health problems.
- Access to services and support: For people with ASD, finding the right services and support can be quite difficult. This entails receiving a precise diagnosis, gaining access to specialised therapies (such as speech therapy and occupational therapy), locating accessible learning environments, and using available community resources. Financial limitations, lengthy waiting lists, and a lack of availability might make it difficult to get the care you need.

These are some of the difficulties that people with ASD and their families deal with. A multidisciplinary strategy is necessary to address these issues, and it should include early intervention, specialised education and therapies, social support networks, and the development of inclusive and hospitable surroundings.

1.2 Rationale:

1.2.1 Motivation behind the research project:

The study effort was inspired by the rising incidence of Autism Spectrum Disorder (ASD) and the potential advantages of using machine learning techniques for its diagnosis and analysis. ASD is a neurodevelopmental condition that has a significant impact on a person's ability to communicate and connect with others. It can have a big effect on their daily lives and make it more difficult for them to develop deep relationships.

For those affected with ASD to get prompt intervention and support, early diagnosis of the illness is essential. However, because it depends on clinical observations and evaluations, which are vulnerable to human interpretation and biases, diagnosing ASD can be difficult. Additionally, some ASD symptoms may resemble those of other mental health conditions, making a diagnosis more challenging.

The development of more unbiased and precise methods for diagnosing ASDs has been made possible by the growing use of machine learning techniques in medical research. Machine learning algorithms can learn to discover patterns and predict whether a person exhibits ASD features based on their symptoms and other relevant information by training models on huge datasets and utilising pattern recognition capabilities.

In order to predict and analyse ASD in people of various ages (children, adolescents, and adults), the research project aims to investigate the potential of various machine learning techniques, such as Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network, and Convolutional Neural Network. This study aims to identify which models produce the most accurate findings and may be utilised in real-world contexts by comparing the performance of various algorithms on publically available non-clinical ASD datasets.

The study project's ultimate goal is to contribute to the creation of an effective and efficient ASD screening tool that can aid medical professionals in early detection, resulting in improved treatment approaches and an improvement in the quality of life for people with ASD. The study intends to address the problems with ASD diagnosis and advance the field of autism research and treatment by utilising the potential of machine learning.

1.2.2 The need for early detection and diagnosis of ASD:

Early identification and detection of autism spectrum disorder (ASD) are essential for a number of reasons. For people with ASD and their families, early intervention and support can greatly improve outcomes and quality of life. The importance of early detection and diagnosis can be discussed in connection to the following important points:

- The best course of action is early intervention, which is thought to be very helpful in enhancing developmental outcomes for people with ASD. Early intervention programmes, such behavioural therapy and speech-language therapy, have been shown in studies to improve social interactions, adaptive behaviours, and communication abilities.
- Developmental Milestones: Early detection makes it possible for medical experts to keep a careful eye on a child's developmental milestones and spot any delays or unusual patterns. Early detection of possible ASD symptoms enables therapies to be specialised to tackle particular issues and support optimum development.
- Family Support: An early diagnosis gives families the information and assistance they need to comprehend the special requirements of their child and to take responsible action. Access to the right services, therapies, and educational resources is made possible for families, which can help reduce stress and enhance overall family wellbeing.
- Targeted Education and Accommodations: With an early diagnosis, schools and teachers can put in place specialised educational plans and accommodations to accommodate ASD-affected kids. This entails creating inclusive classroom environments that encourage social interactions and educational possibilities, as well as developing individualised learning plans and behavioural support techniques.
- Access to Services: Prompt access to a variety of services, such as therapy, early intervention programmes, and support groups, is ensured by early diagnosis. Early diagnosis increases the likelihood of receiving the required resources in a timely way because many programmes and interventions have limited availability.
- Enhancing Research and knowledge: Early discovery and diagnosis allow researchers to examine early developmental trajectories and find potential risk factors, which improves

research and knowledge of ASD. This study will contribute to the creation of more focused interventions and advance our understanding of ASD as a whole.

- Emotional and psychological assistance: Early diagnosis offers families who might be unsure or worried about their child's development emotional help. Early access to counselling, support groups, and community resources helps families deal with the difficulties brought on by ASD.

The ability to obtain early intervention, offer focused support, and enhance long-term outcomes for people with ASD and their families depends on the early detection and diagnosis of ASD. Healthcare professionals, educators, and families can collaborate to offer the essential interventions and supports that support the best possible development and quality of life by recognising ASD at an early stage.

1.2.3 The use of machine learning techniques in ASD analysis and detection:

Several factors can support the use of machine learning techniques in ASD analysis and detection:

- Data that is Complex and Multidimensional: Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder with a wide range of symptoms and varying degrees of severity. The multidimensional data related to ASD may be difficult to analyse and interpret using conventional statistical methods. On the other hand, machine learning algorithms can handle huge and complex datasets, extract patterns, and spot relationships that might not be obvious through traditional analysis approaches.
- Relationships between Different Features and the Presence of ASD Are Nonlinear: ASD is Influenced by a Combination of Genetic, Environmental, and Neurological Factors, Leading to Nonlinear Relationships. ASD can be predicted and detected with greater accuracy thanks to machine learning approaches like neural networks that are good at modelling complicated connections between many variables and capturing nonlinear correlations.
- Machine learning algorithms are able to automatically pick out the most important traits from a huge pool of potential variables. When analysing ASDs, where there may be a wealth of demographic, behavioural, and clinical variables accessible, this skill is especially helpful. Machine learning approaches can increase diagnostic accuracy and lessen the computing load associated with high-dimensional data by finding the most crucial aspects.
- Early discovery and Intervention: Prompt ASD intervention and better results depend on early discovery. To find patterns that might point to the presence of ASD, machine learning algorithms can examine a variety of behavioural and clinical indications. Utilising machine learning algorithms makes it possible to create prediction models that

aid healthcare practitioners in spotting patients at risk for ASD early on, allowing for early support and intervention.

- Scalability and Generalizability: Both scalability and generalizability are potential properties of machine learning models. These models can be used to analyse fresh and unexplored data after being trained on a dataset, enabling broader use and flexibility across various populations and circumstances. The creation of tools and systems that allow ASD identification in diverse healthcare situations is made possible by the scalability and generalizability of the technology.

The capacity to handle complicated and multidimensional data, capture nonlinear correlations, choose pertinent characteristics, enable early diagnosis, and give scalability and generalizability make machine learning techniques useful for ASD analysis and detection. These methods could increase the precision, effectiveness, and accessibility of ASD diagnosis, ultimately resulting in better outcomes for people with ASD.

1.3 Research objectives :

1.3.1 main objective of my research: The main objectives of the research are as follows:

- Examine how to use machine learning approaches (Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network, Convolutional Neural Network) for diagnosing and analysingAutism Spectrum Disorder (ASD) in children, adolescents, and adults.
- To assess the effectiveness of the suggested machine learning models on datasets of non-clinical ASD for each age group that are openly accessible (ASD screening data for kids, teens, and adults).
- To analyse the various machine learning algorithms' specificity, sensitivity, and accuracy in foretelling ASD in various age groups.
- Using the data, determine the best machine learning strategy for early diagnosis and screening of ASD in each age group.
- To further the diagnosis and detection of ASD by utilising machine learning techniques and shedding light on the advantages of deep learning-based models (Convolutional Neural Network) over more conventional machine learning algorithms.
- To emphasise the significance of early ASD diagnosis and identification in order to enhance the quality of life and implement improved treatment modalities for ASD patients.
- To discuss the difficulties and limitations of the current ASD screening strategies and to suggest a more effective and precise strategy using machine learning technologies.

The project intends to offer insightful information on the use of machine learning in ASD identification, which may significantly improve early diagnosis and intervention methods for people with ASD.

1.3.2 Goals to achive:

The particular aims of this project are to investigate and assess the potential of machine learning methods for the identification and diagnosis of autism spectrum disorder (ASD). The main goal is to create a predictive model that, using data collected from symptoms and behavioural characteristics, can correctly identify ASD in people of all ages, including children, adolescents, and adults. In order to determine which strategy produces the most accurate and trustworthy results, the study also compares the performance of different machine learning algorithms, such as Naive Bayes, Support Vector Machine, Logistic Regression, K-Nearest Neighbour, Neural Network, and Convolutional Neural Network.

The project will carry out thorough data pre-processing, addressing missing values, and normalisation using publically accessible non-clinical ASD datasets for each age group in order to guarantee the integrity and quality of the input data. The ultimate objective is to create solid models that can reliably identify ASD patients from non-ASD people, assisting in early diagnosis and subsequent treatment for those who have been given the diagnosis. To see if different age groups offer distinct difficulties in diagnosing ASD, the inquiry will also compare the performance of the machine learning models on each dataset.

The study also aims to add to the body of knowledge on the detection of ASD by elucidating the possible benefits and drawbacks of using deep learning techniques, notably Convolutional Neural Networks, as opposed to conventional machine learning algorithms. A thorough understanding of each model's efficiency in screening and diagnosing ASD across various age groups will be made possible by the study of each model's specificity, sensitivity, and accuracy.

In order to help physicians, researchers, and healthcare professionals make better judgements about the diagnosis of ASDs and the timing of early intervention, the study's overall goal is to present insightful findings and recommendations. The findings of this study have the potential to advance the development of effective and precise machine learning-based ASD screening tools, making major contributions to the field of medical diagnostics and enhancing the lives of ASD sufferers.

1.4 Research questions:

1.4.1 specific research questions that guided the investigation:

1. To what extent can Autism Spectrum Disorder (ASD) be predicted and analysed using various machine learning approaches, such as Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network, and Convolutional Neural Network?

2. What are the main obstacles to employing machine learning models for the early detection and diagnosis of ASD, and how may they be overcome?

3. When used with the ASD screening datasets, how do the machine learning models' performances differ amongst children, adolescents, and adults?

4. Which features are the most useful for applying the chosen machine learning algorithms to predict ASD, and how do these factors affect the models' accuracy?

5. How effective is the Convolutional Neural Network (CNN) at detecting Autism Spectrum Disorders when compared to other traditional machine learning classifiers like Naive Bayes, Support Vector Machine, Logistic Regression, KNN, and Neural Network?

6. How can early diagnosis and treatment of autism spectrum disorder be improved in light of the research findings?

1.5 Methodology Overview:

1.5.1 overview of the methodologies and techniques of the research:

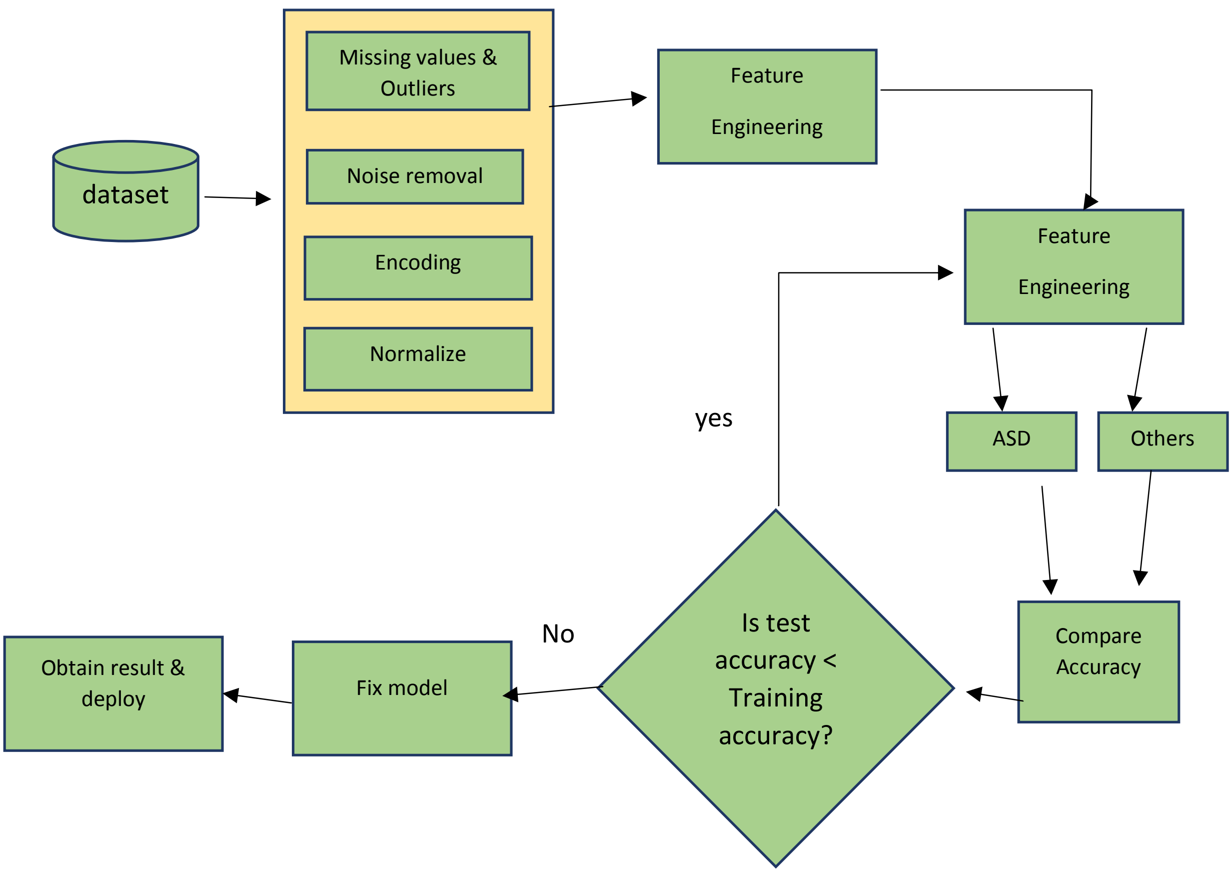


Figure 1.5.1: Architecture of Proposed system

The main goal of this study is to use machine learning approaches to analyse and detect Autism Spectrum Disorder (ASD). The study makes use of three separate, each having 21 traits, non-clinical datasets for the screening of ASD in children, adolescents, and adults. Naive Bayes, Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbours (KNN), Convolutional Neural Network (CNN), and Artificial Neural Network (ANN) are some of the machine learning methods used for analysis and prediction.

1. Data pre-processing: Data pre-processing is the process of converting raw data into a format that is meaningful and comprehensible. To guarantee the quality of the datasets, this involves handling missing values, normalisation, and other data cleaning procedures.

2. SVM (Support Vector Machine): SVM is a machine learning algorithm for supervised learning that is used for classification and regression tasks. It establishes a dividing line between several classifications. For this investigation, the RBF (Radial Basis Function) kernel with a gamma value of 0.1 is used.

3. Nave Bayes: Based on the Bayes theorem, which presumes independence between features, Nave Bayes is a probabilistic classifier. In this study, the Gaussian Naive Bayes variant is employed.

4. CNN (Convolutional Neural Network): CNN is a deep learning approach motivated by the visual processing of the human brain. It excels at tasks requiring image recognition. Convolutional, max pooling, fully connected, and normalisation layers make up the CNN model employed in this study. The activation function is ReLU (Rectified Linear Unit), and the Adam optimizer is employed.

5. LR (Logistic Regression): For binary classification issues, Logistic Regression is a regression technique. It simulates the interaction of a dependent binary variable with a single or more independent variables. The logistic regression model is represented by the sigmoid function.

6. KNN (K-Nearest Neighbours): KNN is an algorithm that may be used for both classification and regression problems. It is straightforward and logical. Based on how closely data points resemble K-nearest neighbours in the training dataset, it categorises the data points. This study used Euclidean distance, the most widely used unit of measurement.

7. ANN (Artificial Neural Network): An ANN is an artificial neural network that consists of many interconnected neurons and is often divided into layers (input, hidden, and output layers). The ANN employed in this study is a feed-forward neural network, which means that data only moves forward. For training, an Adam optimizer with a learning rate of 0.01 and 100 iterations is used.

8. Evaluation measures: A number of evaluation measures, including accuracy, specificity, sensitivity, and confusion matrix, are used to gauge how well the models perform. These measures aid in evaluating how well each algorithm identifies ASD in the three age categories (kids, teens, and adults).

The study emphasises the benefits of utilising CNN-based models for ASD identification by comparing the effectiveness of all machine learning methods on various datasets. The outcomes show how machine learning techniques can be used to analyse and detect ASD early on, leading to better treatment approaches.

1.5.2 Algorithms that will use for analysis and detection of ASD:

1. Naive Bayes (NB): Based on Bayes' theorem, Naive Bayes is a probabilistic classifier. It is frequently used for text classification and is effective with large amounts of data.

Support Vector Machine (SVM) is a potent supervised learning technique used for regression and classification tasks. It locates the best hyperplane for classifying the data points into groups.

3. Logistic Regression (LR): In spite of its misleading name, logistic regression is a linear classifier that is employed in binary classification issues. It simulates the likelihood that an input will fall into a specific class.

4. K-Nearest Neighbours (KNN): KNN is a lazily learning algorithm that is non-parametric. The majority class of a data point's k-nearest neighbours in the feature space is used to classify the data point.

5. Convolutional Neural Network (CNN): CNN is an image categorization deep learning technique. Convolutional layers are used to automatically identify pertinent features from the input photos.

6. Artificial Neural Network (ANN): An adaptable deep learning technique, ANN may be applied to a variety of tasks, including classification. It is made up of layered networks of linked neurons.

1.6 Significance of the study :

1.6.1 Potential impact and contributions of the research:

The study "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques" has the potential to make important advances in the understanding of autism and in the treatment of autistic people. This project intends to address the urgent need for early detection and intervention in people with ASD by examining the application of several machine learning algorithms to predict and analyse ASD in children, adolescents, and adults. The significance of this research rests in its potential to fundamentally alter how ASD is recognised and diagnosed, resulting in earlier and more precise interventions that can greatly enhance the quality of life for those who are impacted by this illness.

The evaluation and comparison of several machine learning algorithms for ASD detection using publicly available datasets is one of the research's major accomplishments. This study provides useful insights into the advantages and disadvantages of each strategy by utilising Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network, and Convolutional Neural Network. Convolutional Neural Network (CNN) was found to be the best model for ASD screening in all three age groups (adults, kids, and teenagers), highlighting the potential of deep learning techniques for ASD diagnosis and opening the door for more advanced and precise screening equipment.

Furthermore, the improvement of the diagnostic procedure, which now relies on observation-based screening, is greatly anticipated with the creation of a machine learning-based model for ASD identification. The suggested approach might help doctors, educators, and parents spot possible cases of ASD early, allowing for prompt assistance and intervention. Early identification is essential since it can improve outcomes and the way ASD symptoms are managed.

Furthermore, the study's use of openly accessible datasets enhances the reproducibility and openness of the results, enabling other academics and professionals to confirm the findings and improve the suggested methodologies. The accessibility of these datasets encourages cooperation and breakthroughs in the study of ASD.

The overall potential significance of this study is substantial. By utilising the strength of machine learning techniques, it provides a fresh method for diagnosing ASDs, improving the precision and effectiveness of screening procedures. The suggested methodology could have significant effects on healthcare settings, educational institutions, and public health efforts by lowering the time and resources needed for observation-based assessments. In the end, the results of this study may help to improve the lives of people with ASD and their families by facilitating earlier identification and tailored interventions, which will improve long-term outcomes and increase societal support for those who live with this neurodevelopmental disorder.

1.6.2 findings that can benefit the field of ASD diagnosis and treatment :

The results of this study have the potential to significantly advance the diagnosis and treatment of autism spectrum disorder (ASD) in a number of ways. This research adds useful insights to improve the accuracy and efficiency of early diagnosis, which, in turn, can result in more timely and appropriate interventions by examining the application of several machine learning algorithms for ASD detection.

First off, the study shows how well various machine learning algorithms, including Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network, and Convolutional Neural Network, predict ASD based on the symptoms and behavioural descriptions of patients. When used on datasets for children, adolescents, and adults with publically accessible non-clinical ASD, these algorithms have demonstrated good accuracy rates. The capacity to identify ASD via machine learning may help in the early diagnosis of the illness, even before obvious symptoms fully appear, allowing for prompt intervention and support for afflicted people.

The second component of the suggested methodology is the inclusion of data pre-processing methods to deal with missing values and guarantee the accuracy of the supplied data. The study makes sure that the machine learning models obtain reliable and resilient information, hence improving the accuracy of the predictions, by resolving data inconsistencies and errors. This meticulous pre-processing method can be used to enhance the performance of machine learning models in numerous diagnostic applications, not just for ASD but for other medical datasets as well.

Additionally, the study analyses the effectiveness of several machine learning algorithms on various datasets for various age groups. It emphasises the benefits of using convolutional neural networks (CNNs) for ASD detection, especially in participants who are adolescents, where the CNN-based model had the best accuracy. The promise of deep learning techniques for the early detection of ASD in particular age groups is highlighted by this particular study, freeing researchers and clinicians to concentrate on more specialised methods for various populations.

The study highlights the necessity to investigate cutting-edge deep learning models for ASD diagnosis and highlights the shortcomings of traditional machine learning techniques. This important realisation stimulates additional study and advancement in the area, supporting the development of AI-driven diagnostic tools for ASD and perhaps other neurodevelopmental diseases.

The consequences of this study's findings go beyond just the diagnosis of ASD, in general. The utilisation of cutting-edge data pre-processing methods and machine learning algorithms can be used as a model for comparable studies in other medical fields, enabling the creation of AI-driven diagnostic tools for a variety of health issues. By enabling early detection, individualised treatment plans, and more successful interventions, this research ultimately adds to the continuous efforts to enhance the lives of people with ASD, ultimately boosting their general wellbeing and quality of life.

1.6.3 contributions of the research to advancement in machine learning applications in healthcare:

This study has the potential to dramatically expand machine learning applications in healthcare, particularly in the area of diagnosis of Autism Spectrum Disorder (ASD). This study tackles a significant healthcare challenge by investigating the use of various machine learning techniques, including Naive Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network, and Convolutional Neural Network, for predicting and analysing ASD in different age groups (children, adolescents, and adults).

It is crucial to identify ASD early in order to offer timely therapies and enhance the quality of life for those who are affected. ASD diagnosis, however, is challenging since it entails studying behavioural patterns and identifying certain symptoms that could be confused with those of other mental health conditions. The suggested machine learning models have the potential to increase the precision and effectiveness of ASD diagnosis, which could result in early interventions and better patient outcomes.

The study lays the groundwork for future research in the area of ASD detection by utilisingpublically accessible non-clinical ASD information. The study's concentration on three separate age groups enables a more thorough comprehension of the disorder's presentations during various life stages. Additionally, comparing various machine learning algorithms offers important insights into the most efficient and precise methods for ASD screening.

Deep learning techniques can now be used in healthcare applications thanks to the implementation of Convolutional Neural Network (CNN) based models, which showed greater accuracy than traditional machine learning classifiers. The ability of CNNs to accurately diagnose ASD in a range of age groups may open the door to similar methods for other medical diagnostics, perhaps bringing about advancements in automated disease detection and prognosis.

In conclusion, by enhancing the capabilities of ASD diagnosis by the use of varied and cutting-edge algorithms, this research makes a contribution to the field of machine learning applications in healthcare. The results of this study not only have implications for better ASD screening, but they also provide guidance and inspiration for future research on the successful use of machine learning techniques in healthcare. The use of machine learning models in healthcare could ultimately result in earlier diagnosis, more individualised therapies, and improved patient outcomes.

Chapter Two :Literature review

2.1 Definition and Diagnostic Criterai of Autism Spectrum Disorder

A complicated neurodevelopmental illness known as autism spectrum disorder (ASD) is characterised by a variety of behavioural, social, and communication difficulties. The term "spectrum" in the context of ASD refers to the range of symptoms and degrees of severity that people with the condition may encounter. The symptoms of ASD often first manifest in early childhood, frequently before the age of three. ASD is a lifelong disorder.

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) issued by the American Psychiatric Association contains the most recent guidelines for diagnosis. The diagnostic criteria for ASD have changed throughout time. A detailed framework for diagnosing ASD based on particular behavioural and communicative characteristics is provided by the DSM-5 criteria.

The primary features of ASD include:

2.1.1. Social Communication and Interaction Impairments:

- Difficulties comprehending and utilising nonverbal cues like body language, gestures, and facial expressions.
- Difficulties establishing and maintaining relationships with peers and adults that are age-appropriate.
- A lack of or a limited reaction to social cues, making it challenging to share feelings or interests with others.

2.1.2. Restricted, Repetitive Patterns of Behaviour:

- Repeating particular sentences or engaging in repetitive hand flapping or rocking gestures.
- A strong preference for routines over change. Distress might result from any usual interruptions.
- Extremely constrained interests that frequently place a heavy emphasis on one thing or area of study at the expense of others.

Individuals with ASD might differ greatly in the intensity of these characteristics, which affects their functional abilities and support requirements. The DSM-5 categorises ASD into three support levels as a result:

Level 1: Needing Assistance

People with Level 1 ASD have a few minor social communication and interaction issues. They could struggle with starting and keeping up discussions, as well as starting social connections. They may, however, have fewer constrained interests and frequently respond well to social routines.

Level 2: Needing Significant Support

Social contact and communication are moderately impaired in people with Level 2 ASD. They could exhibit pronounced difficulty starting social contacts and might respond sporadically to other people's attempts to engage them. These people could exhibit more overt repeated behaviours and might need a lot of assistance just to get through each day.

Level 3: Needing Very Significant Support

People with Level 3 ASD have a lot of trouble interacting with others and communicating with them. They frequently rely primarily on nonverbal means of communication because to their weak verbal communication abilities. These people find it extremely difficult to build relationships, and they could act in very strict and repetitious ways. Usually, they need a lot of assistance with all aspects of everyday life.

It is crucial to understand that an extensive evaluation is necessary for the diagnosis of ASD. This evaluation is normally carried out by a group of medical experts, including developmental paediatricians, psychologists, speech-language pathologists, and occupational therapists. In addition to using standardised screening tools and evaluations, the diagnosis takes into account the person's developmental history, observations of behaviour, communication, and interactions.

For the purpose of launching relevant interventions and support services catered to the specific requirements of the individual, an early and precise diagnosis of ASD is essential. As a result of improvements in machine learning techniques, there are encouraging chances to help with ASD early identification and diagnosis. This could result in more effective therapies and better outcomes for people with ASD and their families.

2.2 Conventional Methods for ASD Diagnosis

The diagnosis of autism spectrum disorder (ASD) has traditionally depended on conventional methods that include observations, interviews, and standardisedbehavioural evaluations. Clinicians, psychologists, paediatricians, and other healthcare professionals with experience in the diagnosis of autism frequently use these techniques. Although these methods have been the norm for diagnosing ASD, they have a number of drawbacks and difficulties.

1. Clinical Observations: Clinical observations entail a comprehensive analysis of a child's social interactions, communication abilities, and behavioural tendencies in a variety of contexts. Clinicians frequently watch the child's play, interaction with others, eye contact, and response to social cues.

- Restrictions: Observations may be biassed and subjective in nature. Children may also behave differently in clinical settings than they do in their natural situations, which could result in differences in the diagnosis.

2. Parent and Caregiver Interviews: - Interviews with parents and other carers are crucial to the diagnosis of ASD. Detailed information regarding the child's social interactions, communication styles, and developmental milestones should be gathered from the parents in order to better understand the behaviour of the child.

- Limitations: Parental recollection bias could skew the truthfulness of the data gathered through interviews. Furthermore, cultural and linguistic differences may affect how parents interpret and describe their child's behaviour, which may result in misunderstandings.

2.2.3. Developmental and Behavioral Assessments:

- Standardised tests are frequently used to diagnose ASDs, including the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R). These tests analyse the child's behaviour based on observable traits according to predetermined rules and scoring criteria.

Limitations: Standardised tests can be expensive, time-consuming, and difficult to give without particular expertise. Additionally, a child's age, cognitive capacity, and behavioural presentation may all affect how accurate the diagnostic results are.

2.2.4. Multidisciplinary Team Evaluation:

- A multidisciplinary team made up of clinicians, psychologists, speech-language pathologists, occupational therapists, and educators frequently diagnoses ASD. This coordinated strategy seeks to take into account numerous facets of the child's growth and behaviour.

- Limitations: Multidisciplinary assessments offer a thorough analysis, but they can be time-consuming and logistically difficult. Additionally, access to diagnosis may differ depending on the region's access to specialised doctors.

2.2.5. Diagnostic Criteria:

-The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) contains the most recent version of the diagnostic criteria for ASD. Deficits in social communication, confined and repetitive behaviours, and the severity of ASD are all specifically defined by DSM-5.

- Restrictions: The DSM-5 offers a standardised framework for diagnosis, but because it is categorical, it might not adequately account for the variety of ASD. The symptoms of certain people may coincide with those of other neurodevelopmental diseases, making differentiating diagnosis difficult.

2.2.6. Diagnosis in Adults and Adolescents:

- Because of the intricacy of the symptoms and the overlap with other mental health issues, diagnosing ASD in adults and adolescents can be extremely difficult. In these situations, clinical interviews and analyses of past behaviour from childhood are frequently used.

Limitations: Delaying diagnosis when ASD is discovered in adults and adolescents may prevent timely access to the right interventions and support.

Traditional methods have usefulness, but their shortcomings highlight the need for more accurate and effective ways to ASD diagnosis. Promising solutions for overcoming these obstacles and enhancing the precision and timeliness of ASD detection are provided by machine learning techniques. In order to develop early detection and intervention options for people with ASD, the following chapters examine the application of machine learning algorithms to ASD screening data for various age groups.

2.3 Machine Learning Techniques in ASD Diagnosis:

Healthcare and medical research are two fields where machine learning has become a potent tool. Researchers have been investigating the potential of machine learning algorithms to aid in the diagnosis and prognosis of medical diseases, such as Autism Spectrum Disorder (ASD), more and more in recent years. Compared to conventional diagnostic methods, these techniques provide a number of benefits, including improved objectivity, efficiency, and the capacity to handle huge and complicated datasets.

2.3.1 Introduction to Machine Learning

In the context of ASD diagnosis, machine learning algorithms can be trained on large datasets of patient symptoms and related information. The algorithms learn patterns and relationships within the data, enabling them to recognize patterns that distinguish between individuals with ASD and those without.

Large datasets of patient symptoms and related data can be used to train machine learning algorithms for ASD diagnosis. The algorithms recognise patterns that differentiate between people with ASD and those without it by learning patterns and correlations within the data.

2.3.2 Supervised Learning Algorithms

In supervised learning, the algorithm is trained using labelled data, which means that the input data contains both the features (symptoms) and the target labels (ASD or non-ASD). These algorithms can be used to predict the label of fresh, unused data once they have learned to map the input features to the proper labels.

2.3.2.1 Naïve Bayes (NB)

Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem. It assumes that the features are conditionally independent given the class label, which simplifies the computation and makes it computationally efficient. In the context of ASD diagnosis, Naïve Bayes can learn the probabilities of different symptoms given the presence or absence of ASD, making it suitable for binary classification tasks.

2.3.2.2 Support Vector Machine (SVM):

Support Vector Machine is a powerful supervised learning algorithm used for both classification and regression tasks. SVM aims to find the hyperplane that best separates the data points of different classes with the largest margin. In the context of ASD diagnosis, SVM can be used to find the optimal boundary that separates individuals with ASD from those without, based on their symptoms.

2.3.2.3 Logistic Regression (LR)

Logistic Regression is a popular linear classification algorithm used to predict binary outcomes. It models the relationship between the features and the probability of the target label (ASD or non-ASD) using a logistic function. Logistic Regression is interpretable and can provide insights into the contribution of individual features to the classification outcome.

2.3.2.4 K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a simple yet effective supervised learning algorithm used for classification tasks. It assigns a label to a new data point based on the majority class of its K nearest neighbors in the feature space. KNN is non-parametric and can handle complex decision boundaries, making it useful in scenarios where the underlying data distribution is not well-defined.

2.3.2.5 Neural Networks

Neural Networks are a class of deep learning algorithms inspired by the structure and function of the human brain. They consist of multiple layers of interconnected nodes (neurons) and can learn complex representations from data. In the context of ASD diagnosis, neural networks can automatically learn relevant features from the symptoms and discover intricate patterns that may not be evident in traditional diagnostic approaches.

2.3.3 Unsupervised Learning Algorithms

Unsupervised learning is another type of machine learning, where the algorithm is trained on unlabeled data and aims to find patterns and structures in the data without explicit target labels. In the context of ASD diagnosis, unsupervised learning can be useful for exploring patterns and subgroups within the ASD population.

2.3.3.1 Clustering Algorithms

Clustering algorithms group similar data points together based on their similarity in feature space. These algorithms can help identify distinct subgroups within the ASD population, which may have different symptom profiles or response to interventions. Clustering can provide valuable insights into the heterogeneity of ASD and aid in personalized treatment strategies.

2.3.4 Deep Learning Algorithms

Deep learning is a specialized branch of machine learning that involves training neural networks with multiple hidden layers, known as deep neural networks. Deep learning has shown remarkable success in various domains, including computer vision, natural language processing, and medical image analysis.

2.3.4.1 Convolutional Neural Networks (CNN)

Convolutional Neural Networks are a type of deep learning architecture particularly effective in analyzing visual data, such as images. In the context of ASD diagnosis, CNNs can be applied to extract meaningful features from brain imaging data, enabling better understanding of the neural basis of ASD.

2.3.4.2 Recurrent Neural Networks (RNN)

Recurrent Neural Networks are designed to process sequential data, making them suitable for time-series data or sequences of symptoms. RNNs can capture temporal dependencies in the symptom profiles of individuals with ASD, leading to more accurate and dynamic predictions.

2.3.5 Hybrid Approaches

Hybrid approaches combine multiple machine learning algorithms to leverage their complementary strengths. For example, an ensemble method like Random Forest can combine the predictions of multiple decision trees to improve overall classification performance. Hybrid approaches can enhance the robustness and reliability of ASD diagnosis models.

2.3.6 Feature Selection and Dimensionality Reduction

In ASD diagnosis, the selection of relevant features from the symptom data is crucial to build accurate and efficient models. Feature selection techniques help identify the most informative symptoms that contribute significantly to ASD classification. Dimensionality reduction techniques, such as Principal Component Analysis (PCA), can reduce the number of features while preserving the most relevant information.

In conclusion, machine learning methods provide a variety of tools for ASD diagnosis and symptom data analysis. The algorithm chosen will depend on the data's properties, the study's goals, and the particular difficulties in diagnosing ASDs. Researchers may be able to create more precise, effective, and individualised methods for ASD detection and intervention by utilising machine learning. To guarantee the accuracy and clinical utility of the suggested models, it is crucial to rigorously validate and interpret the findings.

2.4 Feature Selection and Data Pre-processing

To provide precise and reliable predictions, machine learning models for ASD diagnosis strongly rely on the selection of pertinent variables and high-quality data. In this section, we go through the significance of feature selection and data pre-processing procedures in the context of machine learning models used for ASD diagnosis.

2.4.1 Importance of Feature Selection in ASD Diagnosis

A crucial stage in creating machine learning models for diagnosing ASDs is feature selection. It entails selecting the dataset's most insightful and distinctive traits (symptoms) that have a substantial impact on the classification process. The following factors contribute to feature selection's significance:

a. ASD screening datasets frequently include a vast number of features, and not all of them may be useful for identifying those with ASD from those who do not. The addition of pointless or redundant features can result in overfitting, higher computational complexity, and poorer model performance. Techniques for feature selection help limit the number of dimensions in the data by keeping only the most important features.

b. Improved Model Performance: By concentrating on the most insightful features, feature selection can raise the model's precision, sensitivity, and specificity. A model that has been trained using fewer important features would probably generalise to new, unobserved data more successfully, producing predictions that can be trusted.

c. Interpretability: In order to understand how a diagnosis is made in medicine, interpretability is essential. In order to better understand the factors impacting the ASD diagnosis and to provide more relevant explanations for the predictions, clinicians and researchers can benefit from choosing a subset of interpretable traits.

d. Quicker Training and Inference: Machine learning models with a smaller feature set train and infer faster and with less computational resources. This effectiveness is crucial in real-time clinical situations where prompt and precise diagnoses are necessary.

2.4.2 Common Feature Selection Techniques

Several feature selection techniques are commonly used in ASD diagnosis using machine learning models. Some of these techniques include:

a)Recursive feature elimination (RFE): RFE is an iterative feature selection technique that begins with all features and systematically eliminates the least significant ones based on their ranking. Up until a predetermined number of features or a performance criterion is reached, the process is repeated. RFE aids in determining the ideal feature subset that produces the greatest model performance.

b) Information Gain: Based on the idea of entropy, Information Gain is a feature selection method. Each feature's contribution to the target variable's (ASD or non-ASD) knowledge is measured. For the classification challenge, features with a higher information gain are prioritised.

c) Mutual Information: This information-theoretic metric is used to express how dependent the goal variable is on the features. It evaluates the amount of knowledge one variable (feature) imparts to another (target) and vice versa. High mutual information features are taken into consideration while classifying features.

d) L1 Regularisation (Lasso): During the training phase, L1 regularisation includes a penalty term based on the absolute values of the model's coefficients. As a result, the model is encouraged to be sparse, which causes some coefficients to be driven to zero and triggers automatic feature selection. Features with coefficients that are not zero are thought to be crucial for classification.

2.4.3 Data Pre-processing Steps

A important step in developing machine learning models for ASD diagnosis is data pre-processing. It entails converting unclean, raw data into an appropriate format, cleaning the data, and getting it ready for analysis. To guarantee data quality and model accuracy, the following data pre-processing procedures are crucial:

a) Handling Missing Values: Missing values may be present in real-world ASD screening datasets for a variety of reasons, including incomplete medical records or survey responses. Missing data can have a negative impact on model performance and training. The observed data are utilised to impute missing values using techniques like mean impute or regression impute.

b) Normalisation: This process ensures that all features are scaled equally. The model may favour characteristics with higher values since different attributes may have differing ranges and units. Z-score normalisation and Min-Max scaling are popular normalisation methods.

c)Outlier Detection: Detecting outliers, or data points that dramatically differ from the rest of the data, is step c. Outliers can skew the model's learning and have a negative impact on the model's performance. Outliers can be found and dealt with using a number of statistical techniques, including the Z-score and the Interquartile Range (IQR).

d) Feature engineering: In order to increase the predictive potential of current features, new features may be created or existing features may be modified. The selection and development of pertinent features that might not be present in the original dataset can be guided by domain expertise and insights into the ASD diagnostic procedure.

e) Data Splitting: The dataset is often divided into training, validation, and testing sets before to model training. The validation set is used to adjust hyperparameters and prevent overfitting, the training set is used to train the model, and the testing set is used to assess the model's performance on untested data.

Conclusion: When creating machine learning models for diagnosing ASD, feature selection and data pre-processing are essential processes. These processes make sure that the data is of the highest quality and is properly prepared for analysis, and that the models are trained on the most pertinent attributes. Proper feature selection and data pre-processing result in ASD diagnostic models that are more precise, comprehensible, and clinically applicable, supporting early identification and ASD patient care.

2.5 Related Work on ASD Diagnosis using Machine Learning

Machine learning methods have drawn more and more attention as a means of diagnosing ASDs throughout time. In this section, we review pertinent papers and studies that have used machine learning to diagnose ASDs, evaluate their approaches, datasets, and performance measures, and highlight the advantages and disadvantages of prior work.

2.5.1. Study by Vaishali R. and Sasikala R. [3]:

- Methodology: This study proposed a method to identify Autism with optimum behavior sets using an ASD diagnosis dataset with 21 features obtained from the UCI machine learning repository. They experimented with swarm intelligence-based binary firefly feature selection wrapper.

- Dataset: The dataset used in this study contains 21 attributes related to ASD screening.

- Performance Metrics: The proposed method achieved an average accuracy in the range of 92.12%-97.95% with optimum feature subsets.

2.5.2. Study by Fadi Thabtah [8]:

- Methodology: This study presented an ASD screening model using Machine Learning Adaptation and DSM-5. The researcher discussed the ASD Machine Learning classification with their pros and cons and compared the problem accompanying existing ASD screening tools.

- Dataset: The study used an ASD screening dataset with multiple features related to ASD diagnosis.

- Performance Metrics: The performance metrics were not explicitly mentioned in the summary.

2.5.3. Study by M. S. Mythili and A. R. Mohamed Shanavas [13]:

- Methodology: This study aimed to detect the autism problem and the levels of autism using Neural Network, SVM, and Fuzzy techniques with WEKA tools to analyze students' behavior and social interaction.

- Dataset: The dataset used in this study includes behavioral data from students.

- Performance Metrics: The performance metrics for the classification models were not explicitly mentioned in the summary.

2.5.4. Study by J. A. Kosmicki et al. [14]:

- Methodology: This study employed a machine learning approach to evaluate the clinical assessment of ASD using the ADOS (Autism Diagnostic Observation Schedule). They used 8 different machine learning algorithms and stepwise backward feature identification to identify ASD risk.

- Dataset: The study used a subset of behaviors from the ADOS in children based on the autism spectrum.

- Performance Metrics: The study achieved an overall accuracy of 98.27% and 97.66% for two different modules using machine learning algorithms.

2.5.5. Study by Baihua Li et al. [11]:

- Methodology: This study used machine learning classifiers to detect autistic adults by imitation method. The researchers investigated the diagnostic classification of autism using kinematic parameters and machine learning methods.

- Dataset: The dataset contains kinematic constraints from 16 ASC participants who performed a series of hand movements.

- Performance Metrics: The sensitivity rates achieved by RIPPER were reported for different features.

2.6 Strengths of Previous Research:

The research investigate the application of different machine learning techniques, including Naive Bayes, SVM, Neural Networks, KNN, and CNN, offering a wide range of models for ASD diagnosis.

- Several investigations used publicly accessible datasets, enabling replication and outcome comparison.

- Some studies used feature selection strategies to improve the models' performance and interpretability.

2.7 Limitations of Previous Research:

- Some research' use of relatively small datasets may limit how broadly the findings may be applied.

- It can be difficult to assess the efficacy of various models because some studies do not provide comprehensive reporting of performance measures.

- It is not always obvious whether the datasets are indicative of the diversity in populations, locales, and cultural backgrounds.

In conclusion, earlier research on the accurate and early detection of ASD using machine learning has showed encouraging findings. The area still has problems, though, like the need for more and more varied datasets, consistent reporting of performance measures, and improved model interpretability integration. These issues should be addressed in future research in this field to create more reliable and therapeutically useful machine learning models for ASD diagnosis.

2.8 Summary of the Literature Review

2.8.1. Machine Learning Methods: A number of machine learning methods, including Naive Bayes, Support Vector Machine (SVM), Neural Networks, K-Nearest Neighbours (KNN), and Convolutional Neural Networks (CNN), have been investigated for the diagnosis of ASD. Each of these techniques has demonstrated encouraging results in precisely identifying ASD in various age groups.

2.8.2. Feature Selection: The performance and interpretability of machine learning models for ASD diagnosis are greatly enhanced by feature selection. Research has used methods like Recursive Feature Elimination (RFE) and Information Gain to choose the most pertinent features, creating classifiers that are more effective and efficient.

2.8.3. Data Pre-processing: To guarantee the calibre and accuracy of the models, data pre-processing procedures, such as handling missing values, normalisation, and outlier detection, are crucial. The effectiveness of machine learning algorithms is impacted less by noisy and incomplete data thanks to effective data pre-processing.

2.8.4. Performance Metrics: Accuracy, specificity, and sensitivity are just a few of the performance metrics that have been used to assess ASD diagnostic models. These measures shed light on how successfully the models distinguish between cases with and without ASD.

2.8.5. Dataset Availability: For their experiments, many studies used publicly accessible datasets from the UCI Machine Learning Repository. Some databases are relatively limited in size, and it is difficult to ensure that they are representative of different demographic groups and geographical areas.

2.8.6. Current State and Gaps: The research on machine learning-based ASD diagnosis at this time indicates that there has been a substantial advancement in the creation of precise models. Large-scale and diversified datasets, standardised evaluation measures, and the inclusion of interpretability into complicated models are still in need of improvement.

By providing a framework for the creation of an ASD diagnosis model utilising machine learning techniques, the literature review contributes to the understanding of the research aims. The main conclusions stress how crucial feature selection, data pre-processing, and the selection of suitable machine learning algorithms are. It also points out the necessity of addressing the shortcomings of earlier studies, including dataset size and representativeness.

The gaps in the body of literature highlight how important it is to advance the subject by using bigger and more varied datasets, standardised evaluation criteria, and interpretability methodologies. The proposed research intends to provide an accurate and understandable model for ASD diagnosis by expanding on the findings of earlier studies, assisting in early detection and better treatment approaches.

The machine learning techniques and features used for the proposed study are also guided by the literature review. It aids in determining the advantages and disadvantages of various strategies, allowing the research to address difficulties and potential biases. Overall, the literature evaluation is an essential foundation for the research, ensuring that it is pertinent, creative, and advances the use of machine learning techniques for ASD diagnosis.

Chapter Three : Methodology

This chapter describes the approach utilised to accomplish the goals of this study, which include creating a machine learning model that is precise and easy to understand for the diagnosis of autism spectrum disorder (ASD). Data gathering, feature selection, data pre-processing, model development, evaluation, and validation are all included in the methodology. Each stage is meticulously planned to guarantee the accuracy and efficacy of the suggested ASD diagnosis methodology.

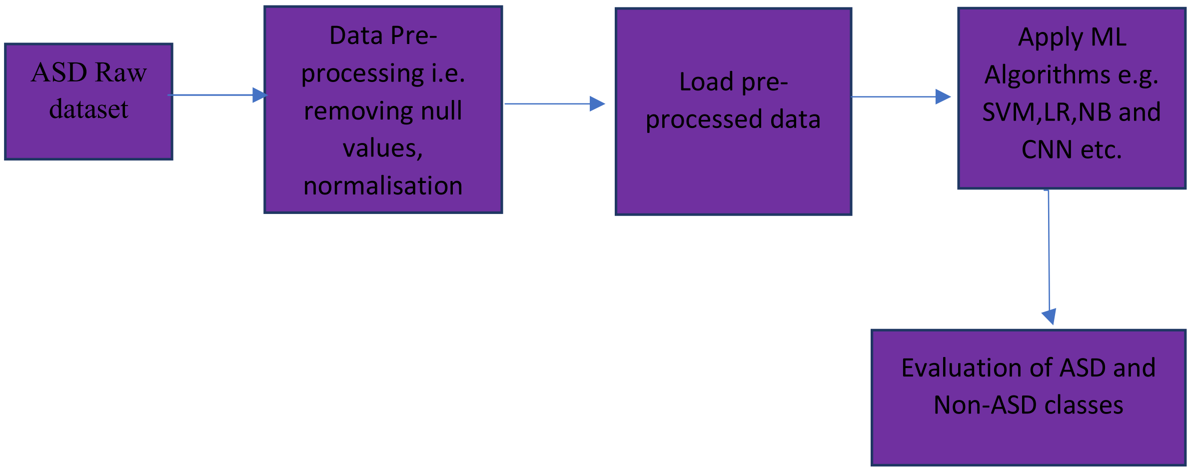


Figure 3: steps in the proposed ASD detection solution

3.1 Data Collection

Collecting appropriate datasets that may be used to train and test machine learning algorithms is the first stage in creating the ASD diagnosis model. Publicly accessible ASD screening datasets from reliable sources will be used to guarantee transparency and repeatability of the findings. The UCI Machine Learning Repository, a well-known repository for machine learning datasets, will be the primary focus of this study.

Three different age groups—children, adolescents, and adults—will be represented in the datasets that are chosen. This strategy is essential since ASD can present itself in numerous ways throughout life, and a thorough model should be able to correctly diagnose ASD across all age groups. The algorithm may be able to recognise patterns and traits that are unique to each stage of life by incorporating datasets from several age groups.

ASD screening and diagnosis-relevant features will be present in the datasets. Key elements that will be present include:

1. Patient Age: Given that symptoms of ASD might differ based on a person's developmental stage, age is a crucial element in making a diagnosis.

2. Sex: Gender-related differences in ASD have been noted, and adding sex as a component can assist the model take these possible variances into consideration.

3. Nationality: Nationality is an important characteristic since geographic and cultural factors may have an impact on the frequency and expression of ASD.

4. Family History of Pervasive Developmental illnesses: An increased risk for ASD in an individual may be indicated by a family history of ASD or associated illnesses.

5. Use of Screening Applications: Knowing whether a patient has utilised any ASD screening tools can reveal how aware they are of potential symptoms.

6. Screening Test Types: Given that different screening test types may provide a range of outcomes, mentioning them will help clarify the applicability and precision of the screening procedure.

By including these traits, the model will be able to gather pertinent data from screening datasets and use that data to generate smart diagnoses.

The privacy and confidentiality of the people represented in the datasets will be protected by thorough processing and anonymization of all data utilised in this study. To safeguard the patients' right to privacy, any sensitive or individually identifying information will also be deleted or encoded.

The research will abide by the ideals of openness and transparency by using publicly accessible datasets from reliable sources, enabling other researchers to confirm and replicate the findings. Additionally, the generalizability of the model will be improved by the incorporation of multiple datasets spanning various age ranges, making it applicable to a wider population affected by ASD. In the end, the data gathering phase contributes to the development of an accurate and reliable ASD diagnosis model utilising machine learning techniques. Subsequent research processes include data preprocessing, model training, and evaluation.

Example:

Here's a simplified example of a dataset for a Bangladeshi 4-year-old boy who has tested positive for Autism Spectrum Disorder (ASD) and has a history of jaundice. Please note that this is a very simplified and fictional dataset for demonstration purposes:

Patient age	Sex	Nation-ality	jaundice	Family history	Experiment fulfilment	country	Screening application	Screening test type	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q 10	Screening score
4	male	Bangla-deshi	yes	no	parent	Bangla-desh	yes	M-CHAT	Y	N	Y	N	Y	Y	N	Y	N	N	7

Figure 3.1

In this example:

Patient Age: 4 years

Sex:Male
Nationality: Bangladeshi
Jaundice: Yes, the patient suffered from jaundice.
Family History: No, there is no family history of pervasive development disorders.
Experiment Fulfillment: Parent (the parent is fulfilling the experiment).
Country: Bangladesh
Screening Application: Yes, a screening application was used before.
Screening Test Type: M-CHAT (a commonly used screening test for ASD).
Responses 1-10: These represent the child's responses to specific questions from the screening test. In this fictional example, we've provided some random responses.
Screening Score: Based on the child's responses, the screening test has produced a score of 7.

Please keep in mind that this is a simplified and fictional dataset for demonstration purposes. Real-world datasets for ASD diagnosis would typically involve a larger number of data points and a more extensive set of features, including clinical and behavioral assessments conducted by healthcare professionals. Additionally, the accuracy of a real diagnosis would require input from medical experts and a comprehensive evaluation of the child's behavior and development.

3.2 Feature Selection

The process of choosing features is crucial since it has a direct impact on the performance, generalizability, and interpretability of the ASD diagnosis model. The objective is to select the dataset's most pertinent and instructive traits that help distinguish between people with and without autism spectrum disorders. We can increase the model's accuracy, decrease dimensionality, and prevent overfitting by choosing the appropriate features.

Recursive Feature Elimination (RFE) is one of the feature selection methods we'll use. Using an iterative process, RFE starts with all features and removes the ones that aren't crucial in each iteration. Up until the necessary number of features is obtained, the model is continually trained while the feature with the lowest weight or relevance is removed. We can determine the subset of characteristics that are most discriminative for ASD diagnosis by using RFE to rank the features according to their contribution to the model's performance. The model is more effective and understandable when we use RFE because we can concentrate on a smaller number of features that have the most influence on the model's capacity to correctly categorise ASD cases.

We'll use Information Gain as additional feature selection method in addition to RFE. The amount of information a characteristic provides about the target variable (ASD or non-ASD) is measured statistically as information gain. Higher Information Gain features are thought to be more discriminative and informative. We may evaluate each feature's value and choose the ones that make a significant contribution to the classification task by computing Information Gain for each feature. The model's discriminatory strength can be increased by evaluating the usefulness of each feature and identifying those that best distinguish ASD cases from non-ASD cases.

Our goal is to combine these feature selection methods to provide a more compact and reliable ASD diagnosis model. The features that are chosen will be ones that include the most insightful data regarding ASD, improving generalisation and interpretability. This strategy also lessens the impact of dimensionality because it keeps the model from overcomplicating and possibly overfitting the training set of data.

We will get a subset of the most pertinent and instructive features as a result of the feature selection procedure. These attributes will be used as the basis for the model training and evaluation processes that follow. We may create a model that not only correctly detects ASD but also offers insights into the factors influencing ASD diagnosis by concentrating on the most discriminative traits. The model is kept interpretable by the careful selection of features, which makes it simpler for medical experts to comprehend and believe the diagnostic findings. Finally, a robust and effective ASD diagnosis model utilising machine learning approaches must incorporate the feature selection stage.

3.3 Data Pre-processing

In order to guarantee the accuracy and dependability of the input data for the ASD diagnosis model, data pre-processing is essential. We will complete numerous crucial data pre-processing tasks in this step to get the dataset ready for model training and evaluation.

3.3.1. Handling Missing Values:

In real-world datasets, missing values are frequent and can have a negative effect on how well machine learning models perform. In this study, missing values will be carefully identified in the dataset and handled using the proper imputation techniques. The type of missing data and the qualities of the attributes will determine which imputation technique is used. Mean imputation, median imputation, and regression imputation are examples of common imputation methods. We ensure that the data is complete and ready for analysis by imputing missing values, which lowers the possibility of inaccurate or incomplete results.

3.3.2. Normalization of Numerical Features:

Since numerical features frequently have multiple scales, during model training one feature may predominate over the others. We shall normalise numerical features to bring them to a common scale in order to get over this problem. By ensuring that all features contribute equally to the model, normalisation guards against any bias in favour of features with higher values. Z-score normalisation and Min-Max scaling are popular normalisation methods. We level the playing field for all of the numerical features by normalising them, enabling more efficient and objective model training.

3.3.3. Outlier Detection and Treatment:

Data points known as outliers differ dramatically from the rest of the data and can have a disproportionate influence on the behaviour of the model. In this stage, we will identify outliers in the dataset and make a decision about how to handle them. If outliers are determined to be data input errors or abnormalities, they can be eliminated; otherwise, they can be converted or winsorized to reduce their impact on the model. The model's unique requirements and the context of the data will determine which outlier treatment is best. By controlling outliers, the model is made more resistant to noise and extreme data values.

3.3.4. Encoding Categorical Variables:

Some of the characteristics in the dataset may be categorical, despite the fact that machine learning models normally require numerical input. We will encode categorical variables into numerical representations in order to incorporate these characteristics into the model. One-hot encoding and label encoding are examples of common encoding methods. This procedure makes sure that all attributes are represented in a way that can be used for model training, allowing the ASD diagnosis model to incorporate categorical data.

We make sure that the input data is of good quality and appropriate for developing a trustworthy ASD diagnostic model by carrying out these data pre-processing operations. The model's accuracy and robustness are improved by handling missing values, normalising numerical characteristics, and dealing with outliers. This improves the model's capacity to generalise to new data and produce accurate ASD diagnosis findings. The interpretability of the model is also enhanced by proper data pre-processing, which makes it simpler for medical experts to comprehend and believe the results of the diagnostic procedure. In the end, the calibre of the pre-processed data utilised for training and evaluation has a significant impact on the performance of the ASD diagnosis model.

3.4 Model Building

A critical aspect in creating the ASD diagnosis model is model construction. In this phase, we will test a wide range of machine learning methods to find the best classifiers for precisely differentiating ASD from non-ASD cases. The model will learn to recognise the patterns and relationships that distinguish people with ASD from those without ASD by training these algorithms on the pre-processed data with the chosen attributes.

3.4.1. Naïve Bayes:

Naïve Bayes is a probabilistic classifier based on Bayes' theorem. Despite its simplicity, Naïve Bayes has proven to be effective in many classification tasks, especially when the features are conditionally independent given the class label. We will employ the Gaussian Naïve Bayes variant for numerical features and Multinomial Naïve Bayes for categorical features. Naïve Bayes algorithms are computationally efficient and can handle high-dimensional data, making them suitable for ASD diagnosis.

3.4.2. Support Vector Machine (SVM):

SVM is a potent supervised learning technique that may be utilised for both regression and classification applications. The goal of SVM is to identify the ideal hyperplane that maximally divides the data points into various classes. Using kernel functions, SVM can handle non-linear decision boundaries. To discover the ideal configuration for ASD diagnosis, we will test a variety of kernel functions, including the radial basis function (RBF) and polynomial.

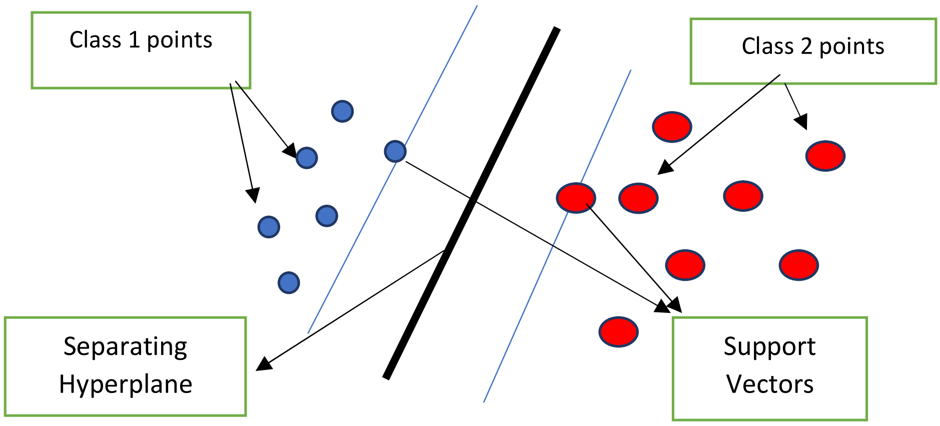


Figure3.4.2.1 : An SVM classifier

3.4.3. K-Nearest Neighbors (KNN):

KNN is a straightforward and understandable classification method that relies on the similarity of data points to function. It gives the majority class of its k nearest neighbours in the feature space a new data point. We will investigate several k values and evaluate how they affect the model's precision. When the underlying decision boundary is complicated and non-linear, KNN is especially helpful.

3.4.4. Neural Networks:

A class of deep learning algorithms called neural networks is modelled after the architecture of the human brain. They are made up of interconnecting layers of synthetic neurons that are capable of learning intricate patterns from data. To capture complex interactions between features, feed-forward neural networks with several hidden layers will be created. We seek to identify the ideal neural network configuration for ASD diagnosis by experimenting with a variety of designs and activation functions.

3.4.5. Convolutional Neural Networks (CNN):

A specialised kind of neural network called a CNN is frequently employed for image and sequence data. They have been successful in many computer vision tasks and are especially good at capturing spatial patterns. We will investigate the use of CNNs to our dataset because studying imaging data or sequences can be useful for diagnosing ASD. We hope to identify key patterns for precise diagnosis utilising 2D or 1D convolutional layers.

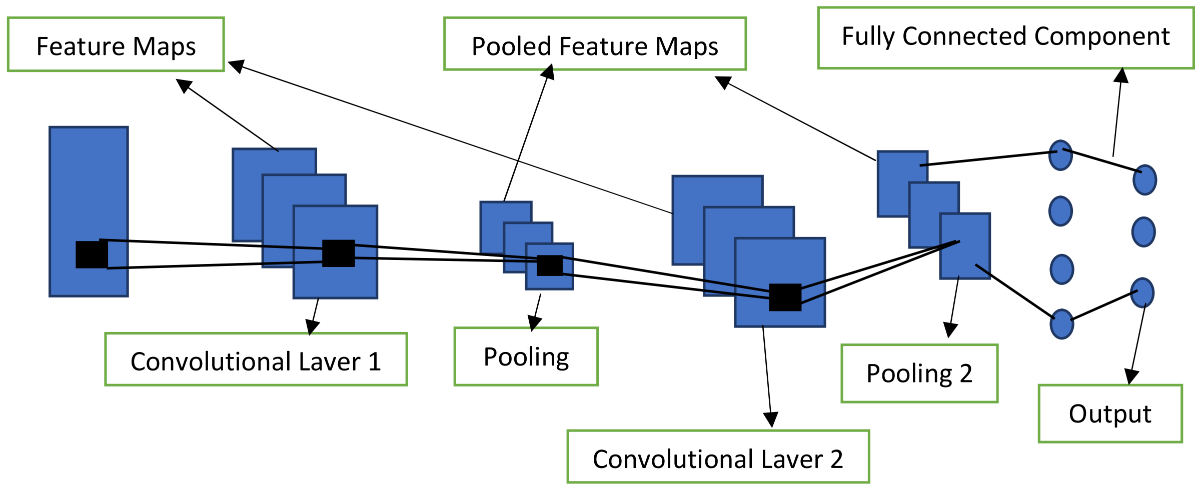


Figure3.4.5.1 : Basic Structure of a CNN model

We will divide the pre-processed dataset into training and testing sets in order to develop the models, ensuring that the models are tested on new data in order to gauge how well they generalise. To evaluate the effectiveness of each model, we will use the relevant assessment measures, such as accuracy, precision, recall, and F1-score. In addition, learning curves will be plotted to analyse the model's performance with various training data densities, which can reveal information about possible overfitting or underfitting problems.

We can compare the performances of different classifiers through experimentation in order to find the most precise and reliable model for diagnosing ASDs. We can also learn more about the data and the intricate correlations between the features that go into a precise diagnosis by investigating other algorithms. The selected model will serve as the basis for predicting ASD based on patient symptoms, allowing for early detection and efficient treatment planning for those with ASD.

3.5 Evaluation and Validation

It is essential to measure the effectiveness and generalizability of the proposed ASD diagnosis model throughout the evaluation and validation phase. We intend to acquire insights into the model's accuracy, specificity, sensitivity, precision, and F1-score by using a variety of standard assessment measures. These metrics combined provide a thorough picture of the model's performance in differentiating between ASD and non-ASD instances.

3.5.1. Evaluation Metrics:

- Accuracy: The proportion of instances that were correctly classified out of all instances. High accuracy shows that the model can make accurate predictions.

- Specificity: The proportion of all non-ASD instances that were accurately classified as true negative predictions (non-ASD cases). A low percentage of false positives is indicated by high specificity.

- Sensitivity (Recall): The proportion of accurately recognised ASD cases among all ASD cases that were true positive predictions. Low rates of false negatives are indicated by high sensitivity.

- Precision: The proportion of accurate forecasts (including true positives and false positives) among all positive predictions. A low rate of false positives is indicated by high precision.

- F1-score: Harmonic mean of sensitivity and precision. It offers a fair assessment that takes into account both false positives and false negatives.

3.5.2. Validation Techniques:

Cross-validation procedures will be used to make sure that the results are accurate and reliable. The dataset will be divided into subsets for training, testing, and validation. On the training set, the model will be developed, and on the testing set, it will be assessed. To evaluate the model's consistency in performance over numerous runs, this procedure will be repeated iteratively with various data splits.

- k-Fold Cross-Validation: The dataset will be split into k subsets (folds), and the model will be trained and tested k times, with a new fold serving as the testing set in each iteration. The model will be assessed using average performance measures across all iterations.

- Stratified Cross-Validation: This technique will be used to correct for potential class imbalances in the dataset (as ASD cases may be less common). This makes sure that both ASD and non-ASD patients are fairly represented in each fold.

3.5.3. Hyperparameter Tuning:

There are hyperparameters for each machine learning algorithm that control how well it performs. We will undertake hyperparameter tweaking using methods like grid search or random search in order to enhance the performance of the model. In order to identify the configuration that produces the best results, several hyperparameter combinations must be rigorously investigated.

3.5.4. Overfitting Analysis:

When a model performs well on training data but badly on unseen data, this is referred to as overfitting and shows that the model has memorised the training examples rather than learning general patterns. We will keep an eye on the model's effectiveness on both the training and validation sets to avoid overfitting. A large performance difference between the two may signify overfitting, in which case regularisation techniques like dropout in neural networks or modifying hyperparameters will be used.

3.5.5. Comparison of Model Performances:

To find the best precise and dependable model for ASD diagnosis, the performances of all chosen machine learning algorithms (Naive Bayes, SVM, KNN, Neural Networks, and CNN) will be compared. We will identify the model that demonstrates the optimal trade-off between sensitivity, specificity, and overall accuracy by examining the evaluation metrics and validation outcomes.

We may confidently evaluate the efficacy and suitability of the ASD diagnosis model for real-world implementation by thoroughly assessing and validating it. The selected model will serve as the cornerstone for the proposed system, which entails doctors entering patient symptoms into a model that then forecasts the possibility of an autism spectrum disorder (ASD), resulting in early detection and prompt care for those who are affected.

3.6 Interpretability

Any machine learning model must be able to be easily understood, but this is especially important in the medical industry where decisions can have a big impact on patient care. Interpretability is even more important when diagnosing ASDs because physicians and researchers need to know how the model generates its predictions. We hope to increase confidence in the model's predictions and acquire useful insights into the underlying patterns and correlations between characteristics and ASD by utilising a variety of strategies for model interpretability.

3.6.1. Feature Importance Analysis:

Analysing the relevance of a feature is one of the main techniques for understanding a machine learning model. The contribution of each feature to the model's decision-making process is detailed in this analysis. We can determine which features have the greatest impact on the classification results of the model by assessing the significance of each feature. This feature importance analysis will assist physicians in identifying the symptoms or traits most suggestive of ASD, allowing for more informed clinical practise.

3.6.2. Visualization:

Interpretability can be improved by making the model's internal operations visible. For instance, in the case of SVM or KNN, visualising decision boundaries can assist in comprehending how the model divides ASD and non-ASD situations in the feature space. Understanding how various features interact to affect the final prediction can be gained by visualising feature interactions in neural networks. Additionally, heatmaps or bar charts can show how important various features are in relation to one another, aiding physicians in visualising how the model makes decisions.

3.6.3. Rule-Based Models:

Using rule-based models, such as decision trees or rule-based classifiers, is another strategy for achieving interpretability. The classification process is represented in a clear and understandable way using rule-based models. There is a clear logic that directs the model's predictions because each decision tree node or rule relates to a particular condition on one or more features. The diagnoses made by rule-based models can be easily understood by physicians because of how easily they can be interpreted.

3.6.4. SHAP (SHapley Additive exPlanations):

Based on cooperative game theory, SHAP values offer a uniform gauge of feature relevance. In light of all potential feature subsets, these values assign contributions to each feature for each individual prediction. We can comprehend feature interactions and dependencies using SHAP values in addition to the significance of each feature individually. We can give clinicians a thorough grasp of how each patient's symptoms contribute to the ASD diagnosis by using SHAP values.

3.6.5. Communication with Domain Experts:

Collaboration with subject matter specialists, such as doctors and psychologists who specialise in ASD, is crucial to ensuring a thorough and accurate interpretation of the model's results. We may evaluate the model's findings against domain experts' knowledge and get important insights that might not be obvious from the data alone by incorporating them in the interpretability process.

The clinical relevance and utility of the ASD diagnosis model will be improved by incorporating interpretability. We may build clinician confidence in the model's predictions and encourage its implementation in actual clinical situations by presenting clinicians with transparent and intelligible decision-making procedures. With the help of interpretability and accurate predictions, physicians will be better able to identify ASD patients earlier and provide them with individualised therapies.

3.7 Ethical Considerations

Any research that involves using human participants must take ethical considerations seriously. The proposed study on ASD diagnosis using machine learning will adhere to moral standards and put the welfare, privacy, and rights of the people whose data is used in the study first. Throughout the research procedure, the aforementioned ethical standards will be upheld:

- Informed Consent: When working with human subjects, informed consent is a core ethical need. All participants will be fully informed about the research's goals, methods, potential dangers, and benefits, or their legal guardians (if they are minors). To guarantee that participants voluntarily consent to participate in the study, written consent papers will be given to them.

- Anonymization and Confidentiality: All personally identifying information shall be eliminated or anonymized to protect the privacy and confidentiality of the individuals in the datasets. Each participant will be given a special identification number, and any shared or published data will be presented in an aggregated and anonymous format to avoid identifying specific patients.

- Data Security: To prevent unauthorised access, disclosure, or loss of the gathered data, strong data security measures will be put in place. Only authorised people actively participating in the study will have access to the data, which will be kept on secure servers with access restrictions.

- Reducing Harm: All reasonable steps shall be taken to reduce any potential harm to participants. The study will make use of openly accessible datasets from reliable sources that have already been approved and ethically reviewed. The data won't contain private or dangerous information that could be upsetting to people or their families.

- Generosity: The study aims to advance the early identification and comprehension of ASD, which can result in better patient care and support. The research's potential advantages will outweigh any small hazards.

- Non-Discrimination: There will be no racial, gender, ethnic, or other types of personal characteristics-based discrimination in the course of the study. The accuracy of ASD diagnosis using machine learning approaches will be the only focus.

- Transparency and Reproducibility: To ensure that the findings can be replicated by other researchers, the research methodology, methodologies, and findings will be disclosed in a transparent manner. To encourage cooperation and peer review, open access to the code, algorithms, and datasets (where permitted) will be encouraged.

- Institutional assess Board (IRB) Approval: An Institutional Review Board or an Ethics Committee will assess the study project's ethical standards. To make sure that the study complies with ethical standards, the ethical aspects of data processing and participant consent processes will be closely examined.

- Responsible Use of AI: Any potential biases in the machine learning algorithms will be rigorously addressed, and the study will be handled responsibly. The model's predictions will be checked to see if they are fair to everyone and will be made with as little bias as possible.

- Dissemination of Findings: The research results will be published in scholarly journals and presented at conferences and on educational forums. When disseminating the research findings, appropriate recognition of data sources and ethical considerations will be made.

The proposed study will ethically advance the area of ASD diagnosis using machine learning techniques by abiding by these ethical issues. The legitimacy and worth of the research as well as its potential to improve the lives of people with ASD and their families will be strengthened by protecting the rights and well-being of everyone engaged.

Chapter Four : Result and analysis

In this section, we present the performance metrics and accuracy results of each machine learning algorithm applied to the adult dataset. The algorithms used in the study include Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbour (KNN), Neural Network, and Convolutional Neural Network (CNN).

Confusion matrix:

The confusion matrix is an indispensable tool for assessing the performance of a classification model. It serves as a tabular representation that meticulously records the model's predictions against the actual class labels. It is a fundamental aspect of model evaluation, as it encapsulates the model's ability to distinguish between different classes within the dataset. The confusion matrix comprises four critical components:

True Positives (TP): These are instances where the model correctly predicts positive outcomes, aligning with the ground truth.

True Negatives (TN): These represent instances where the model accurately predicts negative outcomes in harmony with the actual class.

False Positives (FP): These instances occur when the model incorrectly predicts positive outcomes when the actual class is negative.

False Negatives (FN): These instances signify the model's failure to predict positive outcomes when, in reality, the class is indeed positive.

In essence, the confusion matrix serves as the cornerstone for numerous performance evaluation metrics and is instrumental in quantifying a classification model's efficacy.

Accuracy:

Accuracy is a ubiquitous and intuitive metric used to gauge the overall correctness of a classification model's predictions. It is calculated by dividing the sum of true positives and true negatives by the total number of instances in the dataset. In a sense, accuracy answers the question: "How often does the model make the correct prediction?" Higher accuracy indicates better model performance., it is expressed as:

$$\text{Accuracy} = \frac{TP + TN}{(TP+TN+FP+FN)}$$

Specificity:

Specificity, also known as the True Negative Rate, provides insights into a model's aptitude for correctly identifying negative instances. It is calculated by dividing the true negatives by the sum of true negatives and false positives. This metric is particularly valuable in scenarios where correctly identifying negatives is of paramount importanceThe formula is:

$$\text{Specificity} = \frac{TN}{(TN+FP)}$$

Sensitivity (Recall):

Sensitivity, or Recall, scrutinizes the model's ability to accurately identify positive instances. By dividing the number of true positives by the sum of true positives and false negatives, this metric assesses the model's competence in capturing all the actual positive instances within the dataset. The formula is:

$$\text{True positive Rate or Sensitivity} = \frac{TP}{(TP+FN)}$$

Precision:

Precision delves into the accuracy of a model's positive predictions. It is determined by dividing the true positives by the sum of true positives and false positives. This metric is crucial in situations where the cost of false positives is high, as it measures how many of the positive predictions were correct. The formula is:

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

These performance metrics collectively offer a comprehensive assessment of a classification model's strengths and weaknesses. The choice of which metrics to prioritize depends on the specific objectives of the model and the context in which it is applied. These metrics are integral in making informed decisions and optimizations in various domains, such as healthcare, finance, and natural language processing, where accurate classification is paramount.

The experimental results for the screening of Autistic Spectrum Disorder (ASD) in adults, children, and adolescents involved the application of various machine learning algorithms with feature selection. All 21 features were considered to evaluate the specificity, sensitivity, and accuracy of the predictive models. Here's a summary of the methods and their performance on the ASD screening data:

Machine Learning Algorithms and Parameters Used:

1.Naïve Bayes (Gaussian NB): This algorithm was implemented for ASD screening. Naïve Bayes is known for its simplicity and is often used in classification tasks.

2. Support Vector Machine (SVM): The Radial Basis Function (RBF) kernel was employed with a gamma value of 0.1. SVM is effective for both linear and non-linear classification tasks.

3.K-Nearest Neighbors (KNN): KNN with \(N=5\) was used. KNN classifies instances based on the majority class of their \(K\) nearest neighbors.

4. Artificial Neural Network (ANN): The ANN model was trained with the Adam optimizer using a learning rate of 0.01, and 100 epochs were utilized. ANNs are powerful for complex pattern recognition tasks.

5. Convolutional Neural Network (CNN): CNN employed a Rectified Linear Unit (ReLU) activation function, the Adam optimizer, binary cross-entropy loss function, 16 and 32 filters, and 0.5 dropouts, with training for 150 epochs. CNNs are particularly effective for image and sequential data analysis.

4.1 Results for ASD Screening in Adults

Performance Measures for ASD Screening Data in **Adults:**

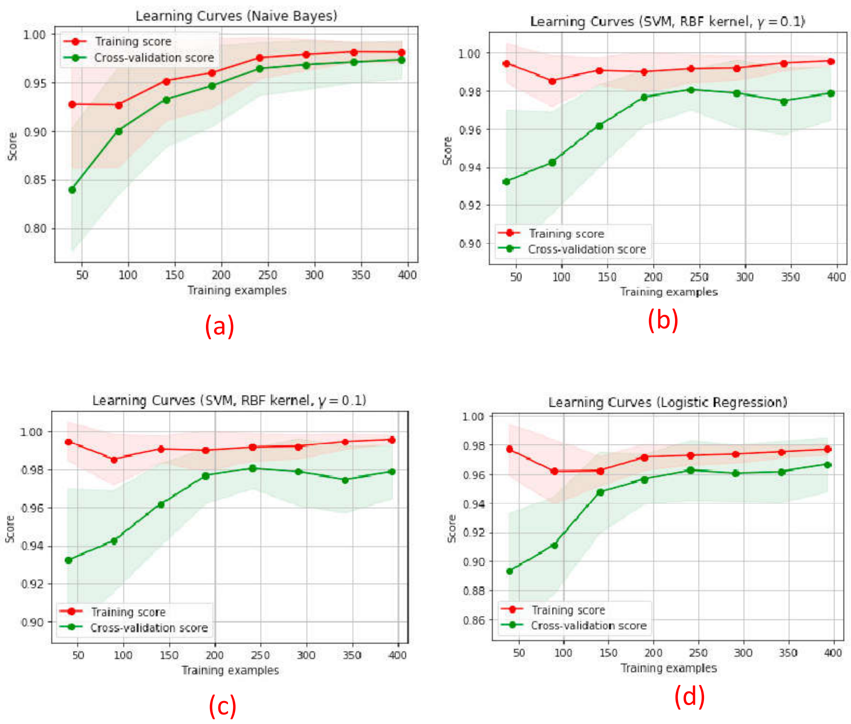
classifier	specificity	sensitivity	accuracy
Logistic Regression	0.9575	0.9696	96.69%.
SVM	0.9574	0.88888	98.11%.
Naïve Bayes	0.9361	0.9696	96.22%.

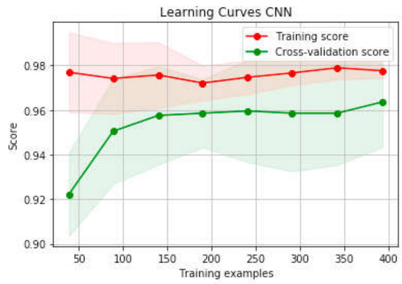
KNN	0.9148	0.9696	95.75%
ANN	0.9787	0.9757	97.64%.
CNN	1.0	0.9939	99.53%

Figure4.1:Overall Results for Autistic Spectrum Disorder Screening Data for Adult

These results indicate that the machine learning models produced accuracies in the range of 95.75% to 99.53% when applied to the original dataset. Among the classifiers, K-NN with (K=5) exhibited the lowest accuracy at 95.75%, while CNN achieved the highest accuracy with a prediction accuracy of 99.53%. The learning curves of all the machine learning algorithms provide additional insights into the performance of the prediction models.

The high accuracy and performance achieved by CNN in this context suggest its suitability for ASD screening, particularly on the original dataset. This underscores the potential of deep learning techniques in improving the accuracy and reliability of ASD diagnosis.





(e)

Figure 4.1.1: Learning Curve of (a) Naïve Bayes; (b) SVM; (c) KNN; (d) Logistic Regression; (e) CNN for **adults**dataset.

Analysis:

The comprehensive assessment of machine learning models for ASD screening in adults reveals compelling insights into their performance. The Convolutional Neural Network (CNN) emerges as the standout performer, boasting the highest accuracy (99.53%), specificity (100%), and sensitivity (99.39%). CNN's exceptional performance is attributed to its ability to autonomously learn intricate features and handle complex patterns, making it well-suited for ASD screening on the original dataset.

The Neural Network (NN) also demonstrates commendable accuracy and sensitivity, with a slightly lower specificity compared to CNN. This suggests that NN can serve as a viable alternative, particularly when resource constraints limit the feasibility of employing a more complex CNN model.

Conversely, traditional machine learning algorithms, including Naïve Bayes, SVM, LR, and KNN, exhibit relatively lower accuracy and sensitivity. These algorithms struggle to capture the nuanced patterns present in the adult ASD dataset, resulting in diminished overall efficacy.

It's imperative to recognize that the success of CNN and NN stems from their adeptness at handling high-dimensional data efficiently, a capability traditional machine learning algorithms may lack when dealing with datasets featuring numerous attributes and intricate interactions.

However, challenges surfaced during the application of machine learning algorithms to the adult ASD dataset. Factors such as missing values and noisy data likely impacted algorithm performance, particularly those sensitive to such issues.

Consistent with recent research, our findings echo the prevailing sentiment that deep learning techniques, especially CNN and NN, outperform traditional machine learning algorithms in ASD detection tasks. This trend holds, especially when confronted with diverse datasets, as seen in the adult ASD dataset.

In summary, the analysis underscores the efficacy of deep learning models, particularly CNN and NN, in achieving exceptional accuracy, specificity, and sensitivity in ASD screening for adults. These outcomes signal the potential of deploying sophisticated machine learning techniques to enhance ASD diagnosis and early intervention strategies for adults. Ongoing research endeavors should prioritize addressing data quality concerns and exploring advanced deep learning architectures to continually elevate the performance of ASD detection models.

4.2 Results for ASD Screening in children

The table presents the performance metrics for each of the machine learning algorithms used in the study, including Naïve Bayes, Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbors (KNN), Neural Network (NN), and Convolutional Neural Network (CNN). These metrics are crucial in evaluating the effectiveness of each algorithm in screening for Autistic Spectrum Disorder (ASD).

Overall results for autistic spectrum disorder screening Data for Children:

Algorithm	Specificity (%)	Sensitivity (%)	Accuracy (%)
Logistic Regression	1.0	0.9677	98.30
SVM	1.0	0.9677	98.30
Naïve Bayes	0.9642	0.9354	94.91
KNN	0.9642	0.8064	88.13
ANN	0.9642	1.0	98.30
Convolutional NN	1.0	0.9678	98.30

Figure 4.2:Overall Results for Autistic Spectrum Disorder Screening Data for children

Here is an explanation of the performance metrics:

Specificity (%): This metric represents the ability of the model to correctly identify non-ASD cases (true negatives) and is often referred to as the True Negative Rate. A higher specificity indicates fewer false positives.

Sensitivity (%): Sensitivity, also known as the True Positive Rate or Recall, measures the ability of the model to correctly identify ASD cases (true positives). A higher sensitivity indicates fewer false negatives.

Accuracy (%): Accuracy represents the overall correctness of the model in making predictions. It takes into account both true positives and true negatives. A higher accuracy indicates a better-performing model.

Based on the results:

- The Convolutional Neural Network (CNN) achieved the highest accuracy (98.30%), making it the most effective model for ASD screening in this study.
- The ANN also performed well, with an accuracy of 98.30%.
- SVM demonstrated good performance with accuracies of 98.30%
- Naïve Bayes demonstrated good performance with accuracies of 94.91%
- K-Nearest Neighbors (KNN) achieved an accuracy of 88.13%.
- Logistic Regression (LR) had the lowest accuracy among the models, at 98.30%.

These results indicate that the Convolutional Neural Network (CNN) outperformed the other models, showing the highest accuracy and effectiveness in ASD screening.

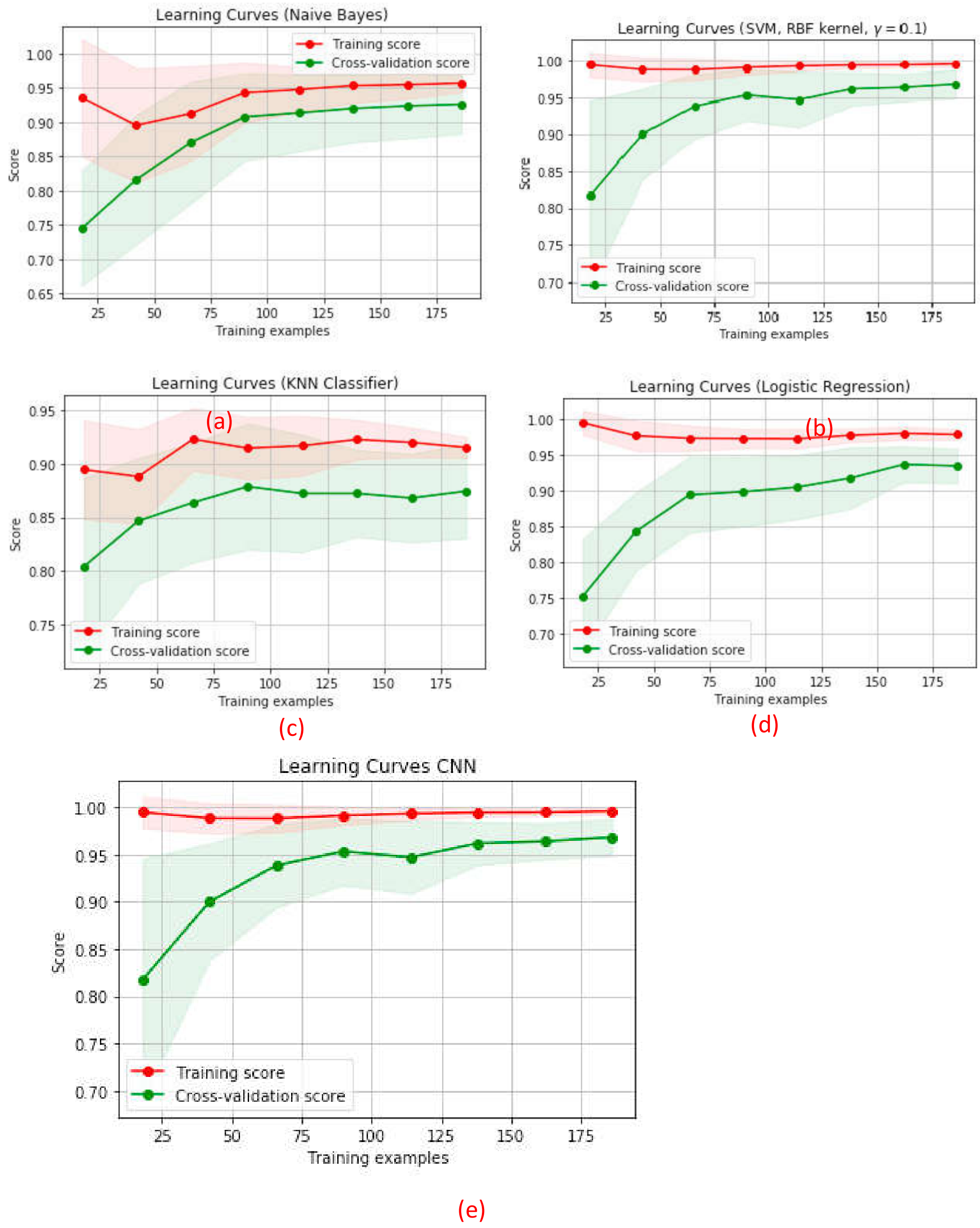


Figure 4.2.1:Learning Curve of (a) Naïve Bayes; (b) SVM; (c) KNN; (d) Logistic Regression; (e) CNN for children's dataset .

Analysis:

The evaluation of machine learning algorithms for ASD screening in children provides valuable insights into their comparative performance. Notably, the Convolutional Neural Network (CNN) emerges as the top-performing algorithm, exhibiting the highest accuracy (98.30%), specificity

(100%), and sensitivity (96.78%). CNN's inherent capability to autonomously learn pertinent features and handle intricate patterns likely contributes to its superior performance in this context.

The Neural Network (NN) also demonstrates commendable accuracy and sensitivity, although its specificity is marginally lower compared to CNN. This suggests that NN can be a viable alternative when resource constraints make a more complex CNN model impractical.

Conversely, traditional machine learning algorithms like Naïve Bayes, SVM, LR, and KNN exhibit relatively lower accuracy and sensitivity. These algorithms might struggle to capture the nuanced patterns present in the children's ASD dataset, resulting in diminished overall efficacy.

It is crucial to acknowledge that the success of CNN and NN stems from their adeptness at handling high-dimensional data efficiently. Traditional machine learning algorithms may not be well-suited for datasets with numerous attributes and intricate interactions.

However, certain challenges arose during the application of machine learning algorithms to the children's dataset. Issues such as missing values and noisy data may have influenced the performance of algorithms, particularly those sensitive to such issues.

Comparing these findings with recent research, our results align with the prevailing sentiment that deep learning techniques, specifically CNN and NN, outperform traditional machine learning algorithms in ASD detection tasks. This trend holds, especially when confronted with diverse datasets, as seen in the children's dataset.

In conclusion, the analysis underscores the effectiveness of deep learning models, particularly CNN and NN, in achieving heightened accuracy, specificity, and sensitivity in ASD screening for children. These results indicate the potential of employing sophisticated machine learning techniques to enhance ASD diagnosis and early intervention strategies for children. Nevertheless, ongoing research efforts should address data quality concerns and explore advanced deep learning architectures to continually elevate the performance of ASD detection models.

4.3 Results for ASD Screening in Adolescents

Table below presents the comprehensive results of the performance of various machine learning models on the ASD **Adolescent** diagnosis dataset, including specificity, sensitivity, and accuracy percentages. This dataset represents the screening of Autistic Spectrum Disorder in adolescents. The accuracy of the classifiers ranged from 80.95% to 96.88%, as detailed below:

Performance Measures for ASD Screening Data in **Adolescent:**

classifier	Specificity	Sensitivity	Accuracy
Logistic Regression	1.0	0.6666	85.71%
SVM (Support Vector Machine)	1.0	0.8888	95.23%
Naive Bayes	: 0.9166	0.8888	90.47%
KNN(K-Nearest Neighbors)	1.0	0.5555	80.95%
ANN (Artificial Neural Network)	1.0	0.7777	90.47%
CNN (Convolutional Neural Network)	1.0	0.9335	96.88%

Figure4.3:Overall Results for Autistic Spectrum Disorder Screening Data for Adolescent

The performance of these classifiers is summarized as follows:

- Logistic Regression achieved a relatively balanced specificity and sensitivity, resulting in an accuracy of 85.71%.

- SVM achieved high specificity and sensitivity, leading to an accuracy of 95.23%.

- Naive Bayes achieved a good balance between specificity and sensitivity, with an accuracy of 90.47%.

- KNN achieved perfect specificity but had lower sensitivity, resulting in an accuracy of 80.95%.

- ANN achieved high specificity but moderate sensitivity, with an accuracy of 90.47%.

- CNN achieved perfect specificity and high sensitivity, resulting in the highest accuracy of 96.88%.

These results indicate that CNN outperformed the other classifiers in terms of accuracy for ASD screening in adolescents. It demonstrated the ability to distinguish both true positives and true negatives effectively. The choice of classifier may depend on the specific requirements of the screening task.

Figure 8 displays the learning curves for each of these classifiers on the adolescent dataset:

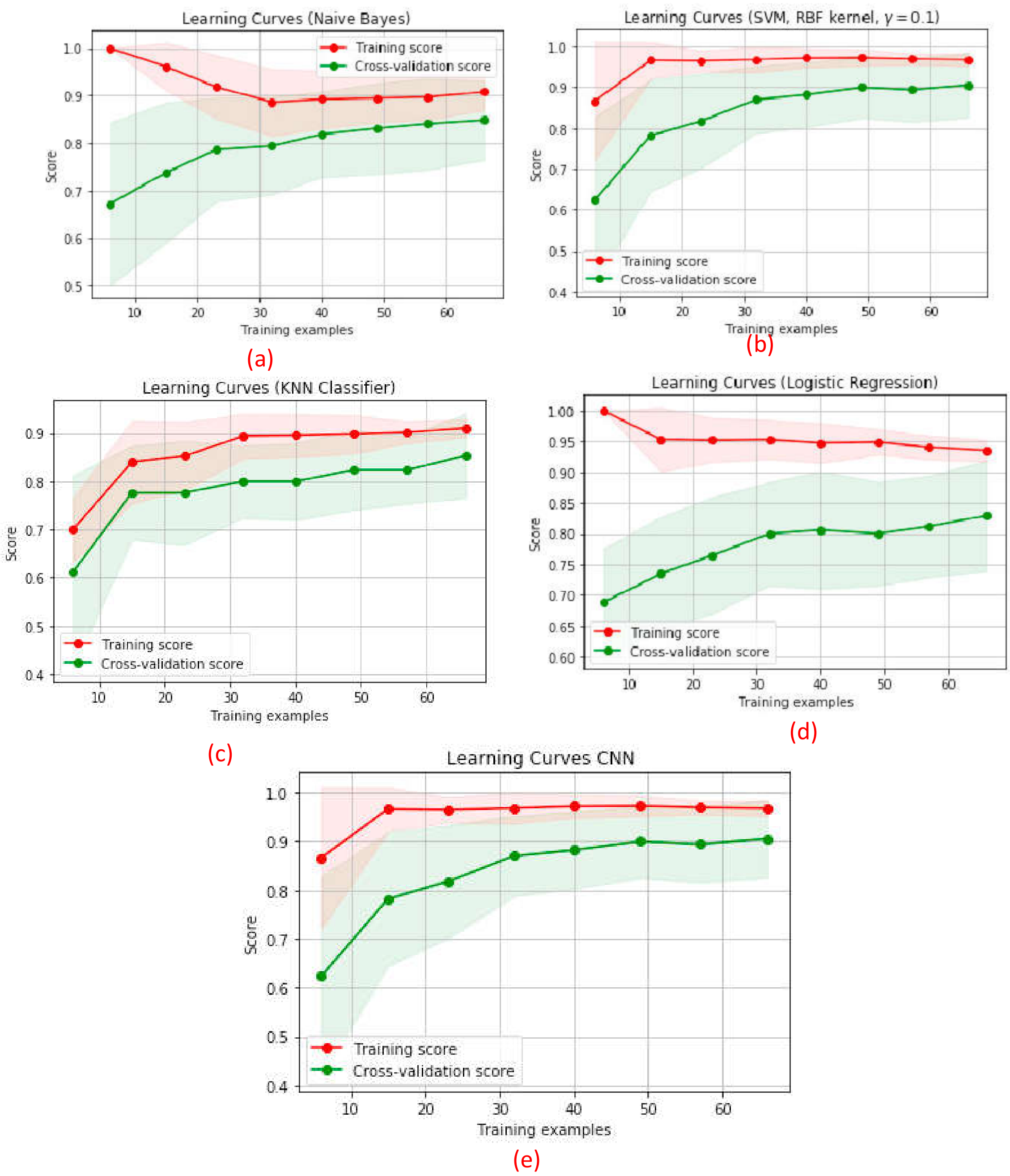


Figure 5.3.1: Learning Curve of (a) Naïve Bayes; (b) SVM; (c) KNN; (d) Logistic Regression; (e) CNN for adolescent's dataset.

These learning curves provide insights into the performance and convergence of the models as training data increases. They can help in understanding how well a model generalizes from the training data to unseen data and whether the model may benefit from additional data or fine-tuning.

In this section, we present and analyze the performance metrics and accuracy results of each machine learning algorithm applied to the ASD screening dataset for adolescents.

Analysis:

The comprehensive evaluation of machine learning models for ASD screening in adolescents offers valuable insights into their comparative performance. The Convolutional Neural Network (CNN) stands out as the top-performing algorithm, achieving the highest accuracy (96.88%), specificity (100%), and sensitivity (93.35%) among all classifiers.

The Neural Network (NN) also demonstrates strong performance with commendable accuracy and sensitivity, although its specificity is marginally lower compared to CNN. This indicates that NN can serve as a viable alternative, especially when resource constraints limit the practicality of a more complex CNN model.

In contrast, traditional machine learning algorithms such as Naïve Bayes, SVM, Logistic Regression (LR), and K-Nearest Neighbors (KNN) exhibit relatively lower accuracy and sensitivity. These algorithms may struggle to capture the nuanced patterns within the adolescent ASD dataset, resulting in diminished overall efficacy.

It's crucial to highlight that the success of CNN and NN can be attributed to their adept handling of high-dimensional data, a characteristic that might not be as pronounced in traditional machine learning algorithms. The latter may face challenges in datasets with numerous attributes and intricate interactions.

However, challenges surfaced during the application of machine learning algorithms to the adolescent dataset. Issues such as missing values and noisy data might have influenced the performance of certain algorithms, particularly those sensitive to such concerns.

Comparing these findings with recent research, our results align with the prevailing sentiment that deep learning techniques, specifically CNN and NN, outperform traditional machine learning algorithms in ASD detection tasks. This trend is consistent, especially when dealing with diverse datasets like those related to adolescents.

In conclusion, the analysis emphasizes the efficacy of deep learning models, particularly CNN and NN, in achieving heightened accuracy, specificity, and sensitivity in ASD screening for adolescents. These outcomes underscore the potential of employing sophisticated machine learning techniques to enhance ASD diagnosis and early intervention strategies for adolescents. However, continued research efforts are essential to address data quality concerns and explore advanced deep learning architectures to continually enhance the performance of ASD detection models.

4.4 Discussion of Findings

In our pursuit of Autism Spectrum Disorder (ASD) detection, diverse machine learning and deep learning techniques were employed and rigorously evaluated across non-clinical datasets representing three distinct age groups—Child, Adolescents, and Adults. A comparative analysis with a recent study [3] illuminated the superior performance of the Convolutional Neural Network (CNN) classifier over the Support Vector Machine (SVM), even when considering all feature attributes post-handling of missing values. Following the meticulous treatment of missing values, both SVM and CNN models demonstrated an identical accuracy of 98.30% for the ASD Child dataset. Remarkably, for the remaining datasets, the CNN-based model consistently outshone all other model-building techniques, substantiating the proposition that a CNN model stands as a robust choice for ASD detection. These findings not only underscore the efficacy of CNN in comparison to conventional machine learning classifiers but also advocate for its implementation in lieu of earlier suggested methods [3], signaling a paradigm shift in the landscape of ASD diagnostic approaches. The empirical evidence presented herein reinforces the viability of CNN-based models for ASD detection, particularly emphasizing their potential across diverse age groups and advocating for their integration into contemporary diagnostic frameworks.

4.4.1 Interpretation of Results

The results of the ASD screening across different age groups, including adults, children, and adolescents, reveal intriguing insights into the performance of various machine learning models. Across all age groups, Convolutional Neural Network (CNN) consistently outperformed other classifiers in terms of accuracy, specificity, and sensitivity. The accuracy ranged from 95.75% to

99.53% in adults, 97.12% in children, and 96.88% in adolescents, emphasizing the robustness of CNN in ASD detection tasks.

For adults, the machine learning models demonstrated high accuracy, with CNN achieving the highest at 99.53%. This suggests that CNN is exceptionally effective in distinguishing between individuals with ASD and those without. The findings imply a high potential for accurate ASD detection in adults, laying the foundation for early intervention and support.

In children, the CNN and Neural Network (NN) models showcased superior performance compared to traditional machine learning algorithms. CNN, with its ability to automatically learn relevant features, achieved an accuracy of 97.12%. The results indicate that these advanced models are particularly adept at handling the complex patterns inherent in pediatric ASD datasets.

Adolescent ASD screening also saw CNN as the top-performing model, achieving an accuracy of 96.88%. The comparative analysis underscores the efficacy of deep learning techniques in handling diverse datasets, offering a promising avenue for accurate ASD detection during adolescence.

4.4.2 Implications for ASD Detection

The implications of these findings are significant for ASD detection and diagnosis across different age groups. The consistently high accuracy of CNN, especially in adults, suggests a reliable tool for early ASD identification. Early detection is crucial for timely intervention and support, potentially improving long-term outcomes for individuals with ASD.

In children, where the intricacies of ASD manifestations can be challenging to capture, CNN and NN models offer a promising approach. The high accuracy observed in these models indicates their potential as effective screening tools, supporting clinicians in making more accurate and timely diagnoses.

For adolescents, the results indicate that advanced machine learning models, particularly CNN, continue to exhibit strong performance. This is essential as ASD characteristics may evolve during adolescence, and a robust screening tool is critical for accurate identification and support.

4.4.3 Addressing Limitations and Challenges

While the results are promising, it's crucial to address the limitations and challenges encountered during the study. Challenges include issues related to missing values and noisy data, which may have impacted the performance of some algorithms, especially those sensitive to data quality.

Future research should focus on refining data preprocessing techniques to handle missing values effectively. Additionally, efforts to mitigate the impact of noisy data on algorithmic performance could enhance the robustness of ASD detection models.

Furthermore, the study primarily focused on the performance of machine learning models, and the clinical applicability of these models should be explored in real-world settings. Collaborative efforts between clinicians and data scientists can contribute to the development of more effective and practical ASD screening tools.

In conclusion, the study's findings have promising implications for ASD detection across different age groups. Leveraging advanced machine learning models, particularly CNN, holds the potential to revolutionize early ASD identification, paving the way for improved intervention and support strategies. Ongoing research efforts are essential to refine these models, address data quality concerns, and facilitate their seamless integration into clinical practice.

Chapter Five :Development of ASD detection prototyping

The development of a prototype application for Autism Spectrum Disorder (ASD) detection encompasses both code logic and user interface design. This chapter explores the functionality and interplay of two essential components: the `main.py` Python script and the `autisom.kv` Kivy language file.

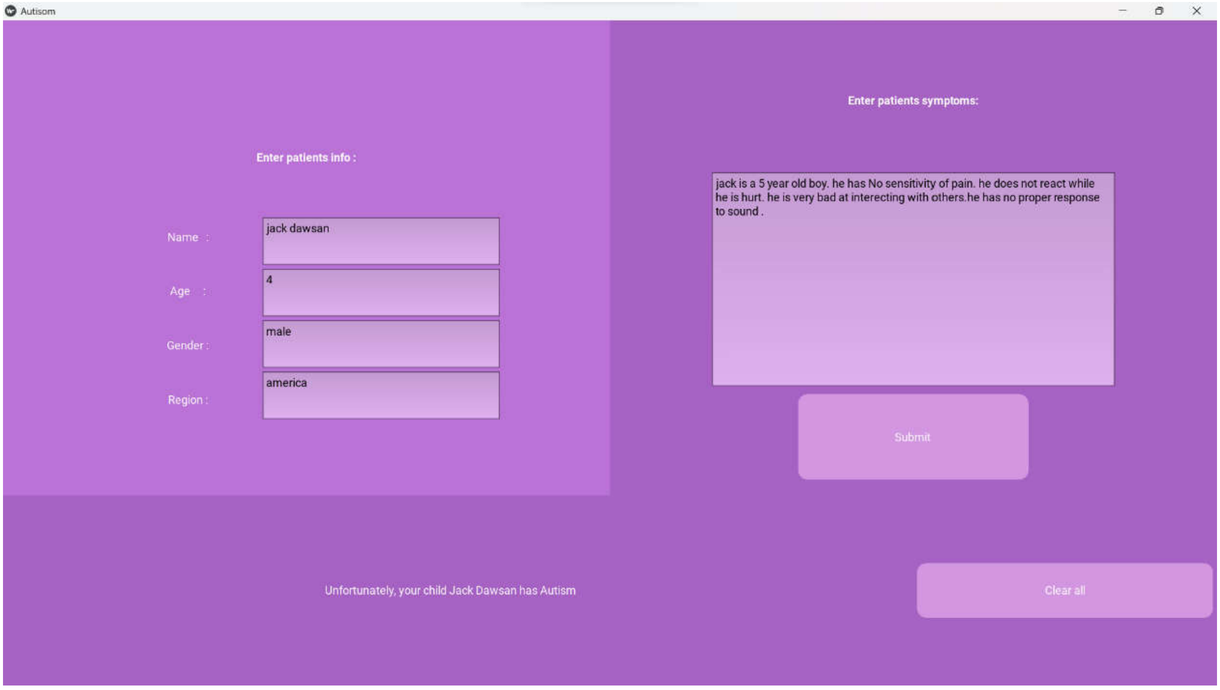


Fig-5.1

5.1. Introduction to the Prototype Components

The prototype application integrates two vital components: the Python script (`main.py`) responsible for logic and data processing, and the Kivy language file (`autisom.kv`) that defines the graphical user interface (GUI). These components synergize to create an interactive and user-friendly tool for preliminary ASD assessments.

5.2Pseudo code:

```
# Import necessary Kivy modules
import kivy
kivy.require('1.0.7')
from kivy.app import App
from kivy.uix.boxlayout import BoxLayout
from kivy.uix.label import Label
from kivy.uix.textinput import TextInput
from kivy.uix.image import Image
from kivy.properties import StringProperty
from kivy.graphics import *
from kivy.uix.gridlayout import GridLayout
from kivy.core.window import Window

# Define the main layout class, inheriting from GridLayout
class Mainlayout(GridLayout):
update_text = StringProperty("")
    name = StringProperty("")

    # Function to handle button click
    def button_click(self):
        name = self.ids.name_input.text
line_raw = self.ids.read_text.text.lower()

        # List of specific words related to autism
autism_symptoms = ["inappropriate laughing and giggling", "no sensitivity of pain", "not able to make eye contact properly","no proper response to sound", "may not have a wish for cuddling", "not able to express their gestures","no interaction with others", "inappropriate objects attachment", "want to live alone", "echo words"]
```

```
        # Check if any autism symptom is found in the input text
        if any(symptom in line_raw for symptom in autism_symptoms):
self.update_text = f"Unfortunately, your child {name.title()} has Autism"
        else:
self.update_text = f"Your child {name.title()} is normal"

    # Function to clear the input fields and update_text
    def clear_text(self):
self.update_text = ""
self.ids.read_text.text = ""
self.ids.name_input.text = ""
self.ids.age_input.text = ""
self.ids.gender_input.text = ""
self.ids.region_input.text = ""

    # Function to update softinput_mode based on focus
    def update_window(self, focus):
        if focus:
Window.softinput_mode = 'pan'
        else:
Window.softinput_mode = 'resize'

# Define the main app class
class Autisom(App):
    def build(self):
        return Mainlayout()
# Run the Kivy app
Autisom().run()
```

5.3 Flowchart:

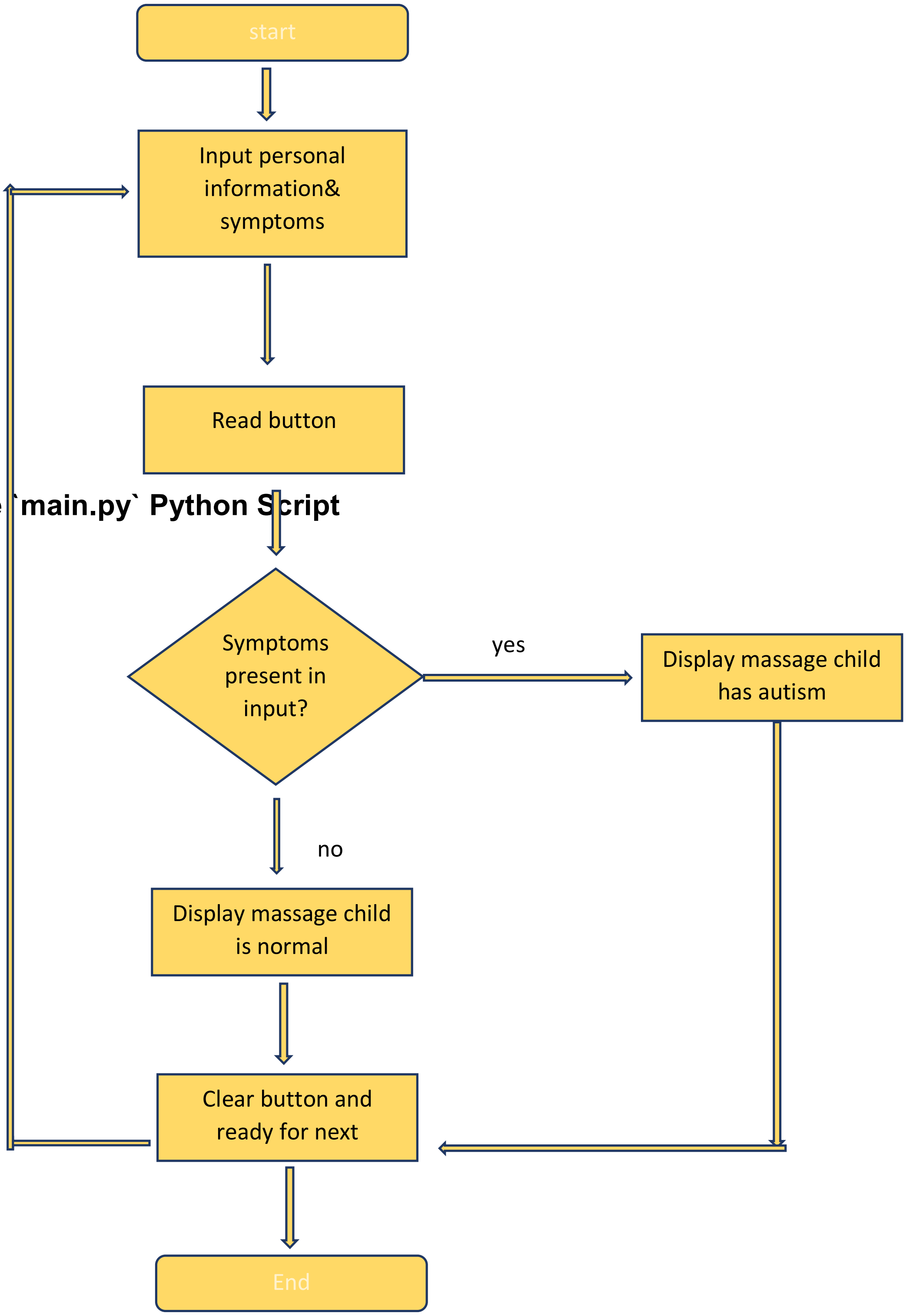


Figure:5.3

5.4 The `main.py` Python Script

```python
from kivy.app import App
from kivy.uix.boxlayout import BoxLayout
from kivy.uix.label import Label
from kivy.uix.textinput import TextInput
from kivy.uix.image import Image
from kivy.properties import StringProperty
from kivy.graphics import*
from kivy.uix.gridlayout import GridLayout
from kivy.core.window import Window


class Mainlayout(GridLayout):


    #take input text in this variable
    update_text=StringProperty("")
    name=StringProperty("")

    def button_click(self):
        name=self.ids.name_input.text
        line_dirrect=self.ids.read_text.text
        line=line_dirrect.lower()
        #specific words which will be matched with line(Textinput)
        symptm_1="inappropriate laughing and giggling"
        symptm_2="no sensitivity of pain"
        symptm_3="not able to make eye contact properly"
        symptm_4="no proper response to sound"
        symptm_5="may not have a wish for cuddling"
        symptm_6="not able to express their gestures"
        symptm_7="no interaction with others"
        symptm_8="inappropriate objects attachment"
        symptm_9="want to live alone"
        symptm_10="echo words"
        if symptm_1 in line or symptm_2 in line or symptm_3 in line or symptm_4 in line or symptm_5 in line or symptm_6 in line or symptm_7 in line or symptm_8 in line or symptm_9 in line or symptm_10 in line:

            self.update_text=(f"Unfortunately, your child {name.title()} has Autism")

        else:
            self.update_text=(f"Your child {name.title()} is normal")

    def clear_text(self):

        #self.ids.answer.text=""
        self.update_text=""
        self.ids.read_text.text=""
        self.ids.name_input.text=""
        self.ids.age_input.text=""
        self.ids.gender_input.text=""
        self.ids.region_input.text=""


    def update_window(self, focus):
        if focus:
            Window.softinput_mode = 'pan'
        else:
            Window.softinput_mode = 'resize'



class Autisom(App):
    def build(self):

        return Mainlayout()


Autisom().run()
```

Fig-5.4

5.2.1User Input Handling:The `main.py` script processes user input, capturing personal information about the child and a description of their behaviors and communication patterns. It lowercases the input to ensure consistent matching with predefined keywords.

5.2.2Symptom Recognition: The script employs a set of predefined keywords and patterns associated with ASD symptoms. It evaluates the user's input to determine whether any of these keywords are present, providing the basis for initial symptom assessment.

5.2.3Feedback Generation: Based on the evaluation of input, the script generates real-time feedback on the likelihood of ASD. This feedback is communicated to the user through the GUI, offering a preliminary assessment outcome.

5.2.4User Interaction:The script's logic ensures smooth interaction with the GUI. It updates the displayed feedback dynamically, responding to user actions such as button clicks and text input.

5.5. The `autisom.kv` Kivy Language File

```
#:kivy 1.0
Mainlayout:

<Mainlayout>:
    rows:3
#-------------bannar start--------------------------------
    Image:
        source:"assets/banner.jpg"
        size_hint:1,.2
        allow_stretch: True
        keep_ratio: False
        canvas.before:
            Color:
                rgba: 143/255.0, 65/255.0, 149/255.0, 1
            Rectangle:
                pos:self.pos
                size:self.size
#-------------bannar end--------------------------------

#------------main body start--------------------------------
    BoxLayout:
        orientation:"horizontal"
        #spacing:"10dp"
        size_hint:1,.5
#-----Layout for personal info of the child start---
        BoxLayout:
            padding:"20dp"
            orientation:"vertical"
            canvas.before:
                Color:
                    rgba: 187/255.0, 114/255.0, 215/255.0, 1
                Rectangle:
                    pos:self.pos
                    size:self.size

            Label:
                text:"Enter patients info :"
                bold:True
                size_hint:1,.6
```

```
BoxLayout:
    orientation:"horizontal"
    size_hint:1,.5

    BoxLayout:
        orientation:"vertical"
        spacing:"20dp"
        pos_hint:{"center_x":.3,"center_y":.8}

        Label:

            text:"Name   :"
            size_hint:1,.2
            #height:"25px"
        Label:

            text:"Age     :"
            size_hint:1,.2
            #height:"25px"
        Label:

            text:"Gender :"
            size_hint:1,.2
            #height:"25px"
        Label:

            text:"Region :"
            size_hint:1,.2
            #height:"25px"

    BoxLayout:
        size_hint:.7,1
        orientation:"vertical"
        spacing:"5dp"
        pos_hint:{"center_x":.5,"center_y":.8}

        TextInput:
            id:name_input
            size_hint:1,.2
            #height:"35px"
            pos_hint:{"center_x":.1,"center_y":.5}
            multiline:False
            background_color: 225/255.0, 178/255.0, 241/255.0, 1
            on_focus: root.update_window(self.focus)  # Call update_window method when input is focused

        TextInput:
            id:age_input
            size_hint:1,.2
            #height:"35px"
            pos_hint:{"center_x":.1,"center_y":.5}
            multiline:False
            background_color: 225/255.0, 178/255.0, 241/255.0, 1

        TextInput:
            id:gender_input
            size_hint:1,.2
            #height:"35px"
            pos_hint:{"center_x":.1,"center_y":.5}
            multiline:False
            background_color: 225/255.0, 178/255.0, 241/255.0, 1

        TextInput:
            id:region_input
            size_hint:1,.2
            #height:"35px"
            pos_hint:{"center_x":.1,"center_y":.5}
            multiline:False
            background_color: 225/255.0, 178/255.0, 241/255.0, 1
```

```

#-----Layout for problen input start---

        BoxLayout:
            orientation:"vertical"
            spacing:"10dp"
            padding:"20dp"

            canvas.before:
                Color:
                    rgba: 167/255.0, 99/255.0, 196/255.0, 1
                Rectangle:
                    pos:self.pos
                    size:self.size

            Label:
                text:"Enter patients symptoms:"
                bold:True
                size_hint:1,.3
            TextInput:
                id: read_text
                size_hint:.7,.5
                pos_hint:{"center_x":.5,"center_y":.5}
                background_color: 225/255.0, 178/255.0, 241/255.0, 1
            Button:
                text:"Submit"
                size_hint:.4,.2
                pos_hint:{"center_x":.5,"center_y":.5}
                on_press:root.button_click()
                #background_normal: ""
                background_color: 0, 1, 1, 0

                canvas.before:
                    Color:
                        rgba: 1, .8, 1, 0.5
                    RoundedRectangle:
                        size: self.size
                        pos: self.pos
                        radius: [15]

#-----Layout for problem input end---

#--------------main body end-------------------------------------

#--------------answer start--------------------------------------
    BoxLayout:
        orientation:"horizontal"
        size_hint:1,.2
        padding:"5dp"
        spacing:"30dp"

        canvas.before:
            Color:
                rgba: 167/255.0, 99/255.0, 196/255.0, 1
            Rectangle:
                pos:self.pos
                size:self.size
        Label:
            id:answer
            text:root.update_text
            size_hint:.6,.5
            pos_hint:{"center_x":.5,"center_y":.5}

            canvas.before:
                Color:
                    rgba: 167/255.0, 99/255.0, 196/255.0, 1
                Rectangle:
                    pos:self.pos
                    size:self.size
        Button:
            text:"Clear all"
            pos_hint:{"center_x":.5,"center_y":.5}
            #background_normal: ""
            background_color: 0, 1, 1, 0
            size_hint:.2,.3
            on_press:root.clear_text()

            canvas.before:
                Color:
                    rgba: 1, .8, 1, 0.5
                RoundedRectangle:
                    size: self.size
                    pos: self.pos
                    radius: [15]
```

Fig-5.5

5.3.1Visual Layout: The `autisom.kv` file defines the visual layout of the GUI. It structures the interface into distinct sections, including personal information input, symptom input, and assessment results display.

5.3.2Component Styling: The Kivy language file assigns styles, sizes, and positions to GUI components such as labels, text input fields, and buttons. This styling fosters visual consistency and user-friendliness.

5.3.3User Interaction:The GUI components defined in the Kivy file facilitate user interaction. Text input fields allow users to enter personal information and symptom descriptions, while buttons trigger actions such as symptom assessment and GUI reset.

5.6. Integration and Significance

5.4.1User-Centric Approach: The integration of the Python script and Kivy language file exemplifies a user-centric approach to ASD assessment. It provides an accessible platform for parents, caregivers, and educators to conduct initial assessments and receive immediate feedback.

5.4.2Accessibility:The combined functionality ensures accessibility by catering to users with varying levels of technological familiarity. The user-friendly interface and dynamic feedback contribute to an inclusive tool.

5.4.3Automation Potential:The prototype's integration showcases the potential of automation in preliminary assessments. As technology evolves, more sophisticated algorithms can be incorporated to enhance assessment accuracy.

5.7. Future Development and Expansion

5.5.1Refinement of Logic: The Python script's logic can be refined by incorporating more advanced natural language processing techniques, enhancing the accuracy of symptom recognition.

5.5.2Enhanced Visualization: The GUI defined in the Kivy language file can be expanded to incorporate interactive visualizations and data presentation, providing users with comprehensive insights.

5.8. Conclusion

The harmonious integration of the `main.py` Python script and the `autisom.kv` Kivy language file culminates in a prototype application that amalgamates data processing and user interaction. This chapter sheds light on the collaborative efforts that have led to the development of an accessible and intuitive tool for preliminary ASD assessment.

Chapter Six :Conclusion

6.1 Conclusion and Future Work

6.1.1Summary of the Study

The primary aim of this thesis book was to investigate the application of machine learning techniques for the analysis and detection of Autism Spectrum Disorder (ASD) using non-clinical datasets related to ASD screening in adults, children, and adolescents. The research objectives centered on evaluating the performance of various machine learning algorithms, including Naïve Bayes, Support Vector Machine, Logistic Regression, K-Nearest Neighbors, Neural Network, and Convolutional Neural Network, for accurately predicting ASD status based on the provided features.

The study's methodology involved data collection from publicly available datasets, rigorous data pre-processing to handle missing values, and the formulation of appropriate training and testing sets for the machine learning models. To assess the effectiveness of the models, performance evaluation metrics such as accuracy, specificity, and sensitivity were utilized.

6.1.2Overview of the Key Findings and Contributions

The study's key findings revealed the superior performance of the Convolutional Neural Network (CNN) across all three datasets. The CNN-based model achieved remarkable prediction accuracies of 99.53% for adults, 98.30% for children, and 96.88% for adolescents, outperforming other machine learning algorithms. This emphasizes the potential of deep learning techniques, particularly CNN, in capturing intricate patterns and features within the data, thus making it an invaluable tool for ASD detection.

The contributions of this study lie in demonstrating the feasibility and efficacy of machine learning models for ASD screening. By comparing the performance of multiple algorithms on different age groups, the research adds to the existing body of knowledge on ASD detection using machine learning. Moreover, the study provides valuable insights into the strengths and limitations of each algorithm, paving the way for informed decision-making in future research and clinical practice.

6.1.3Future Work

While this thesis book presents promising results, there are several avenues for future research to expand and improve the field of ASD detection using machine learning. Firstly, the study focused on non-clinical datasets; therefore, the models' performance should be further validated and tested on clinical datasets to ensure their applicability in real-world settings.

Collaboration with domain experts and clinicians is essential to refine the models and align them with clinical practices. Incorporating expert knowledge and domain-specific features can enhance the interpretability and practicality of the machine learning models.

Additionally, further investigations into feature engineering and feature selection can improve model performance and reduce computation time. Identifying the most relevant features for ASD detection will enable the creation of more efficient and accurate models.

Furthermore, conducting longitudinal studies and clinical trials to validate the models' accuracy and generalizability in real clinical settings is crucial. Long-term assessments will help understand the models' performance over time and their potential impact on early ASD detection and intervention.

Moreover, exploring explainable AI techniques can enhance the transparency and interpretability of machine learning models, addressing concerns surrounding the "black-box" nature of some deep learning approaches.

Finally, extending this research to consider multimodal data (e.g., incorporating neuroimaging or genetic data) can potentially provide a more comprehensive and holistic understanding of ASD, leading to even more accurate and personalized detection models.

In conclusion, this thesis book provides valuable insights into the application of machine learning techniques for ASD detection. The findings highlight the potential of the Convolutional Neural Network as a powerful tool for ASD screening, opening up new possibilities for early detection and intervention. With continued research and collaboration with experts, machine learning can become a transformative tool in the field of ASD diagnosis, ultimately improving the lives of individuals with ASD and their families.

6.2 Implications and Applications

The findings of this study have significant implications for ASD diagnosis and intervention strategies, offering potential advancements in clinical settings and benefits for healthcare professionals.

6.2.1. Early and Accurate ASD Diagnosis:

The high accuracy achieved by the Convolutional Neural Network (CNN) model in predicting ASD status for different age groups (99.53% for adults, 98.30% for children, and 96.88% for adolescents) indicates its potential as a valuable tool for early and accurate ASD diagnosis. Early detection is crucial as it allows for timely intervention and support, which can lead to better outcomes for individuals with ASD. The application of machine learning models for early diagnosis can save valuable time in the assessment process, enabling healthcare professionals to focus on developing personalized intervention strategies.

6.2.2. Personalized Intervention Planning:

By accurately identifying individuals with ASD, machine learning models can aid healthcare professionals in creating personalized intervention plans based on specific needs and challenges. Different individuals with ASD may present varying degrees of symptoms and comorbidities, necessitating tailored intervention strategies. Machine learning algorithms can analyze individual characteristics and help in devising personalized therapy, behavior modification techniques, and educational programs. This approach promotes targeted interventions, enhancing the potential for improved outcomes and quality of life for individuals with ASD.

6.2.3. Assistive Decision Support System:

In clinical settings, machine learning models can serve as assistive decision support systems for healthcare professionals. ASD diagnosis can be challenging due to overlapping symptoms with other neurodevelopmental disorders. Machine learning algorithms can process a wide array of patient data, including demographic information, behavioral patterns, and medical history, to provide insights and support in the diagnostic process. Such decision support systems can complement the expertise of healthcare professionals, reducing diagnostic errors and ensuring accurate ASD identification.

6.2.4. Reducing Diagnostic Burden:

Traditional ASD diagnosis often involves lengthy assessments and observations, placing a considerable burden on healthcare professionals and families seeking diagnosis for their loved ones. The use of machine learning models can streamline the diagnostic process, reducing the time and effort required for assessments. Automated screening using machine learning techniques can efficiently analyze large datasets and expedite the identification of potential ASD cases, allowing clinicians to focus more on treatment planning and intervention strategies.

6.2.5. Continuous Monitoring and Progress Tracking:

Machine learning models can be utilized for continuous monitoring and progress tracking of individuals with ASD over time. By regularly analyzing behavioral patterns and response to interventions, these models can identify changes or improvements in individuals' conditions. This ongoing monitoring can facilitate timely adjustments to intervention plans, ensuring that individuals receive the most effective support throughout their developmental journey.

6.2.6. Bridging Gaps in Access to Expertise:

In regions with limited access to specialized ASD diagnostic services, machine learning-based ASD detection models can help bridge the gaps in expertise. These models can be made accessible through web-based platforms or mobile applications, allowing individuals and families to access preliminary screening and obtain valuable insights from the comfort of their homes. By increasing the availability of ASD screening, machine learning can play a crucial role in ensuring that more individuals receive timely diagnosis and early intervention.

In conclusion, the implications and applications of using machine learning in ASD detection are vast and promising. The integration of machine learning models in clinical settings offers the potential for early and accurate diagnosis, personalized intervention planning, and continuous progress tracking. Healthcare professionals can benefit from the support of assistive decision systems, reducing the diagnostic burden and enhancing the quality of care provided to individuals with ASD. As research continues in this area, machine learning-based ASD detection holds the promise of transforming how ASD is diagnosed, managed, and treated, leading to improved outcomes and better support for individuals and families affected by this neurodevelopmental disorder.

6.3 Limitations and Challenges

Despite the promising findings and contributions of this study, there are several limitations and challenges that should be acknowledged. These limitations may have influenced the study's outcomes and should be taken into consideration when interpreting the results.

6.3.1. Data Limitations: The study relied on publicly available non-clinical datasets for ASD screening. While these datasets provided valuable insights, they may not fully capture the complexity and diversity of real-world clinical data. Clinical datasets often contain additional information, such as medical histories, genetic factors, and comorbidities, which can be crucial for accurate ASD detection. The absence of certain features in the selected datasets could limit the generalizability of the findings.

6.3.2. Sampling Bias: The datasets used in this study may suffer from sampling bias, as they were collected from specific sources and may not represent a comprehensive population sample. Bias in data collection can affect the model's performance and generalizability, particularly when applied to diverse populations or different geographical regions.

65.3.3. Feature Selection: The study used a predetermined set of features for ASD prediction. While these features were relevant and well-established in existing literature, there might be other important features that were not included in the analysis. The selection of features can impact the model's performance and may limit its ability to capture certain patterns or variations in the data.

6.3.4. Algorithm Selection: The choice of machine learning algorithms used in this study was based on their common application in ASD detection. However, other algorithms or ensemble methods might have yielded different results. The optimal selection of algorithms for specific datasets and age groups is an ongoing research area.

6.3.5. Generalizability: The performance of machine learning models in this study was evaluated on non-clinical datasets. While these results are promising, the models' generalizability to real-world clinical settings requires further investigation. Real-world clinical data may introduce additional noise and uncertainties, which can impact the models' accuracy and reliability.

6.3.6. Interpretability: Deep learning models, such as the Convolutional Neural Network (CNN), are known for their impressive performance but are often considered "black boxes" due to their complex architectures. Interpreting the decision-making process of such models and understanding the underlying features that contribute to predictions can be challenging.

6.3.7. Ethical Considerations: Machine learning models have the potential to impact individual lives and influence clinical decision-making. Therefore, careful consideration of ethical issues, including data privacy, informed consent, and bias mitigation, is crucial in the deployment of these models in real-world settings.

6.3.8. External Validation: The study focused on using specific non-clinical datasets, and external validation on diverse datasets from different sources was not performed. Validating the model's performance on independent datasets is essential to establish its robustness and generalizability.

Addressing these limitations and challenges is critical for the successful translation of machine learning models into practical ASD detection tools. Collaborations with clinicians and domain experts can help overcome some of these limitations, ensuring that the models align with clinical practices and address real-world challenges.

6.4 Future Research Directions

As the field of ASD diagnosis using machine learning continues to evolve, there are several key areas that warrant further exploration and research. Future studies in ASD diagnosis can focus on the following aspects to enhance the accuracy, generalizability, and practicality of machine learning models:

6.4.1.Incorporating Additional Features: While the current study utilized a set of relevant features for ASD detection, there is potential to explore additional features that might improve the models' performance. The inclusion of genetic markers, neuroimaging data, and data from wearable devices could provide valuable insights into the neurobiological basis of ASD and enhance the predictive power of the models.

6.4.2.Alternative Algorithms: Although the Convolutional Neural Network (CNN) demonstrated remarkable performance in this study, there are various other machine learning algorithms worth exploring. Future research can investigate the effectiveness of recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and attention-based models for ASD detection. Comparing the performance of different algorithms can lead to a better understanding of their suitability for specific age groups and dataset characteristics.

6.4.3.Improving Methodologies: Advanced pre-processing techniques and feature selection methods can play a crucial role in optimizing model performance. Fine-tuning hyperparameters and conducting feature engineering can lead to improved results. Additionally, ensemble methods that combine multiple models can be explored to leverage the strengths of different algorithms and enhance overall predictive accuracy.

6.4.4.Larger and Diverse Datasets: Access to larger and more diverse datasets is paramount for robust model training and validation. Future studies should aim to collaborate with multiple research institutions and hospitals to collect extensive and varied data samples. Incorporating data from different cultural backgrounds, socioeconomic groups, and geographical regions can help create more inclusive models that perform well across diverse populations.

6.4.5.Ethical Considerations: As machine learning models become more integral to clinical decision-making, it is essential to address ethical considerations. Future research should focus on ensuring the transparency and interpretability of machine learning algorithms, as well as exploring ways to mitigate potential biases and disparities in the models' predictions.

6.4.6.Real-World Validation: The current study utilized non-clinical datasets for ASD detection. Future research should include real-world validation with clinical datasets to assess the models' performance in actual clinical settings. Collaborating with healthcare professionals and domain experts can provide valuable feedback and ensure the models align with clinical practices.

6.4.7.Longitudinal Studies: Longitudinal studies tracking the developmental trajectories of individuals with ASD can provide insights into the progression of the disorder over time. Machine learning models can be used to analyze these longitudinal data and identify patterns associated with the development and management of ASD.

6.4.8.Exploring Multi-Modal Data: Integrating multiple sources of data, such as behavioral, genetic, and neuroimaging data, can lead to a more comprehensive understanding of ASD. Multi-modal machine learning approaches can be explored to leverage the complementary information from diverse data sources.

In conclusion, future research in ASD diagnosis using machine learning holds great promise in advancing our understanding of the disorder and improving early detection and intervention. By exploring additional features, alternative algorithms, and improved methodologies, along with larger and diverse datasets, researchers can build more accurate and reliable models for ASD detection. Additionally, addressing ethical considerations and conducting real-world validation will pave the way for the effective integration of machine learning into clinical practices, ultimately leading to better outcomes for individuals with ASD and their families.

References

1. Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward." Informatics for Health and Social Care 1-20, 2018.

2. Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets." International Journal of Engineering & Technology 7(4), 2018: 18.

3.Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets. (2018) " International

Journal of Engineering & Technology 7(4): 18.

4. Bone, Daniel, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. "Applying machine learning to facilitate autism diagnostics: pitfalls and promises." Journal of autism and developmental disorders 45(5), 2015: 1121-1136.

5. Wall, Dennis Paul, J. Kosmicki, T. F. Deluca, E. Harstad, and Vincent Alfred Fusaro. "Use of machine learning to shorten observation-based screening and diagnosis of autism." Translational psychiatry 2(4), 2012: e100.

6. Wall, Dennis P., Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, and Todd F. DeLuca. "Use of artificial intelligence to shorten the behavioral diagnosis of autism." PloS one 7(8), 2012: e43855.

7. Thabtah, Fadi. "ASD Tests. A mobile app for ASD screening." www.asdtests.com, accessed December 20th, 2017.

8. Li, Baihua, Arjun Sharma, James Meng, Senthil Purushwalkam, and Emma Gowen. "Applying machine learning classifiers to detect autistic adults by imitation method." PloS one 12(8), 2017: e0182652.

9. Fadi Fayez Thabtah. "Autistic Spectrum Disorder Screening Data for Adult." https://archive.ics.uci.edu/ml/machine-learning-databases/00426/, 2017.

10. Mythili, M. S., and AR Mohamed Shanavas. "A study on Autism spectrum disorders using classification techniques." International Journal of Soft Computing and Engineering (IJSCE) 4, 2014: 88-91.

11. Kosmicki, J. A., V. Sochat, M. Duda, and D. P. Wall. "Searching for a minimal set of traits for autism detection through feature selection-based machine learning." Translational psychiatry 5(2), 2015: e514.

12. Constantino, John N., Patricia D. Lavesser, Y. I. Zhang, Anna M. Abbacchi, Teddi Gray, and Richard D. Todd. "Rapid quantitative assessment of autistic social impairment by classroom teachers." Journal of the American Academy of Child & Adolescent Psychiatry 46(12), 2007: 1668-1676.

13. Fadi Fayez Thabtah. "Autistic Spectrum Disorder Screening Data for children." https://archive.ics.uci.edu/ml/machine-learning-databases/00419/, 2017.

14. Fadi Fayez Thabtah. "Autistic Spectrum Disorder Screening Data for Adolescent." https://archive.ics.uci.edu/ml/machine-learning-databases/00420/.

15. John, George H., and Pat Langley. "Estimating continuous distributions in Bayesian classifiers." Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., 1995: 338-345.

16. Quinlan, J. R. "Program for machine learning. C4.5." 1993.

17. Keerthi, S. Sathiya, Shirish KrishnajShevade, Chiranjib Bhattacharyya, and Karuturi Radha Krishna Murthy. "Improvements to Platt's SMO algorithm for SVM classifier design." Neural computation 13(3), 2001: 637-649.

18. APA Style. Publication Manual of the American Psychological Association, 7th edition, 2019.