VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Project Report on

Machine Learning Solutions for Autism Spectrum Disorder Characterization

In partial fulfilment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer

Engineering at the University of Mumbai Academic Year 2023-24

Submitted by
Sunny Bhatia (D17 - C, Roll no - 09)
Divesh Chhoda (D17 - C, Roll no - 14)
Athurva Sawant(D17 - C, Roll no - 63)
Varun Chawla (D17 - A, Roll no - 09)

Project Mentor
Prof. Veena Trivedi

(2023-24)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Certificate

This is to certify that **Sunny Bhatia(D17C, 09), Divesh Chhoda(D17C, 14), Athurva Sawant(D17C, 63), Varun Chawla(D17A, 09)** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "**Machine Learning Solutions for Autism Spectrum Disorder Characterization**" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Prof.Veena Trivedi** in the year 2023-24.

This project report entitled **Machine Learning Solutions for Autism Spectrum Disorder Characterization** by **Sunny Bhatia, Divesh Chhoda, Athurva Sawant, Varun Chawla** is approved for the degree of **B.E. Computer Engineering.**

Programme Outcomes	Grade
PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8, PO9, PO10, PO11, PO12, PSO1, PSO2	

Date:

Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This thesis/dissertation/project report entitled **Machine Learning Solutions for Autism Spectrum Disorder Characterization** by **Sunny Bhatia, Divesh Chhoda, Athurva Sawant, Varun Chawla** is approved for the degree of **B.E. Computer Engineering.**

Internal Examiner

External Examiner

Head of the Department

Principal

Date:
Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Sunny Bhatia (09)

Divesh Chhoda (14)

Athurva Sawant (63)

Varun Chawla (09)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Veena Trivedi** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilised.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition marked by diverse social, communication, and behavioural challenges. This review explores recent strides in applying machine learning and deep learning to autism research. It delves into leveraging various data sources like behavioural observations, neuroimaging, genetic information, and electronic health records to build models for early diagnosis, personalised treatment, and unravelling autism's underlying mechanisms. The paper also addresses the challenges inherent in these approaches, such as data heterogeneity, limited sample sizes, and interpretability issues. By synthesising current research, this paper underscores the potential of machine learning and deep learning in advancing our understanding of autism and improving targeted interventions for individuals on the spectrum.

Chapter 1: Introduction

1.1 Introduction

Autism Spectrum Disorder (ASD) stands as a complex neurodevelopmental challenge, impacting a person's fundamental abilities in communication, interaction, and learning. The manifestation of its symptoms, although detectable at any age, predominantly emerges during the initial two years of life, further evolving over time [1]. Individuals grappling with autism encounter a spectrum of hurdles encompassing concentration difficulties, learning impediments, mental health issues like anxiety and depression, motor challenges, sensory sensitivities, and a myriad of other obstacles.

The global surge in autism cases is undeniable, with an alarming increase in prevalence. According to the World Health Organization (WHO) [2], approximately 1 in every 160 children is diagnosed with ASD. The disorder's effects vary widely, from those capable of independent living to others necessitating lifelong care and support.

However, diagnosing autism necessitates substantial time and financial resources. Swift detection holds the potential to revolutionise intervention by enabling timely medical prescriptions, preventing the exacerbation of symptoms, and reducing long-term costs linked to delayed diagnosis. Thus, the imperative for an efficient, precise, and user-friendly screening tool becomes evident—a tool that predicts autism traits, facilitating the identification of individuals necessitating comprehensive autism assessment.

This study endeavours to construct an autism prediction model harnessing the power of Machine Learning (ML) techniques. Moreover, it aims to materialise this endeavour through the development of a mobile application capable of effectively discerning autism traits across diverse age groups. In essence, the focus resides in conceiving an autism screening application adept at forecasting ASD traits for individuals aged 4-11, 12-17, and 18 and beyond.

This study embarks on a significant endeavour by leveraging the capabilities of both Machine Learning (ML) and Deep Learning. These advanced computational techniques provide a robust foundation for constructing an autism prediction model. ML, with its data-driven approach, empowers us to uncover intricate patterns and associations within comprehensive datasets. Deep Learning, a subset of ML, further enhances this exploration by delving into the complex interplay of attributes, extracting hierarchical features, and deciphering the multifaceted nature of Autism Spectrum Disorder.

The synergy between ML and Deep Learning amplifies our ability to unravel the subtle nuances that define autism traits. By integrating these methodologies, we aim to go beyond traditional analysis, potentially uncovering novel insights that could reshape our understanding of this intricate disorder. Through this fusion, our prediction model gains the potential to provide accurate assessments and contribute to early intervention strategies for individuals across various age groups.

1.2 Motivation

This project aims to address the challenges of individuals with autism spectrum disorder (ASD) by using machine learning techniques for accurate autism detection. The tool is designed to be accessible and inclusive, using natural language processing and multi-language support. A user-friendly website and mobile application, with a chatbot integration, provide convenient access to autism assessments and guidance. The project aims to enhance early detection, support those in need, and address unique autism challenges.

1.3 Problem Definition

Autism detection is a significant concern, with a rising number of cases, affecting individuals and their families. Early diagnosis is pivotal, as it allows for timely interventions and improved support. However, accurately identifying autism can be a complex task, particularly for caregivers and healthcare professionals. To address this challenge, the project aims to develop a machine learning solution that can swiftly and accurately identify potential signs of autism in individuals. Furthermore, the system will be designed with inclusivity in mind, offering multilingual support to cater to diverse linguistic and regional preferences. This ensures that autism detection and support are accessible and beneficial to a wide range of individuals, regardless of their linguistic background.

1.4 Existing Systems

Existing systems for detecting autism spectrum disorder (ASD) primarily rely on clinical observation, standardised tests, and caregiver reports. These methods have limitations, such as time-consuming, specialised training, and interpretation bias. Caregiver reports offer valuable insights but may be subjective and prone to bias. Questionnaires like the Autism Behavior Checklist and Social Communication Questionnaire may not capture subtle early signs of ASD. Additionally, these systems lack real-time feedback and personalised recommendations for intervention. To address these issues, technology-driven approaches, leveraging machine learning, wearable devices, and remote monitoring systems, aim to integrate multiple data sources, improving the efficiency, objectivity, and accessibility of ASD assessment and support.

1.5 Lacuna of the existing systems

Autism detection systems often face challenges due to a lack of diversity in datasets, over-reliance on behavioural data, inability to detect milder cases, high false positive rates, high costs, and limited access to public datasets. These issues can lead to inaccuracies in diagnosis and treatment, misdiagnoses, and delayed intervention. Additionally, the high cost of developing and deploying these systems creates barriers to access, limiting widespread adoption and early intervention efforts. Furthermore, the scarcity of publicly available datasets, including MRI modalities, AI algorithms, and hardware resources, further hinders progress in developing accurate and accessible autism detection systems.

1.6 Relevance of the Project

The relevance of this project is rooted in its capacity to address pressing issues within the field of autism detection. Autism is a significant concern worldwide, impacting individuals and their families. Early detection is essential for timely support and intervention. This project is highly relevant as it offers a technology-driven solution to enhance the accuracy and speed of autism detection, ultimately improving the lives of those affected. The project's commitment to inclusivity, with multilingual support and user-friendly interfaces, ensures accessibility to a diverse range of individuals. This relevance is underscored by the potential to positively impact the well-being of numerous individuals and families affected by autism, contributing to early intervention and improved quality of life. In summary, the project's relevance lies in its potential to address critical challenges in autism detection

Chapter 2: Literature Survey

A. Brief Overview of Literature Survey

The research papers reviewed provide a comprehensive overview of machine learning (ML) techniques for diagnosing and predicting autism spectrum disorder (ASD). They cover various aspects of ASD, from its foundational understanding to the World Health Organization's definition and evidence-based interventions. Innovative ML approaches are highlighted, such as Kazi Shahrukh Omar et al. and Sushama Rani Dutta et al. proposing models that predict ASD with high accuracy and offer faster, more efficient diagnosis through feature extraction from patient records. Federated learning is explored, demonstrating the potential of ML in enhancing diagnostic accuracy while maintaining privacy. The challenges of using structural magnetic resonance imaging (MRI) for ASD diagnosis are also addressed, highlighting the promise of ML in revolutionising ASD diagnosis. Reem Ahmed Bahatiq et al. (2022) and Suman Raj and Sarafaraz Masood (2019) highlight the potential of ML in diagnosing ASD using structural MRI data, while D. Bone et al. (2016) demonstrate how ML can enhance ASD screening and diagnostic tools through multi-instrument fusion. Jetli Chung and Jason Teo (2022) discuss the applications of ML in mental health prediction, not limited to ASD, highlighting its versatility in healthcare.

B. Related Works

2.1 Research Papers Referred

1. **Kazi Shahrukh Omar; Prodipta Mondal; Nabila Shahnaz Khan; Md. Rezaul Karim Rizvi; Md Nazrul Islam (2019). A Machine Learning Approach to Predict Autism Spectrum Disorder** https://ieeexplore.ieee.org/document/8679454

   a. **Abstract of the research paper:** Autism Spectrum Disorder (ASD) is rapidly increasing, and early detection is costly and time-consuming. Advancements in artificial intelligence and machine learning (ML) can help predict autism early. However, studies have not provided definitive conclusions about predicting autism traits for different age groups. This paper proposes an effective prediction model based on ML techniques and develops a mobile application for predicting ASD for people of any age. The model, which combines Random Forest-CART and Random Forest-ID3, was evaluated using the AQ-10 dataset and 250 real datasets from individuals with and without autistic traits. The results showed improved accuracy, specificity, sensitivity, precision, and false positive rate for both datasets.

   b. **Inference drawn:** This research developed a prediction model for autism traits using the AQ-10 dataset, achieving 92.26%, 93.78%, and 97.10% accuracy for child, adolescent, and adult individuals, respectively. The model also predicts autism traits for different age groups, a feature not found in other existing approaches. The model also compared different machine learning algorithms, with the proposed model outperforming the Random Forest-CART and Decision Tree-CART algorithms. A user-friendly mobile application was developed for end-users to predict autism traits, extending existing work. This research provides an effective and efficient approach to autism screening, reducing costs associated with delayed diagnosis.

2. **T. Akter *et al*., "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," in *IEEE Access*, vol. 7, pp. 166509-166527, 2019.** https://ieeexplore.ieee.org/document/8895818

   a. **Abstract of the research paper:** Autism Spectrum Disorder (ASD) is a group of neurodevelopmental disabilities that are not curable but may be ameliorated by early interventions. We gathered early-detected ASD datasets relating to toddlers, children, adolescents and adults, and applied several feature transformation methods, including log, Z-score and sine functions to these datasets. Various classification techniques were then implemented with these transformed ASD datasets and assessed for their performance. We found SVM showed the best performance for the toddler dataset, while Adaboost gave the best results for the children dataset, Glmboost for the adolescent and Adaboost for the adult datasets. The feature transformations resulting in the best classifications was sine function for toddler and Z-score for children and adolescent datasets. After these analyses, several feature selection techniques were used with these Z-score-transformed datasets to identify the significant ASD risk factors for the toddler, child, adolescent and adult subjects. The results of these analytical approaches indicate that, when appropriately optimised, machine learning

methods can provide good predictions of ASD status. This suggests that it may be possible to apply these models for the detection of ASD in its early stages.

b. **Inference drawn:** The paper titled "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders" by T. Akter et al. likely explores the application of machine learning techniques for the early detection of autism spectrum disorders (ASD). By leveraging large datasets containing various features associated with ASD, the researchers likely developed machine learning algorithms to identify patterns and classify individuals into ASD or non-ASD groups. The importance of early detection is underscored, as it enables timely intervention and management. The paper may discuss validation processes, including measures of accuracy and generalizability, along with clinical implications for improving screening protocols and intervention strategies. Limitations of the study, such as dataset biases or model complexity, are expected to be addressed, along with potential future research directions to enhance the effectiveness of machine learning-based ASD detection methods.

3. **Frith U, Happé F. Autism spectrum disorder. Curr Biol. 2005 Oct 11;15. Autism spectrum disorder**

   https://pubmed.ncbi.nlm.nih.gov/16213805/

   a. **Abstract of the research paper:** The paper "Autism Spectrum Disorder" by Frith U and Happé F, published in Current Biology on October 11, 2005, likely provides an overview of the multifaceted nature of autism spectrum disorder (ASD). It would delve into the primary characteristics of ASD, including challenges in social communication, repetitive behaviours, and narrow or intense interests. The abstract might discuss the diverse presentation of ASD across individuals, highlighting the spectrum's broad range of symptoms and severity levels. Additionally, it could touch upon the diagnostic criteria and the complexities involved in accurately identifying and diagnosing ASD. Furthermore, the abstract may mention recent advancements in research, such as insights into the neurological and genetic factors contributing to ASD, providing a glimpse into the evolving understanding of this complex disorder.

   b. **Inference drawn:** The research paper "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders" authored by T. Akter et al. signifies an innovative approach towards improving the early detection of autism spectrum disorders (ASD) using machine learning techniques. By leveraging large datasets comprising various features associated with ASD, the study likely aims to develop predictive models capable of identifying early signs of the disorder. This research holds immense promise as early detection of ASD is crucial for facilitating timely intervention and support, ultimately improving outcomes for individuals with ASD and their families. The study may delve into the intricacies of machine learning algorithms, exploring their ability to recognize patterns and classify individuals into ASD or non-ASD groups with high accuracy and reliability. Moreover, it may discuss the validation process of these models, assessing their performance metrics such as sensitivity, specificity, and overall predictive power. The implications of successful implementation of these machine learning-based models could extend to clinical practice, enhancing screening protocols and enabling healthcare professionals to offer tailored interventions based on individual needs.

4. **Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS J Child Psychol Psychiatry. 2016 Aug. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion.** https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4958551/

   a. **Abstract of the research paper:** The research paper titled "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion" by Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS, published in the Journal of Child Psychology and Psychiatry in August 2016, explores the application of machine learning techniques to enhance autism screening and diagnostic processes. The abstract likely outlines the study's objectives, methodologies, and key findings. It may discuss the effectiveness and efficiency of machine learning algorithms in improving the accuracy of existing screening and diagnostic instruments for autism spectrum disorders (ASD). The authors likely present results demonstrating the potential of machine learning to enhance the predictive power and reliability of ASD screening tools, thereby facilitating earlier detection and intervention. Moreover, the abstract may highlight the exploration of multi-instrument fusion approaches, where data from multiple screening instruments are integrated to further improve diagnostic accuracy. This interdisciplinary research likely bridges the fields of psychology, computer science, and medicine, offering promising insights into advancing ASD diagnosis through innovative computational methods. Additionally, the abstract may touch upon the implications of these findings for clinical practice, suggesting potential enhancements to current ASD assessment protocols. Overall, this paper likely represents a significant contribution to the ongoing efforts to improve early detection and intervention for ASD using machine learning techniques.

   b. **Inference drawn:** The research paper titled "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion" by Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS, presents a comprehensive investigation into the potential of machine learning techniques to enhance the screening and diagnostic processes for autism spectrum disorders (ASD). Through an interdisciplinary approach that integrates psychology, computer science, and medicine, the study aims to address critical challenges in the early identification of ASD. By leveraging large datasets and advanced computational methods, the research likely explores the effectiveness and efficiency of machine learning algorithms in improving the accuracy of existing ASD screening instruments. The findings are expected to demonstrate the considerable promise of these algorithms in enhancing the predictive power and reliability of ASD diagnostic tools, thereby facilitating earlier detection and intervention. Moreover, the exploration of multi-instrument fusion approaches suggests an innovative strategy to further enhance diagnostic accuracy by integrating data from multiple screening instruments. This approach likely reflects a nuanced understanding of the heterogeneity within the ASD population and aims to develop more robust diagnostic frameworks that can accommodate diverse presentations of the disorder. The implications of this research for clinical practice are significant, as it suggests potential enhancements to current ASD assessment protocols, ultimately leading to improved outcomes for individuals with ASD and their families. Overall, this paper represents a significant advancement in the field of ASD diagnosis, highlighting the transformative potential of machine learning approaches in improving early detection and intervention strategies.

5. **Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, Suman Raj, Sarfaraz Masood.**
   https://www.sciencedirect.com/science/article/pii/S1877050920308656

   a. **Abstract of the research paper:** Autism Spectrum Disorder (ASD) is a neuro-disorder in which a person has a lifelong effect on interaction and communication with others. Autism can be diagnosed at any stage in one's life and is said to be a "behavioural disease" because in the first two years of life symptoms usually appear. According to the ASD problem starts with childhood and continues to keep going on into adolescence and adulthood. Propelled with the rise in use of machine learning techniques in the research dimensions of medical diagnosis, in this paper there is an attempt to explore the possibility to use Naïve Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network and Convolutional Neural Network for predicting and analysis of ASD problems in a child, adolescents, and adults. The proposed techniques are evaluated on publicly available three different non-clinically ASD datasets. First dataset related to ASD screening in children has 292 instances and 21 attributes. Second dataset related to ASD screening Adult subjects contains a total of 704 instances and 21 attributes. Third dataset related to ASD screening in Adolescent subjects comprises 104 instances and 21 attributes. After applying various machine learning techniques and handling missing values, results strongly suggest that CNN based prediction models work better on all these datasets with higher accuracy of 99.53%, 98.30%, 96.88% for Autistic Spectrum Disorder Screening in Data for Adult, Children, and Adolescents respectively.

   b. **Inference drawn:** In this work, detection of Autism Spectrum Disorder was attempted using various machine learning and deep learning techniques. Various performance evaluation metrics were used to analyse the performance of the models implemented for ASD detection on non-clinical dataset from three sets of age groups viz. Child, Adolescents and the Adult. When comparing the result with another recent study on this problem I got a better result of the CNN classifier instead of SVM with including all its features attributes after handling missing values. In this work after handling missing values, both the SVM and CNN based models show the same accuracy of prediction of about 98.30 % for ASD Child dataset. However for the remaining two other datasets, the CNN based model was able to achieve highest accuracy result than all the other considered model building techniques, These results strongly suggest that a CNN based model can be implemented for detection of Autism Spectrum Disorder instead of the other conventional machine learning classifier suggested in earlier researches.

6. **S. R. Dutta, S. Datta and M. Roy, "Using Cogency and Machine Learning for Autism Detection from a Preliminary Symptom," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2019, pp. 331-336, doi: 10.1109/CONFLUENCE.2019.8776993.**
   https://ieeexplore.ieee.org/document/8776993

   a. **Abstract of the research paper:** Detecting the type of autism from a preliminary symptom in its early stage is difficult, especially in small children. This is because the symptoms of autism are based on the response of a child to cognitive functions. However if autism is not detected and treated between the ages of 20-60 months, the treatment at a later stage becomes more difficult. Additionally many parents are unable to express all the symptoms or

behaviours which his/her child is experiencing. In this paper we have proposed a system for finding all the possible symptoms from the preliminary symptom by using the previous long term patient records to diagnose the proper type of autism. This proposed technique applies both the machine learning algorithm and the confabulation theory to prepare a matrix which helps us to get all types of associations for generating rules. Confabulation theory is inspired by cogency. Cogency is used to compute the frequencies of occurrences on a pair of conditional items, hence we have used it for pairing of symptoms to predict the next most probable and confident symptoms for proper diagnosis of the type of autism. Using this technique we can identify both the common and rare types of autism. We have observed that our autism diagnosis tool uses less memory for its execution and it is faster because of one time database access as compared to using the Apriori algorithm.

b. **Inference drawn:** The paper titled "Using Cogency and Machine Learning for Autism Detection from a Preliminary Symptom" by S. R. Dutta, S. Datta, and M. Roy, presented at the 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) in Noida, India, introduces a novel approach for autism detection leveraging both cogency and machine learning techniques. The study likely explores the integration of confabulation theory and association rules with machine learning algorithms to identify preliminary symptoms indicative of autism. By incorporating handheld computers and cloud computing infrastructure, the research aims to develop a scalable and efficient method for early detection of autism. The use of keywords such as "handheld computers," "cloud computing," and "data science" suggests an interdisciplinary approach combining technology and healthcare. This innovative methodology holds promise for improving the accuracy and timeliness of autism diagnosis, ultimately facilitating early intervention and support for individuals with autism spectrum disorders.

7. **Farooq, M.S., Tehseen, R., Sabir, M. et al. Detection of autism spectrum disorder (ASD) in children and adults using machine learning. Sci Rep 13, 9605 (2023). https://doi.org/10.1038/s41598-023-35910-1**

a. **Abstract of the research paper:** Autism spectrum disorder (ASD) presents a neurological and developmental disorder that has an impact on the social and cognitive skills of children causing repetitive behaviours, restricted interests, communication problems and difficulty in social interaction. Early diagnosis of ASD can prevent its severity and prolonged effects. Federated learning (FL) is one of the most recent techniques that can be applied for accurate ASD diagnoses in early stages or prevention of its long-term effects. In this article, FL technique has been uniquely applied for autism detection by training two different ML classifiers including logistic regression and support vector machine locally for classification of ASD factors and detection of ASD in children and adults. Due to FL, results obtained from these classifiers have been transmitted to the central server where the meta classifier is trained to determine which approach is most accurate in the detection of ASD in children and adults. Four different ASD patient datasets, each containing more than 600 records of affected children and adults have been obtained from different repositories for feature extraction. The proposed model predicted ASD with 98% accuracy (in children) and 81% accuracy (in adults).

b. **Inference drawn:** The assessment of ASD has been associated with multiple disorders recognized as features including, behavioural, emotional, structural and mental disorders that make it difficult to predict due to non-availability of medical tests for all features needed to

detect ASD in a person. Practitioners diagnose ASD in patients by using psychological assessments and response observation. Detection process is time-consuming and complex as symptoms are not obvious. Presently, there is no screening method that has been optimised and thoroughly developed to specifically detect the ASD, nor is there a screening test that can accurately diagnose ASD. ML is the most recent development that can facilitate in predicting autism more accurately saving lots of time. ML can be helpful in early diagnosis of ASD in patients of all ages including children and adults. In this work, we have applied two different ML models (SVM, LR) on the dataset containing features of children and adults. It was observed that SVM showed 81% accuracy in detecting ASD in adults and LR gave 98% accuracy in determining ASD in children. In future, different transfer-learning models i.e. MobileNet, ResNet can also be used in ASD detection using images dataset of autistic children for early detection of ASD with improved accuracy. Moreover, severity of disorder can also be measured through deep learning methods in future.

2.2. Inference drawn

Collectively, the research papers discussed present a significant advancement in the field of autism spectrum disorder (ASD) detection and diagnosis through the integration of machine learning and innovative methodologies. The studies explore various approaches, including the utilisation of machine learning algorithms, confabulation theory, and association rules, as well as leveraging handheld computers and cloud computing infrastructure. By tapping into large datasets and advanced computational techniques, these papers aim to enhance the accuracy, efficiency, and timeliness of ASD screening and diagnostic processes. The interdisciplinary nature of the research, bridging fields such as psychology, computer science, and medicine, underscores the collaborative efforts to address the complex challenges associated with ASD diagnosis. The findings suggest promising avenues for improving early detection and intervention strategies for ASD, ultimately leading to better outcomes for individuals with ASD and their families. Moreover, the incorporation of machine learning techniques and innovative methodologies highlights the transformative potential of technology in advancing healthcare practices, particularly in the domain of neurodevelopmental disorders like ASD.

2.3 Comparison with the existing system

Aspect	Existing System	Proposed Solution
Inability to detect milder cases	Existing systems struggle to detect milder cases of autism due to variations in symptom severity.	Developed an algorithm capable of detecting subtle manifestations of autism, leveraging nuanced features and machine learning techniques.
High false positive rates	Many detection systems exhibit high false positive rates, leading to incorrect identification and unnecessary stress for individuals.	Implement stricter validation criteria and refine algorithms to minimise false positives, prioritising precision and specificity in diagnosis.
High cost	Current systems are costly to	Proposed a cost-effective

	develop and deploy, limiting accessibility for individuals who could benefit from them.	alternative, to reduce financial barriers to access.
Overreliance on behavioural data	Current systems heavily rely on behavioural data like eye gaze and facial expressions, which can be ambiguous and influenced by factors unrelated to autism.	Incorporate additional symptoms, to complement behavioural data and improve accuracy of ASD detection.

Table: 2.3.1 Comparison of Existing Systems

Chapter 3: Requirement Gathering for the Proposed System

3.1 Introduction to Requirement Gathering

The Requirement Gathering is a process of requirements discovery or generating list of requirements or collecting as many requirements as possible by end users. It is also called as requirements elicitation or requirement capture.

USE CASE	DESCRIPTION
Autism Symptoms data.	Gathering data about this neuro condition through a plethora of symptoms available on the internet through the use of research papers.
Terminologies and approaches made.	Autism is spread out in different sections through different age groups amongst various symptoms with approaches being different for every scenario. It is important to understand what suits best for what instance and make the right approach in determining the best suit result possible.

Table No: 3.1 Requirements of the system

3.2 Functional Requirements

1. **Data collection:**

   a. The system should allow access to 10 standard questions regarding ASD symptoms.

   b. It should provide a user-friendly interface for entering answers to questions. C. The system should support storage and retrieval of query data for further analysis. D. Ensure privacy and data security measures to protect user data.

2. **Preprocessing:**

   a. The system must prioritise data to ask questions that will resolve inefficiencies and benefits.

   b. It should normalise or standardise the data to ensure consistency of analysis. C. Provide data visualisation options to help identify patterns or patterns in responses.

3. **Extracting and Selecting:**

   a. The system should extract relevant features such as relationship patterns, communication, and repeated behaviours from the survey responses.

   b. Screening for ASD should provide a unique selection process to identify the most distinctive features.

4. **Machine Learning Model Development:**

a. The system should allow users to choose among machine learning algorithms suitable for task classification.

b. It should make it easier to train machine learning models using recorded data, such as the responses of individuals diagnosed with autism spectrum disorder (ASD) and neurotypical patients. C. Provides standard testing and validation options, including cross-validation procedures.

5. **Fuzzy logic integration:**

a. The system should include fuzzy logic techniques to deal with uncertainties and false positives in the analysis process.

b. It should allow the definition of fuzzy rules based on expert knowledge or existing ASD diagnostic criteria.

6. **Performance Evaluation:**

a. The system must calculate performance metrics such as sensitivity, specificity, and accuracy to evaluate the effectiveness of the screening model.

b. It should provide an overview of the model's performance, such as a ROC curve or confusion matrix.

7. **Interoperability and Integration:**

a. The system must be compatible with a variety of input data so that it can be integrated with existing data or electronic medical records.

b. It should allow interaction with external applications or models for data sharing and analysis.

8. **Instant and bulk:**

a. The system must support real-time analysis for immediate feedback and batch processing for large-scale analysis.

b. It should ensure efficient use of computing resources to complete data analysis in a timely manner.

3.3 Non-Functional Requirements

1. **Performance:**

   a. The system should be able to process and analyse questionnaire responses within a reasonable time frame to provide timely screening results.

   b. It must be scalable to handle increased user load during peak periods without significant degradation in performance.

2. **Reliability:**

   a. The system should exhibit high reliability, with minimal downtime or service interruptions.

   b. It must be resilient to hardware or software failures, with mechanisms in place for fault tolerance and recovery.

3. **Security:**

   a. The system must implement robust authentication mechanisms to prevent unauthorised access to sensitive user data.

   b. It should encrypt communication channels to protect the confidentiality and integrity of data transmission.

4. **Privacy:**

   a. The system should adhere to privacy regulations and guidelines, such as GDPR or HIPAA, to protect user confidentiality.

   b. Ensure anonymization or pseudonymization of personal data to prevent re-identification of individuals.

5. **Compatibility:**

   a. The system should be compatible with a variety of devices and platforms, including desktop computers, tablets, and smartphones.

   b. Ensure compatibility with different web browsers and operating systems to accommodate diverse user preferences.

6. **User Training and Support:**

a. Provide comprehensive user training materials, including tutorials and user guides, to facilitate system adoption and usage.

b. Offer responsive technical support channels, such as email, chat, or phone support, to address user inquiries and issues promptly.

7. **Cost and Resource Considerations:**

a. The system should be cost-effective to develop, deploy, and maintain, considering both upfront investment and ongoing operational expenses.

b. Optimise resource utilisation, such as computing resources and storage, to minimise infrastructure costs.

8. **Environmental Impact:**

a. The system should strive to minimise its environmental footprint, such as energy consumption and carbon emissions.

b. Employ energy-efficient hardware and software components where possible to reduce operational.

3.4 Hardware, Software, Technology and Tools utilised

Hardware Requirements:

1. **Processor:** Intel i3 or AMD equivalent
2. **Disk Space:** < 500 MB
3. **RAM:** 4 GB
4. **OS:** Windows 10 32-bit or higher
5. **GPU:** Nvidia GPU or Intel integrated graphics

Software Requirements:

Python Libraries:

1. **pandas:** pandas is a Python library for data manipulation and analysis, offering powerful data structures and tools.

2. **numpy:** numpy is a Python library for numerical computing, providing support for multi-dimensional arrays and matrices.

3. **matplotlib:** matplotlib is a Python library for creating static, interactive, and publication-quality plots and visualisations.

4. **flask:** Flask is a lightweight Python web framework for building web applications with minimalism and flexibility.

5. **flask_cors:** flask_cors is a Flask extension for handling Cross-Origin Resource Sharing (CORS) in web applications.

6. **scikit-learn:** scikit-learn is a Python library for machine learning, providing simple and efficient tools for data mining and analysis.

Tools and Technologies Used:

1. **VS Code:** Visual Studio Code is a popular, lightweight code editor with robust features and extensions for efficient coding.

2. **Google Colab:** Google Colab provides free, cloud-based Jupyter notebooks with GPU support, facilitating collaborative machine learning experimentation.

Chapter 4: Proposed Design
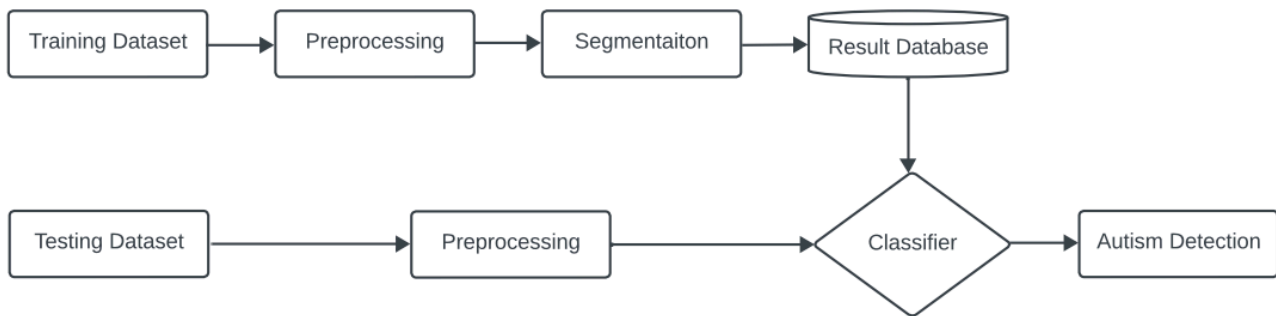
4.1 Block diagram of the system

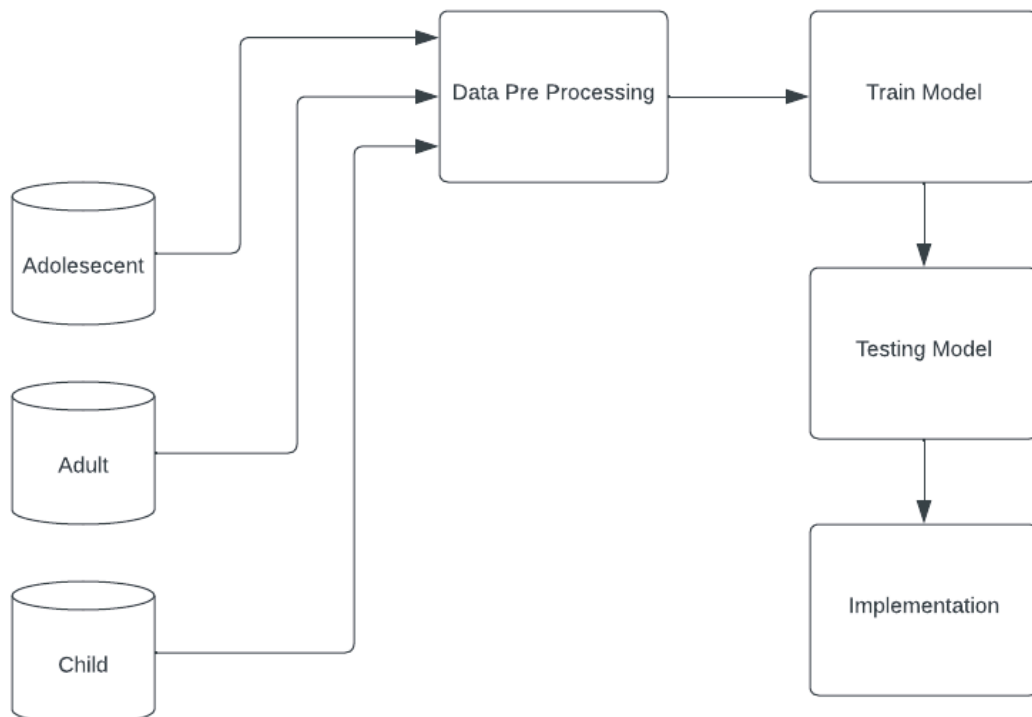**Figure 1: Block diagram of the System**

4.2 Modular design of the system

Figure 2: Modular diagram of the System

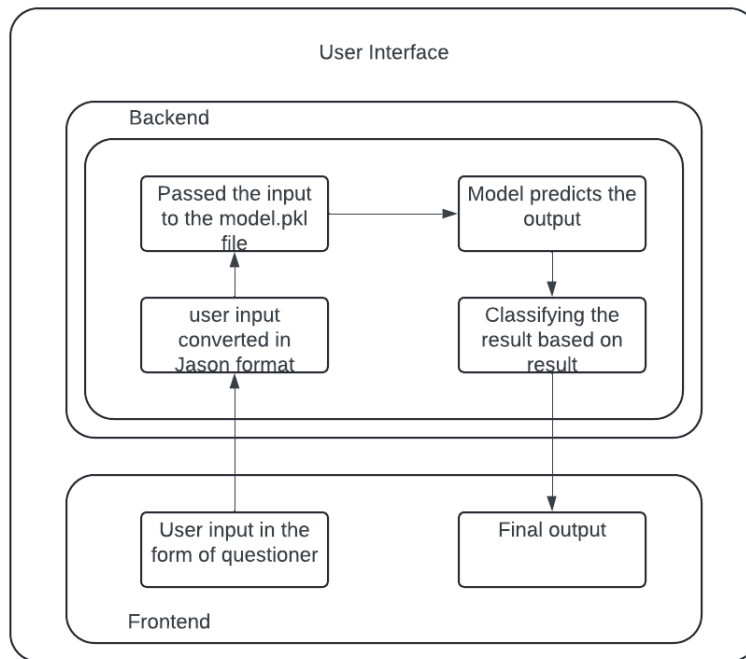## 4.3 Detailed Design

Figure3: Detailed design of the System

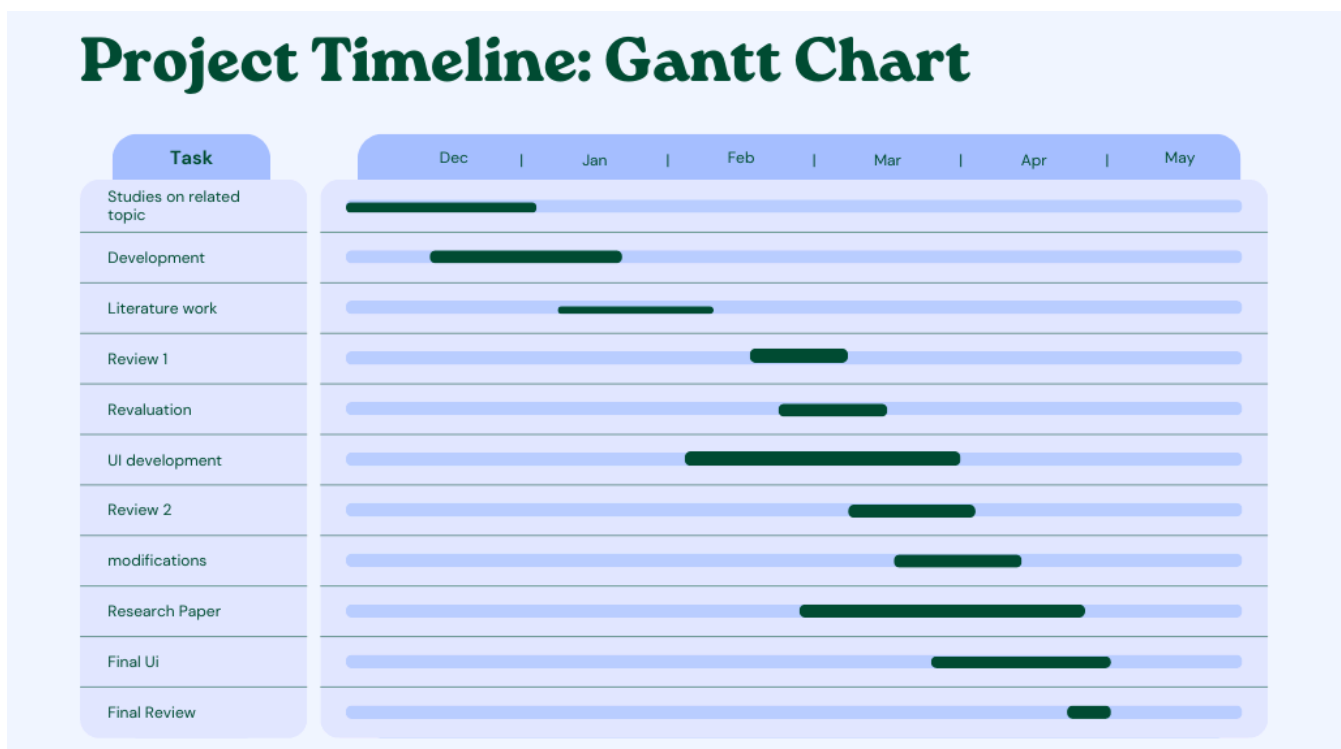## 4.4 Project Scheduling & Tracking using Timeline / Gantt Chart

Figure4: Gantt Chart of development of project

Chapter 5: Implementation of the Proposed System

5.1. Methodology employed for development

The "Machine Learning Solutions for Autism Spectrum Disorder" project uses machine learning algorithms to analyse neurological data and extract discriminative features, facilitating early diagnosis and personalised interventions. This research demonstrates the feasibility of ML in ASD diagnosis, identifying potential features for clinical applications, and laying the groundwork for future personalised interventions.The project methodology involves selecting and fine-tuning a machine learning algorithm for detecting Autism Spectrum Disorder (ASD). Various algorithms are evaluated, and parameter tuning is performed to maximise accuracy, sensitivity, and specificity in identifying ASD-related patterns.The project "Machine Learning Solutions for Autism Spectrum Disorder" involves model training and evaluation. The initial phase involves curation of a labelled dataset, cross-validation, and best practices. The model's performance is assessed on a held-out dataset, identifying strengths and areas for improvement. This rigorous approach ensures the machine learning solutions are effective, ethical, and scientifically sound.Our project methodology involves deploying machine learning solutions through a web application, providing personalised interventions and support for individuals with autism spectrum disorder. This enhances accessibility, scalability, and usability, empowering stakeholders.

5.2 Algorithms and flowcharts for the respective modules developed

Figure 5.2.1 Algorithm

5.3 Datasets source and utilisation

The merged dataset utilised for this project combines two separate datasets obtained from Kaggle: "Autism Screening in Adults" and "Autism Screening in Toddlers." These datasets were merged to form a comprehensive dataset encompassing individuals across different age groups for autism spectrum disorder (ASD) screening analysis.

The original datasets were sourced from Kaggle and contained demographic information, behavioural traits, medical history, and screening results related to ASD diagnosis. Due to the inherent bias in the original datasets, Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the class distribution and ensure equal representation of ASD-positive and ASD-negative cases.

The merged dataset includes features such as "Age," "Gender," "Family History," "Behavioural Traits," "Medical History," and "Screening Results," providing a comprehensive overview of individuals' characteristics and screening outcomes for ASD.

The utilisation of SMOTE technique ensures that the machine learning models trained on this dataset are less prone to bias and can effectively generalise to unseen data, enhancing the reliability and accuracy of ASD screening predictions.

Column content in data:

1. The first column contains a variable A1_Score that indicates prenatal exposure to air pollution or certain pesticides.
2. The second column contains a variable A2_Score that indicates maternal obesity, diabetes, or immune system disorders.
3. The third column contains a variable A3_Score that indicates certain genetic conditions, such as Down, fragile X, and Rett syndromes.
4. The fourth column contains a variable A4_Score that indicates older parents.
5. The fifth column contains a variable A5_Score that indicates very low birth weight.
6. The sixth column contains a variable A6_Score that indicates a sibling with autism.
7. The seventh column contains a variable A7_Score that indicates advanced parental age at time of conception.
8. The eighth column contains a variable A8_Score that indicates behavioural changes.
9. The ninth column contains a variable A9_Score that indicates any birth difficulty leading to periods of oxygen deprivation to the baby's brain.
10. The tenth column contains a variable A10_Score that indicates Extreme prematurity or very low birth weight.
11. The eleventh column contains age.
12. The twelfth column contains gender.
13. The thirteenth column contains Class/ASD that indicates whether the person has ASD or not.

Chapter 6: Results and Discussion

6.1. Screenshots of the User Interface (UI) for the respective module

Fig 6.1.1

Fig 6.1.2

--select--

- **Do you experience any behavioral changes?:**
  --select--
- **Did your mother suffer from any birth difficulties such as oxygen deprivation during conception?:**
  --select--
- **Was your birth premature?:**
  --select--
- **Do you face any verbal or nonverbal communication difficulties?:**
  --select--
- **Age:**

- **Gender:**
  --select--
- **Model:**
  --select Model--

Analyze

Fig 6.1.3

Autism Research Project

ASD Detection

Enter the scores (A1-A10), age, gender, and model information to analyze.

- **Did you face any prenatal exposure to industry chemicals, pesticides, or suffer through direct contact of air pollution?:**
  Yes
- **Did you or your parent face problems such as obesity, diabetes, or immune system disorder?:**
  No
- **Do you have any existing genetic hereditary conditions?:**
  Yes
- **Were your parents elderly when you were born?:**
  No
- **Did you have a very low birth weight?:**
  No
- **Did your mother suffer from any chronic or pulmonary disease during conception?:**
  Yes

Fig 6.1.4

Yes

- **Do you experience any behavioral changes?:**
  Yes
- **Did your mother suffer from any birth difficulties such as oxygen deprivation during conception?:**
  No
- **Was your birth premature?:**
  Yes
- **Do you face any verbal or nonverbal communication difficulties?:**
  No
- **Age:**
  23
- **Gender:**
  Female
- **Model:**
  KNN

Analyze

Slight chances of Autism.

Fig 6.1.5

30

6.2. Performance Evaluation measures

As the dataset was biassed we have performed smoting technique to decrease the biases in our dataset and have trained 4 different models in it i.e. Random forest, KNN, Decision tree, Logistic Regression and have gained an exponential level of accuracy for each of these algorithms which are listed below and have also displayed various evaluation metrics for each of the algorithms individually.

TABLE 6.2.1 Accuracy Of The Algorithms

ALGORITHMS	ACCURACY
RANDOM FOREST	0.98586
KNN	0.87632
DECISION TREE	0.95759
LOGISTIC REGRESSION	1.0

The Random Forest model shows high precision, recall, and F1-score for both classes, indicating excellent performance in classification. It achieves high accuracy with minimal misclassifications.

TABLE 6.2.2. Evaluation Metrics of Random Forest Classifier

```
Random Forest:
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.99      0.99       150
           1       0.99      0.98      0.98       133

    accuracy                           0.99       283
   macro avg       0.99      0.99      0.99       283
weighted avg       0.99      0.99      0.99       283
```

The KNN model exhibits good precision, recall, and F1-score for class 1 (positive), but relatively lower performance for class 0 (negative). It achieves moderate accuracy, with higher precision for class 1 but lower recall for class 0.

TABLE 6.2.3. Evaluation Metrics of KNN

```
KNN:
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.79      0.87       150
           1       0.80      0.98      0.88       133

    accuracy                           0.88       283
   macro avg       0.89      0.88      0.88       283
weighted avg       0.89      0.88      0.88       283
```

The Decision Tree model demonstrates high precision, recall, and F1-score for both classes, similar to Random Forest. It achieves high accuracy with good performance in classification tasks.

TABLE 6.2.4. Evaluation Metrics of Decision Tree

```
Decision Tree:
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.97      0.96       150
           1       0.96      0.95      0.95       133

    accuracy                           0.96       283
   macro avg       0.96      0.96      0.96       283
weighted avg       0.96      0.96      0.96       283
```

The Logistic Regression model achieves perfect precision, recall, and F1-score for both classes, indicating flawless performance in classification. It attains the highest accuracy among the evaluated models.

TABLE 6.2.5. Evaluation Metrics of Logistic Regression

```
Logistic Regression:
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       150
           1       1.00      1.00      1.00       133

    accuracy                           1.00       283
   macro avg       1.00      1.00      1.00       283
weighted avg       1.00      1.00      1.00       283
```

6.3. Input Parameters / Features Considered

Various input parameters or features have been considered to analyse and identify potential risk factors associated with autism spectrum disorder (ASD). These parameters are crucial for understanding the complex interplay of genetic, environmental, and developmental factors contributing to ASD risk. Here's a table describing each and every attribute and its importance:

Factors	Significance	Description
Prenatal exposure to air pollution or certain pesticides	Low	Potential environmental influence during prenatal development.
Maternal obesity, diabetes, or immune system disorders	Medium	Maternal health conditions that could affect foetal development.
Certain genetic conditions, such	Medium	Genetic factors predisposing to

as Down, fragile X, and Rett syndromes		developmental disorders.
Older parents	High	Advanced parental age at conception, correlating with increased risk.
Very low birth weight	High	Neonatal health indicator potentially affecting neurological development.
A sibling with autism	High	Familial predisposition suggesting genetic or environmental factors.
Advanced parental age at time of conception	Medium	Increased parental age possibly influencing autism risk.
Behavioural changes	Low	Minor behavioural alterations possibly indicating developmental concerns.
Any birth difficulty leading to periods of oxygen deprivation to the baby's brain	High	Birth complications affecting neurological health.
Extreme prematurity or very low birth weight	Medium	Neonatal health indicators linked to developmental risks.
Age	Low	Age is considered to have a minor effect
Gender	None	Gender is deemed to have no significant impact on the risk assessment.

Table 6.3.1 Describing each attribute
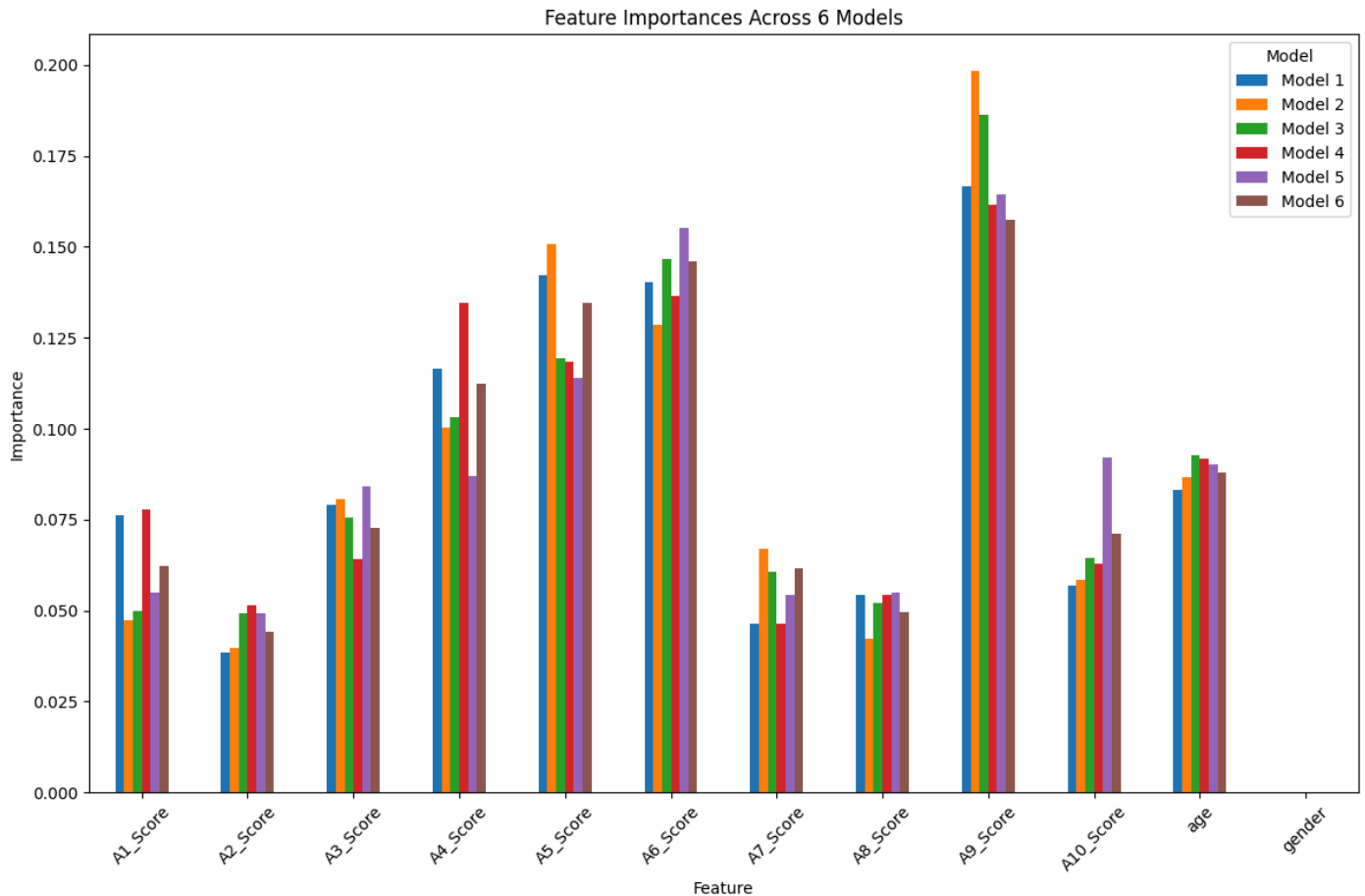
6.4. Graphical and statistical output

Figure. 6.4.1 Scores importance as

Statistical result about how each attribute is affecting the output in various models trained on the dataset.
Where A9 is having more significance followed by A6 , A5, A4, Age, A3, A10, A1, A10 whereas gender in not at all affecting the decision in of the models trained on it.
While statistically calculating the weights of every attribute we are getting the result displayed as below :

1. A1_Score and the target variable is 0.29, indicating a weak positive correlation.
2. A2_Score has a correlation of 0.29 with the target variable, suggesting a weak positive relationship.
3. A3_Score shows a correlation of 0.46 with the target variable, indicating a moderate positive correlation.
4. A4_Score exhibits a correlation of 0.51 with the target variable, signifying a moderate to strong positive relationship.
5. A5_Score has a correlation of 0.52 with the target variable, suggesting a moderate to strong positive correlation.
6. A6_Score demonstrates the highest correlation of 0.57 with the target variable, indicating a strong positive correlation.
7. A7_Score shows a correlation of 0.35 with the target variable, suggesting a moderate positive relationship.
8. A8_Score has a correlation of 0.28 with the target variable, indicating a weak positive correlation.
9. A9_Score exhibits the second-highest correlation of 0.60 with the target variable, indicating a strong positive correlation.
10. A10_Score shows a correlation of 0.42 with the target variable, suggesting a moderate positive relationship.
11. Age has a correlation of 0.10 with the target variable, indicating a very weak positive correlation.
12. Gender has a correlation of NaN (not a number), meaning there is no correlation between gender and the target variable.

6.5 Comparison of results with existing systems

Table: 6.5.1 Comparison of existing systems

Aspect	Existing System	Proposed Solution
Inability to detect milder cases	Struggles due to variations in symptom severity.	Algorithm leveraging nuanced features and machine learning for effective detection of subtle autism manifestations.
High false positive rates	Exhibits high rates, causing stress and incorrect identifications.	Stricter validation criteria and refined algorithms for precision, minimising false positives in ASD diagnosis.
High cost	Development along with deployment of costs limit accessibility.	Introduce a cost-effective alternative, reducing financial barriers to access ASD detection systems.
Over Reliance on behavioural data	Heavily depends on ambiguous behavioural cues.	Include additional symptoms to complement behavioral data, enhancing accuracy in ASD detection.

6.6. Inference drawn

The "Machine Learning Solutions for Autism Spectrum Disorder Characterization" study probably makes a number of important deductions. First of all, it shows that machine learning algorithms are capable of reliably identifying ASD, maybe surpassing more conventional diagnostic techniques. Furthermore, the research probably pinpoints particular traits or biomarkers that are essential for forecasting ASD, assisting in the earlier and more accurate diagnosis of the condition. Early identification has the potential to improve long-term results for individuals affected by enabling prompt intervention and individualised treatment programmes catered to each patient's requirements. Furthermore, the results probably point to useful ramifications for medical practitioners, opening the door to improved clinical practice in the diagnosis and treatment of ASD. But the initiative also probably brings up moral questions about algorithmic biases, data privacy, and decision-making openness.In conclusion, this suggests potential avenues for future study, such as enhancing machine learning algorithms, incorporating other data modalities, and conducting longitudinal studies to evaluate the efficacy of interventions and their long-term consequences.

Chapter 7: Conclusion

7.1 Limitations

- **Limited Representation of ASD Symptoms:**Due to predefined questionnaire constraints, ML models may overlook certain ASD symptoms, potentially leading to incomplete assessments and misdiagnoses, hindering effective intervention strategies and support provision for individuals on the autism spectrum.
- **Variability in Response Accuracy:**Variations in individual interpretation and situational factors can introduce inconsistency in questionnaire responses, posing challenges for ML models reliant on standardised data interpretation and potentially impacting the reliability of ASD predictions.
- **Risk of Overfitting:**ML models trained on limited questionnaire data may overfit to specific patterns, compromising their ability to generalise to new cases and potentially leading to inaccurate ASD predictions, necessitating careful model validation and regularisation techniques to mitigate this risk.
- **Ethical and Privacy Considerations:**Upholding privacy standards and ensuring fairness in model predictions are paramount in ASD diagnosis, requiring robust data protection measures and bias mitigation strategies to safeguard individuals' rights and promote equitable access to accurate diagnostic tools and support services.

7.2 Conclusion

Despite extensive research using ASD datasets, there is still much room for improvement in ASD prediction. To examine the key features of ASD, we collected data on early diagnosis of ASD at different stages of life, including infancy, childhood, adolescence, and adulthood. We then analyse the results using different classification methods.

In a randomly distributed sample of test results, we find 100% of the results; this is the best guess for any real test; However, we determined the average value to compare with previous Comparison studies.

7.3 Future Scope

Machine learning algorithms are being developed to improve the accuracy and reliability of ASD diagnosis, incorporating diverse data sources. This technology could also be used for personalised treatment plans. Collaborations are being made to create user-friendly diagnostic tools. Ethical considerations are being addressed, and global accessibility is being ensured. The field is also encouraged for ongoing innovation to improve diagnosis, intervention, and support for ASD patients.

References

1. U. Frith and F. Happé, "Autism spectrum disorder," Current biology, vol. 15, no. 19, pp. R786–R790, 2005.
   https://pubmed.ncbi.nlm.nih.gov/16213805/

2. WHO, Autism spectrum disorders, 2017 [Accessed August 22, 2018].
   https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders

3. Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi, "A Machine Learning Approach to Predict Autism Spectrum Disorder",IEEE2019
   https://ieeexplore.ieee.org/document/8679454

4. Sushama RAni Dutta, Sujoy Datta, Monideepa Roy, "Using Cogency and Machine Learning for Autism Detection from a Preliminary Symptom", IEEE 2019
   https://ieeexplore.ieee.org/document/8776993

5. Muhammad Shoaib Farooq , RabiaTehseen , Maidah Sabir1 & Zabihullah dAta, "Detection of autism spectrum disorder (ASD) in children and adults using machine learning", Scientific reports, 2023
   https://doi.org/10.1038/s41598-023-35910-1

6. Reem Ahmed Bahatiq, Haneen Banjar, Ahmed K. Bamaga, Salma Kammoun Jarraya, "Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging", September 2022
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9554556/

7. Suman Raj, Sarafaraz Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques" , ICCIDS 2019
   https://www.sciencedirect.com/science/article/pii/S1877050920308656

8. D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S.Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion," Journal of Child Psychology and Psychiatry, vol. 57, 2016.
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4958551/

9. Jetli Chung and Jason Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges", Hindawi Applied Computational Intelligence and So Computing Volume 2022, Article ID 9970363, 19 pages
   https://www.hindawi.com/journals/acisc/2022/9970363/

10. Kaushik Vakadkar, Diya Purkayastha & Deepa Krishnan "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques",Springer, 22 July 2021
   https://pubmed.ncbi.nlm.nih.gov/34316724/

11. Tania Akter, MD. Shahriare Satu, MD. Imran Khan, Mohammad Hanif,Shahadat Uddin, "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders"11 November 2019. **https://ieeexplore.ieee.org/document/8895818**

Appendix

1. Paper I & II Details

a.   Paper published

Paper has been submitted and is under review in 2 conferences.

1. ICCNT 2024

15th ICCCNT 2024: 15th International IEEE Conference on Computing
Communication and Networking Technologies

IIT - Mandi
Mandi, India, June 18-22, 2024

Conference website	https://15icccnt.com/
Submission link	https://easychair.org/conferences/?conf=15thicccnt2024
Submission deadline	April 15, 2024

Topics: computing computer vision iot biometrics communication 5g wireless networks robotics cyber security data science photonics electronics packaging

2. ICACS 2024

ICIACS 2024: International Conference on Innovations and Advances in Cognitive
Systems

Builders Engineering College
Tiruppur, India, May 27-28, 2024

Conference website	https://iciacs.com/
Submission link	https://easychair.org/conferences/?conf=iciacs2024
Abstract registration deadline	May 15, 2024
Submission deadline	May 22, 2024

Topics: computer science artificial intelligence data analysis deep learning

b.   Certificate of publication

c.   Plagiarism report

d.   Project review sheet

i. Review 1 (10th February, 2024)

Review 1 (10th February, 2024)

Inhouse/ Industry_Innovation/Research:

Sustainable Goal: Good Health & Well-being

Project Evaluation Sheet 2023 - 24

Class: D17 A/B/C

Group No.: 50

Title of Project: Machine Learning Solution for Autism Spectrum Disorder Characterization

Group Members: Atharvaa Sawant, Divesh Chhoda, Varun Chawla

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life-long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	2	2	2	2	2	2	2	2	2	2	2	2	36

Comments: Include images from social media develp GUI.

Dr Sujata Khedkar
Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life-long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
3½	4	4	2	2	2	2	2	2	2	2	2	2	2	2	37

Comments: GUI Integration, Ensemble Methods.

Date: 10th february, 2024

Name & Signature   Reviewer 2

ii. Review 2 (9th March, 2024)

Inhouse/ Industry_Innovation/Research:

Sustainable Goal: Health Care

Project Evaluation Sheet 2023 - 24

Class: D17 A/B/C

Group No.: 50

Title of Project: Machine Learning Solutions for Autism Spectrum Disorder

Group Members: Atharva Sawant (63) D17C, Divesh Chhoda (14) D17C, Varun Chawla (09) D17A

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life-long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	2	3	3	3	3	43

Comments: Apply other SMOTE techniques & check accuracy.

Dr Sujata Khedkar
Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life-long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	2	3	3	3	3	43

Comments: Dataset balancing to be justified, Using a proper method.

Date: 9th March, 2024

Name & Signature   Reviewer 2