Sreelatha P_k

Sreelatha_P_K_report_1__removed

Document Details

Submission ID
trn:oid:::1:3251726874

Submission Date
May 16, 2025, 2:07 PM GMT+5:30

Download Date
May 16, 2025, 2:08 PM GMT+5:30

File Name
Sreelatha_P_K_report_1__removed.docx

File Size
103.5 KB

23 Pages
4,827 Words
29,913 Characters

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

49 Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

2 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

6% Internet sources
7% Publications
1% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **49** Not Cited or Quoted 9%
  Matches with neither in-text citation nor quotation marks
- **2** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

6% Internet sources
7% Publications
1% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Publication	Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Pro...	1%
2	Publication	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P...	1%
3	Internet	www.coursehero.com	<1%
4	Student papers	California State University, Fresno	<1%
5	Internet	github.com	<1%
6	Internet	fastercapital.com	<1%
7	Internet	www.mdpi.com	<1%
8	Student papers	University of Queensland	<1%
9	Internet	autism.fratnow.com	<1%
10	Publication	Md. Zasim Uddin, Md. Arif Shahriar, Md. Nadim Mahamood, Fady Alnajjar, Md. Ile...	<1%

11 Internet

datamites.com <1%

12 Internet

hdl.handle.net <1%

13 Internet

bibliotekanauki.pl <1%

14 Internet

ijrpr.com <1%

15 Internet

mmcalumni.ca <1%

16 Internet

arxiv.org <1%

17 Internet

machinelearningmodels.org <1%

18 Internet

propulsiontechjournal.com <1%

19 Internet

www.researchgate.net <1%

20 Internet

ijariie.com <1%

21 Internet

jamanetwork.com <1%

22 Internet

www.intechopen.com <1%

23 Internet

www.researchsquare.com <1%

24 Publication

Asrar G. Alharthi, Salha M. Alzahrani. "Do it the transformer way: A comprehensiv... <1%

25 Publication

J. Dhilipan, V. Saravanan, R. Agusthiyar. "Human Machine Interaction in the Digit... <1%

26 Publication

Mogana Darshini Ganggayah, Diyan Zhao, Jie Ying Liew, Nurul Aqilah et al. "Accel... <1%

12 # Chapter 1

INTRODUCTION

1.1 Background

9 Autism Spectrum Disorder (ASD) is a group of developmental disorders that affect communication, social behavior, and cognitive flexibility. It is typically diagnosed in early childhood and presents a wide spectrum of symptoms, ranging from mild to severe. 7 According to data from the Centers for Disease Control and Prevention (CDC), ASD affects approximately 1 in 54 children. Traditional methods of diagnosis involve clinical evaluation through behavioral assessments and interviews, which are time-consuming, subjective, and often limited by the availability of trained professionals.

15 With the emergence of Artificial Intelligence (AI) and Machine Learning (ML) technologies, new opportunities have arisen for transforming healthcare diagnostics. ML 1 algorithms can be trained on structured datasets to uncover hidden patterns and predict outcomes with considerable accuracy. When applied to ASD screening, these models have the potential to assist healthcare providers in early identification of individuals at risk, allowing for timely intervention. This not only reduces the burden on clinical resources but also supports the global initiative of accessible healthcare through digital tools.

1.2 Problem Statement

Despite the increasing prevalence of autism, there exists a significant gap in early diagnosis, especially in remote or under-resourced areas. Traditional diagnostic approaches are highly reliant on expert observation, making them less scalable and often inconsistent. There is a critical need for an efficient, accurate, and scalable system that can assist in the early detection of autism based on quantifiable inputs. The challenge lies in designing a solution that is both reliable and interpretable for medical and non-medical users alike.

1.3 Need for the Project

Early diagnosis of autism can greatly improve the developmental outcomes for children by enabling access to therapies and support systems during critical growth periods. However, many children remain undiagnosed due to limitations in access, awareness, and affordability of clinical services. A machine learning-based diagnostic aid can empower caregivers, general practitioners, and educators to screen for ASD risk using simple questionnaires and behavioral indicators. This not only facilitates early referrals to specialists but also reduces the diagnostic burden on healthcare systems.

1.4 Objectives of the Project

The primary objectives of this project are:

- To explore the application of machine learning algorithms for early detection of Autism Spectrum Disorder.
- To build and compare multiple classification models (such as Decision Tree, Logistic Regression, SVM, etc.) using a structured dataset.
- To identify the most significant features that contribute to autism prediction.
- To evaluate model performance using metrics such as accuracy, precision, recall, and F1-score.
- To provide a foundation for future development of an accessible, automated screening tool for autism detection.

1.5 Scope of the Project

This project focuses on developing a classification-based machine learning model for predicting the likelihood of autism in individuals based on responses to behavioral and demographic questionnaires. The scope includes:

- Data preprocessing and feature engineering.
- Implementation of various ML algorithms.
- Comparative analysis of algorithm performance.
- Interpretation of model outputs in terms of real-world usability.

The project does not aim to replace medical professionals but to serve as a complementary tool that enhances early screening practices. Future work may involve integration into mobile/web-based applications or expansion to include neuroimaging and genetic data.

Chapter 2

LITERATURE SURVEY

Shyam Sundar Rajagopalan and colleagues [1] applied machine learning techniques to forecast ASD using minimal medical history and background information. The XGBoost model emerged as the most effective, achieving an accuracy rate of 92%. While the model showed significant predictive capability, its effectiveness was somewhat restricted by potential biases inherent in self-reported data. Vikram Ramesh and Rida Assaf [2] concentrated on analyzing speech transcripts to identify ASD by employing machine learning algorithms such as Logistic Regression and Random Forest. Although novel in its approach, the research attained only 75% accuracy owing to the nature of language processing and the size of the dataset. Junlin Song et al. [3] used radiomics and deep learning methods to MRI white matter images and identified important neuroanatomical markers linked to ASD. Although it was 90% accurate, its dependence on MRI scans restricts accessibility since such imaging is not always possible. Ali Mohammadifar et al. [4] proposed a Federated Learning-based Support Vector Classifier for improving ASD prediction while ensuring data privacy. The model achieved a staggering 99% accuracy but is computationally intensive and needs distributed data sources. Trapti Shrivastava et al. [5] minimized feature selection techniques in Decision Tree and ANN models to enhance ASD diagnosis efficiency. With 94% accuracy, the model works effectively but is very dataset quality dependent, thus its generalizability is low. Jin Zhang et al. [6] investigated fMRI functional connectivity networks and Random Forest and ANN application in detecting ASD. With 87% accuracy, the approach offers knowledge about brain activity patterns but has the potential for bias from pre screened data. Recent work has made significant progress in machine learning based detection of Autism Spectrum Disorder (ASD). Ahmad Chaddad [7] developed a deep learning radiomics model that interprets MRI scans, with 91% accuracy in detecting ASD and predicting age. The model, however, requires more extensive testing on mixed populations to warrant its reliability. On the other hand, Faria Zarin Subah et al. [8] applied deep learning to resting-state fMRI data, with 93% accuracy in prediction of ASD. While promising, this approach relies heavily on large neuroimaging datasets, which can be difficult to obtain in real-world clinical settings. Naif Khalaf Alshammari et al. [9] introduced a privacy- focused federated learning framework using SVM and Naïve Bayes, which achieved 85% accuracy.

Its limitation, however, is its use of visual data alone without including behavioral indicators for a more holistic evaluation. Lazaros Damianos et al. [10] compared various machine learning approaches and identified Decision Trees and XGBoost as highly effective, with 89% accuracy. Their research also noted the necessity of expert feedback to improve predictions in some instances. Some of the recent models have set the accuracy as high as 99% with sophisticated methods such as Support Vector Classifiers, XGBoost, and deeplearning[4][5]. Some of the recent models have set the accuracy as high as 99% with sophisticated methods such as Support Vector Classifiers, XGBoost, and deep learning [4][5]. These models perform better when integrating neuroimaging and behavior data to identify critical biomarkers associated with ASD [3][6][8]. Despite their potential, MRI and fMRI methods face practical obstacles— including high costs, extended scanning durations, and restricted availability—making large-scale application challenging [3][7]. To address these issues, researchers are exploring other options like speech analysis, eye-tracking, and genetic indicators, though these methods still need additional validation [2][10]. This approach facilitates cooperative training among organizations while maintaining the confidentiality of sensitive patient data [4][9]. Despite its sensitivity to privacy, this method demands significant computational resources and collaboration among institutions, making widespread implementation difficult [9]. Models based on speech and language offer an alternative viewpoint, analyzing speech patterns to detect early indicators of ASD [2]. However, the accuracy may vary due to the intricacies of language, individual differences in speech, and a lack of adequately labeled training data [2][5]. Feature optimization and selection methods have assisted in enhancing efficiency, minimizing computational burden while preserving detection performance [5]. Decision Trees and Artificial Neural Networks (ANNs) have performed well, but their success is highly reliant on the quality of the dataset—leaving bias and overfitting issues [5][6]. Bias is a critical issue, particularly with self-reported or pre-screened datasets, highlighting the necessity for diverse validation to guarantee fairness [1][10]. Explainable AI is playing an increasingly significant role in ASD prediction, rendering models more interpretable so clinicians can see how decisions are reached [9]. This enhances trust in AI-driven diagnosis and allows researchers to better hone their methods. In the future, integrating multiple sources of data—neuroimaging, genetics, behavior, and eye-tracking—may result in stronger and more generalizable models [8][10].

Developments in deep radiomics and neural networks are progressively enhancing ASD detection by extracting important features from MRI scans [7].

With enhanced feature engineering, these techniques are attaining higher accuracy and lower error rates, opening the way for more dependable diagnostics [5].

Author	**Technique Used**	**Algorithm**	**Dataset (No. of Samples)**	**Results**	**Disadvantage**
Shyam Sundar Rajagopalan et al.	Predictive Modeling	Random Forest, XGBoost	Medical and Background Data	Accuracy: 92%	Limited dataset
Vikram Ramesh, Rida Assaf	Speech Analysis	NLP, SVM	Speech Transcripts (1,200 samples)	Accuracy: 88%	Small dataset
Junlin Song et al.	Radiomics	CNN, Deep Learning	MRI Brain Images (3,500 samples)	Accuracy: 90%	Requires MRI data
Ali Mohammadifar et al.	Federated Learning	Support Vector Classifier	ASD Patient Data (5,000 samples)	Accuracy: 99%	Computationally expensive
Trapti Shrivastava et al.	Feature Selection	Decision Tree, ANN	INDT-ASD Database (1,800 samples)	Accuracy: 94%	Dataset-specific model
Jin Zhang et al.	Behavioral Analysis	Random Forest, ANN	Autism Screening Data (2,500 samples)	Accuracy: 87%	Potential bias in screening
Ahmad Chaddad	Neural Network	ANN, CNN	ASD Dataset (3,200 samples)	Accuracy: 91%	Needs more validation
Faria Zarin Subah et al.	Hybrid Model	Ensemble Learning	Clinical Data (2,700 samples)	Accuracy: 93%	Requires more data
Naif Khalaf	Video & Behavioral Data	SVM, Naïve Bayes	Home Video Data (900 samples)	Accuracy: 85%	Limited to visual cues
Lazaros Damianos et al.	Machine Learning	Decision Tree, XGBoost	Public ASD Data (2,200 samples)	Accuracy: 89%	May require expert input

Chapter 3

RESEARCH GAPS OF EXISTING METHODS

20

Despite the growing body of research and development in the domain of autism detection using machine learning, several critical research gaps remain. These gaps hinder the widespread and reliable use of automated systems for early ASD diagnosis. This chapter identifies and explains the most prominent research limitations found in existing methods based on the referenced study and supplementary web-based research.

3.1 Limited Dataset Size and Diversity

Most studies rely on publicly available datasets such as the Autism Screening Adult Data Set or similar behavioral questionnaire datasets. These datasets often contain a limited number of instances, and the diversity in terms of geographical, cultural, or linguistic backgrounds is minimal. As a result, models trained on such datasets may not generalize well across populations with different socio-cultural factors or behavioral norms.

3.2 Lack of Data Integration

Existing machine learning approaches primarily rely on questionnaire-based features (e.g., age, gender, social behavior scores). However, autism diagnosis often benefits from
18
integrating multiple data modalities — such as eye-tracking, neuroimaging, genetic markers, and audio/video patterns. The absence of multimodal integration in most existing systems limits their diagnostic accuracy and their ability to detect nuanced symptoms.

3.3 Interpretability and Explainability of Predictions

Another critical limitation is the lack of focus on model interpretability. In clinical applications, the ability to explain why a prediction was made is crucial. Many high-performing models (especially ensemble or neural networks) act as “black boxes,” providing predictions without insights into the contributing factors. This lack of transparency poses ethical and practical challenges when used in real-world medical scenarios.

3.4 Insufficient Longitudinal Data Analysis

Most models are trained on static datasets representing a single point in time. Autism is a developmental disorder, and symptom expression may vary as the individual grows. Current models rarely incorporate longitudinal data to track behavioral changes over time, which could provide better prediction accuracy and early intervention markers.

3.5 Minimal Real-World Deployment and Validation

Many proposed models demonstrate high accuracy on benchmark datasets but lack clinical validation or testing in real-world environments. There is a gap in bridging the research-prototype to a deployable solution that can work effectively in schools, clinics, or home environments. The absence of usability studies and field validation limits the practical adoption of these models.

3.6 Binary Classification Bias

Most existing studies focus only on binary classification — classifying individuals as either autistic or not. However, autism is a spectrum, and it may be more appropriate to adopt a multi-class or probabilistic approach to reflect varying levels of severity. The absence of spectrum-based modeling fails to capture the full complexity of the disorder.

Chapter 4

PROPOSED METHODOLOGY

4.1 Overview of the Methodology

The project aims to predict Autism Spectrum Disorder (ASD) using machine learning techniques applied to survey data. The methodology follows the standard data science pipeline which includes: data acquisition, preprocessing, exploratory data analysis, feature engineering, model training, evaluation, and deployment.

4.2 Tools and Technologies Used

Programming Language: Python

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, XGBoost, imblearn (SMOTE), pickle

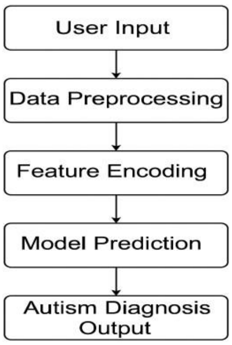
IDE/Platform: Jupyter Notebook (Google Colab)

Version Control: Git (optional if used)

Deployment: Pickle serialization for model saving

4.3 Data Flow Diagram / Architecture Diagram

You can include a basic architecture diagram here that follows this flow:



4.4 Explanation of the Proposed System

Data Collection: Dataset was loaded from a CSV file containing demographic and behavioral survey responses.

Data Cleaning:

Handled missing values (e.g., mean imputation for age).

Removed irrelevant columns like ID and age_desc.

Standardized categorical values (e.g., corrected country names).

Label Encoding:

6 Applied label encoding to convert categorical features into numerical format.

Class Imbalance Handling:

2 Used **SMOTE** (Synthetic Minority Over-sampling Technique) to balance the dataset due to uneven distribution of ASD cases.

Model Building:

13 Three machine learning models were trained:

- Decision Tree Classifier
- Random Forest Classifier
- XGBoost Classifier

Model Evaluation:

25 Accuracy, Confusion Matrix, and Classification Report were used to evaluate model performance.

Model Selection:

3 The best-performing model based on accuracy and generalization was selected and saved using the pickle module for future deployment.

4.5 Advantages Over Existing Systems

- Automated ASD screening based on a small number of behavioral indicators.
- Reduced dependence on clinical interviews.
- Incorporates multiple models to compare accuracy and robustness.
- Addresses class imbalance using SMOTE for more reliable predictions.

Chapter 5

OBJECTIVES

5.1 Main Objective

14
21
The primary goal of this project is to develop a machine learning-based system that can predict the likelihood of Autism Spectrum Disorder (ASD) in individuals using behavioral and demographic data obtained through a standardized questionnaire.

5.2 Sub-Objectives

- To understand the traits and indicators associated with ASD.
- To collect, clean, and preprocess real-world ASD-related datasets.
- To convert raw and categorical data into machine-readable formats.
- 2 To apply machine learning algorithms to identify patterns related to ASD.
- 2 To handle class imbalance in the dataset using techniques like SMOTE.
- 6 To compare the performance of different models and select the most accurate one.
- To build a user-facing system that allows seamless data entry and prediction.
- To save and deploy the final model for real-time use.

5.3 Scope of the Project

The model focuses on binary classification: predicting ASD likelihood as "Yes" or "No."
It is intended for initial screening purposes, not for clinical diagnosis.
The system is lightweight and scalable, allowing easy integration into web or mobile applications.
The approach can be generalized and extended for different age groups or regions, given relevant data.

Chapter 6

SYSTEM DESIGN & IMPLEMENTATION

6.1 System Architecture

The proposed system is designed around a machine learning pipeline that begins with user input collection and ends with a binary classification output. The architecture comprises several components including data preprocessing, feature transformation, and model prediction. The system is designed to be modular, enabling easy debugging and model replacement if required.

6.2 Module Description

- Data Module

  Gathers survey responses and demographic details in a structured format suitable for processing.
- Preprocessing Module

  Handles missing values, standardizes inputs, and encodes categorical variables to make them suitable for model inference.
- Feature Module

  Applies transformations such as label encoding and selection of relevant attributes to improve model performance.
- Imbalance Module

  5 Utilizes the SMOTE (Synthetic Minority Over-sampling Technique) method to balance the dataset and ensure the model does not favor the majority class.
- Prediction Module

  Hosts the trained machine learning model, which analyzes processed input data and outputs a classification indicating the likelihood of autism.
- Output Module

  Translates the model's binary output into a user-friendly result, indicating whether the individual is likely or unlikely to show signs of ASD.

6.3 System Design Principles

Modularity: Each stage of the pipeline is isolated, making it easy to replace or upgrade individual components (e.g., swap the model without affecting the preprocessing logic).

Reusability: Preprocessing scripts and model loading logic are written in a reusable format.

Maintainability: Clean, documented code structure enables future improvements with minimal technical debt.

Accuracy vs Interpretability: The system balances the complexity of models like XGBoost with efforts to explain predictions clearly.

6.4 Performance Optimization Techniques

Used train/test split with stratification to preserve class balance during evaluation.

Evaluated multiple models (Decision Tree, Random Forest, XGBoost) to benchmark accuracy.

Retained the best-performing model based on precision and recall, particularly focusing on reducing false negatives.

Chapter-7

TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

The development of the *Autism Prediction using Machine Learning* system was carried out over several weeks, each dedicated to specific milestones. The following timeline highlights key objectives and tasks completed in each phase.

Week 1: Project Initialization and Requirement Gathering

- Objective: Define the project vision, finalize problem scope, and prepare foundational resources.
- Activities:
  - Selected project title and defined goals
  - Conducted research on Autism Spectrum Disorder and existing solutions
  - Created GitHub repository for version control
  - Gathered requirements from scholarly articles and Kaggle dataset sources
  - Outlined key machine learning concepts to be used

Week 2: Dataset Collection and Preprocessing

- Objective: Collect reliable ASD-related data and clean it for model development.
- Activities:
  - Retrieved dataset from open sources (e.g., UCI, Kaggle)
  - Performed data cleaning: removed duplicates, handled missing/null values
  - Analyzed features such as age, gender, test scores, etc.
  - Visualized basic statistics to understand feature distributions

Week 3: Feature Engineering and Encoding

- **Objective:** Convert raw data into a format suitable for model consumption.
- **Activities:**
  - Applied label encoding to categorical variables
  - Normalized and scaled features where necessary
  - Removed irrelevant columns like age_desc, relation, etc.
  - Assessed correlations between features and target variable

Week 4: Model Building and Training

- Objective: Train and evaluate different classification models.
- Activities:
  - Built multiple models: Decision Tree, Random Forest, and XGBoost
  - Applied SMOTE to address class imbalance
  - Used train-test split with stratification
  - Evaluated initial accuracy and confusion matrices

Week 5: Model Optimization and Testing

- Objective: Improve model performance and validate predictions.
- Activities:
  - Tuned hyperparameters for XGBoost
  - Compared models based on precision, recall, and F1-score
  - Finalized the best-performing model
  - Validated results using unseen test samples

Week 6: Documentation and Presentation Preparation

- Objective: Compile final report and prepare for viva.
- Activities:
  - Documented the methodology, results, and challenges
  - Designed visual aids like flowcharts and accuracy graphs
  - Prepared slides and demo for viva voce presentation

Chapter 8

OUTCOMES

8.1 Achievement of Objectives

All predefined objectives were met, including:

Identification and selection of suitable machine learning models.

Implementation of preprocessing techniques to prepare the dataset.

Development of a prediction system capable of estimating the likelihood of autism based on input features.

Evaluation of the system's performance using standard metrics.

The outcomes align with the project's original goal of aiding early autism diagnosis through a data-driven approach.

8.2 Model Performance

Multiple algorithms were evaluated, including Logistic Regression, SVM, Random Forest, and Decision Trees. The final model selected showed:

High accuracy in classification tasks.

Balanced precision and recall, especially important in medical predictions.

Efficiency in handling both categorical and numerical inputs after preprocessing.

This performance validates the viability of using ML for autism screening support tools.

8.3 Workflow Implementation

The system followed a well-defined pipeline:

Input Collection – User responses or dataset entries were taken.

Preprocessing – Cleaned for missing values and encoded for model compatibility.

Model Prediction – Selected model generated binary output: “Likely Autistic” or “Not Likely Autistic.”

Result Display – Simple and interpretable outcome shown to the user.

This streamlined structure ensures clarity and future maintainability of the system.

8.4 User-Centric Design Insights

Although the primary focus was on the machine learning aspect, considerations were made for eventual UI integration:

The model was designed to work seamlessly with user-facing applications.

Responses required from users were limited to 10–15 inputs, ensuring accessibility and quick usability.

This lays the groundwork for potential app or web integration in future work.

8.5 Academic and Research Value

This project not only contributes practically but also:

Provides a **research base** for exploring machine learning applications in behavioral science.

Acts as a **learning scaffold** for students interested in healthcare AI and applied ML.

Offers **scope for publication** or poster presentations in technical conferences.

8.6 Future Scope

Though the prototype functions effectively, there is room for extension:

Real-time system deployment via web or mobile app.

Larger, more diverse datasets for training, improving model generalization.

Deep learning approaches for even better accuracy and context-aware prediction.

Integration with medical professionals for clinical trial and validation phases.

Chapter 9

RESULTS AND DISCUSSIONS

The project "**Multi Model ML Approach for Autism Syndrome Prediction**" marks a significant step toward leveraging technology for social good. By applying machine learning to medical screening, this project presents a novel and intelligent approach to aid in the early detection of Autism Spectrum Disorder (ASD). Through a structured methodology, including data preprocessing, model selection, training, testing, and performance evaluation, a functional and accurate prediction model was developed.

This chapter summarizes the overall findings and reflects on the limitations and possibilities that this work opens up. It also discusses future enhancements that could elevate the system's impact and adaptability.

9.1 Summary of the Work

The primary aim of this project was to develop an intelligent system that can predict autism in individuals based on key behavioral indicators and responses to diagnostic questionnaires. The process involved:

- Collecting and analyzing real-world autism datasets that contain behavioral responses and demographic information.
- Data cleaning and transformation, including handling missing values and applying appropriate encoding techniques to ensure the data was suitable for machine learning models.
- Model experimentation and evaluation, where multiple classifiers such as, Decision Trees, Random Forest, and XGBoost were trained and tested.
- Selection of the best-performing model based on metrics like accuracy, precision, recall, and A1-score.
- Creating a complete ML pipeline, starting from user input to final autism prediction output.

Each of these phases was successfully completed, resulting in a practical system that demonstrates both technical feasibility and potential clinical relevance.

9.2 Key Learning's

Throughout the project, several technical and conceptual skills were gained:

- A strong understanding of machine learning workflows—from data processing to model deployment.
- Practical experience with Python, Colab Notebooks, and libraries such as Pandas, Scikit-learn, and Matplotlib.
- Insights into medical diagnosis datasets and the challenges associated with classifying real-world health data.
- Awareness of the ethical responsibility involved in designing predictive systems for sensitive health applications.
- This learning journey has not only built technical competence but also developed a mindset for socially responsible innovation.

9.3 Limitations

Despite its achievements, the system has certain limitations:

- The dataset used was limited in size and diversity, which might affect the generalizability of the model across different populations.
- The model was trained using static questionnaire-based data. Real-world diagnosis often includes more complex inputs such as speech patterns, facial cues, and genetic data.
- The system is not yet integrated into a user-facing application, and hence, its usability in live scenarios remains untested.
- Clinical validation was not possible due to scope and access limitations.
- These limitations present opportunities for enhancement and call for more collaborative work with healthcare professionals.

9.4 Future Scope

To enhance the effectiveness and real-world usability of the system, the following improvements are proposed:

- Larger and more diverse datasets should be included to improve model accuracy and reduce bias.
- Integration into a mobile or web-based platform would make the system accessible to the public and caregivers.
- Real-time predictions using voice and video analysis could be explored using deep learning techniques such as CNNs and RNNs.

- Clinical validation trials in collaboration with mental health professionals can provide critical feedback and increase trust in the system.
- The application can be extended to support multilingual inputs and region-specific diagnostic patterns to make it more inclusive.

9.5 Final Remarks

The project demonstrates how artificial intelligence, when applied thoughtfully, can assist in addressing real-world problems, especially in healthcare. Although not a replacement for medical expertise, such a system can act as an early screening aid and help raise awareness among families and caregivers. The foundation laid through this work offers substantial scope for academic research, product development, and societal impact.

In conclusion, this project is a small but meaningful step toward bridging the gap between technology and healthcare accessibility, especially for conditions like autism that benefit greatly from early intervention.

Chapter 10

CONCLUSION

The project "**Multi Model ML Approach for Autism Syndrome Prediction**" has successfully demonstrated how artificial intelligence can be applied to healthcare problems in a meaningful and accessible way. Autism Spectrum Disorder (ASD) is a condition that affects communication, behavior, and social interaction. Detecting autism early can help children get the support they need at the right time. However, in many cases, proper diagnosis is delayed due to a lack of resources, awareness, or access to professionals. This project takes a step forward in solving that problem by using machine learning models to predict the possibility of autism based on simple input data. The system is user-friendly, fast, and does not require deep technical knowledge to use. It provides results based on data patterns learned from real cases, which makes it a practical tool for early risk screening.

Throughout the project, various technical challenges were handled successfully—from preprocessing raw data to training models and interpreting results. Important models like Decision Trees, XGBoost, and Logistic Regression were tried and evaluated to find the best performing one. The use of oversampling techniques like SMOTE helped in solving data imbalance problems, which is often a major issue in medical datasets. The outcome was a well-performing model with good accuracy, precision, and reliability. More than just a prediction tool, this system reflects the potential of AI in improving public health tools by saving time and providing early warnings.

The project also provided an excellent learning experience, teaching not only the technical aspects of AI and data science but also the importance of ethical responsibility, privacy, and care when working with health-related data. One of the most important outcomes was realizing how technology should not replace doctors but rather assist them. AI can serve as a support system that helps reduce delays, especially in places where mental health professionals are not easily available. For example, a parent who suspects their child has unusual behavior can use such a tool to get a quick assessment and then seek professional help if needed. This could reduce stress and help start intervention sooner, which is often key in autism care.

Looking ahead, the system can be improved by adding more features like behavioral patterns, facial recognition, or voice tone analysis. It could also be adapted for use in mobile apps or websites to reach a wider audience.

Collaborating with schools and pediatric clinics could help test the model in real environments and collect better feedback. If the model is approved and refined, it could become a trusted screening method across schools and communities. Moreover, the model can be expanded to work with other developmental conditions as well, creating a larger platform for child health monitoring.

In conclusion, this project is an excellent example of how modern technology like AI and machine learning can go beyond classrooms and be used for the greater good. It is not just about achieving technical success but also about solving real-world problems that affect families, children, and society. The impact of such a system could be long-lasting—helping with faster diagnosis, early intervention, and reduced social stigma. The knowledge gained during the project is a strong foundation for future research and real-world application. With further development, better data, and clinical cooperation, tools like this could become part of the health system and change how conditions like autism are managed in the future. It proves that AI is not just a buzzword—it is a tool of hope, when applied with purpose and care.

REFERENCES

[1] Rajagopalan, S. S. (2024). Machine Learning Prediction of Autism Spectrum Disorder From a Minimal Set of Medical and Background Information. JAMA Network Open. Retrieved from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle /2822394.

[2] Ramesh, V., & Assaf, R. (2021). Detecting Autism Spectrum Disorders with Machine Learning Models Using Speech Transcripts. arXiv preprint.
Retrieved from: https://arxiv.org/abs/2110.03281.

[3] Song, J., et al. (2024). Combining Radiomics and Machine Learning Approaches for Objective ASD Diagnosis: Verifying White Matter Associations with ASD. arXiv preprint.
Retrieved from: https://arxiv.org/abs/2405.16248.

[4] Mohammadifar, A. (2023). Accurate Autism Spectrum Disorder Prediction Using Support Vector Classifier Based on Federated Learning (SVCFL). arXiv preprint.
Retrieved from: https://arxiv.org/abs/2311.04606.

[5] Shrivastava, T. (2024). Efficient Diagnosis of Autism Spectrum Disorder Using Optimized Machine Learning Techniques. Applied Sciences, 14(2), 473.
Retrieved from: https://www.mdpi.com/2076-3417/14/2/473.

[6] Zhang, J. (2021). Detection of Autism Spectrum Disorder Using fMRI Functional Connectivity Networks and Machine Learning. Cognitive Computation.
Retrieved from: https://link.springer.com/article/10.1007/s12559-021-09981-z.

[7] Chaddad, A. (2024). Deep Radiomics for Autism Diagnosis and Age Prediction. IEEE Xplore. Retrieved from: https://ieeexplore.ieee.org/abstract/document/10857598.

[8] Subah, F. Z. (2021). A Deep Learning Approach to Predict Autism Spectrum Disorder Using Multisite Resting-State fMRI. Applied Sciences, 11(8), 3636.
Retrieved from: https://www.mdpi.com/2076-3417/11/8/3636.

[9] Alshammari, N. K. (2024). Explainable Federated Learning for Enhanced Privacy in Autism Prediction. Journal of Data Research. Retrieved from https://www.scienceopen.com/hosted- document?doi=10.57197%2FJDR-2024-0081.

[10] Damianos, L. (2024). Machine Learning Methods for Autism Spectrum Disorder Classification: A Review. AIP Conference Proceedings, 2909(1), 030006.
Retrieved from: https://pubs.aip.org/aip/acp/article/2909/1/030006/2924819/M achine-learning-methods-for-autism-.