



Intelligent Detection Of Autism Spectrum Disorder Using Machine Learning Algorithms

Integrating Machine Learning and Deep Learning Models for Early Detection and Analysis of Autism Spectrum Disorder(ASD)

¹Bheesetty Venkata Kiranmayi, ²K.Tulasi Krishna Kumar

¹Student, ²Training and Placement officer

¹Department of Computer Science and Engineering, ²Department of CS & SE

Department Of Computer Science and Engineering , Sanketika vidya parishad engineering college,
Visakhapatnam, India

Abstract: Autism Spectrum Disorder (ASD) diagnosis has experienced a remarkable evolution with the integration of Artificial Intelligence (AI) and data-driven healthcare analytics. This paper presents an AI-based diagnostic framework that leverages behavioral data, demographic attributes, and machine learning algorithms to predict the likelihood of ASD in children, adolescents, and adults. The research highlights the methodology, experimental design, and comparative analysis of multiple models including Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Convolutional Neural Network (CNN). The proposed system provides accurate and early detection of ASD symptoms using non-clinical datasets from the UCI repository, ensuring reliable predictions for healthcare practitioners. The framework integrates intelligent data preprocessing and deep learning layers to enhance classification accuracy and minimize diagnostic errors. A cloud-based implementation and interactive dashboard improve accessibility, scalability, and usability for researchers, caregivers, and medical professionals across diverse regions.

Index Terms – Autism Spectrum Disorder (ASD), Machine Learning, Data Analysis, Predictive Modeling, Healthcare Analytics, Early Diagnosis, Neurodevelopmental Disorders.

1. INTRODUCTION

The application of Machine Learning (ML) in the healthcare domain has significantly advanced the diagnosis and analysis of complex disorders. Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by challenges in social interaction, communication, and behavioral patterns. This research project focuses on developing a Machine Learning-based system for the analysis and detection of ASD using patient datasets. The proposed model employs data preprocessing, feature extraction, and classification algorithms to predict ASD likelihood accurately. By leveraging supervised learning techniques and medical data analytics, the system aims to assist healthcare professionals in early detection, improving diagnosis accuracy, and supporting timely interventions.

1.1 Research Objectives

- Develop a Machine Learning-based system for the detection of Autism Spectrum Disorder (ASD).
- To analyze medical and behavioral data to identify key features contributing to ASD diagnosis.
- To compare different machine learning algorithms for accurate ASD prediction.
- To design a decision-support tool that assists healthcare professionals in early detection and intervention.

1.2 Research Hypothesis

- H1: Machine Learning algorithms can effectively classify individuals with and without Autism Spectrum Disorder.
- H2: Feature selection and data preprocessing significantly enhance the accuracy of ASD detection models.
- H3: Early diagnosis supported by ML-based prediction systems can improve treatment outcomes and patient support strategies.

2. ABBREVIATIONS AND ACRONYMS

ASD – Autism Spectrum Disorder

ML – Machine Learning

AI – Artificial Intelligence

SVM – Support Vector Machine

ANN – Artificial Neural Network

CNN – Convolutional Neural Network

3. LITERATURE REVIEW

Machine Learning-based diagnostic systems have transformed the healthcare industry by enabling accurate, data-driven detection of complex disorders such as Autism Spectrum Disorder (ASD). Researchers have explored various ML algorithms, including Support Vector Machines (SVM), Random Forests, and Neural Networks, for classifying ASD based on behavioral and biological data. Thabtah (2017) developed a machine learning framework for autism screening using behavioral datasets, demonstrating improved diagnostic accuracy. Duda et al. (2016) emphasized the potential of data-driven models in identifying ASD traits from clinical assessments. Studies by Bone et al. (2015) and Wall et al. (2012) highlight the role of computational techniques in reducing diagnostic time and supporting early intervention. Building on these findings, the proposed research focuses on developing a robust ML-based model for efficient and accurate ASD detection.

3.1 Early Techniques:

- Traditional diagnostic methods for Autism Spectrum Disorder (ASD) relied on clinical observations and questionnaire-based assessments.
- These manual approaches were time-consuming, subjective, and dependent on expert evaluation.
- The lack of automated tools limited early detection accuracy and scalability in large populations.

3.1 Shift to Machine Learning::

- Healthcare diagnostics have shifted from manual assessments to Machine Learning-based prediction models.
- ML algorithms enable automated analysis of medical and behavioral data, improving diagnostic accuracy and efficiency.
- Integration of ML techniques supports early detection, reducing human error and enhancing clinical decision-making.

3.2 Evolution of Data-Driven Diagnosis:

- Early ASD diagnosis methods relied on behavioral checklists and clinical interviews without computational assistance.
- These methods often lacked consistency, leading to delayed or inaccurate diagnoses.
- The introduction of digital health datasets enabled quantitative analysis of ASD traits for more objective evaluation.

3.3 Machine Learning–Based Detection Models:

- Advanced ML models such as SVM, Random Forest, and Neural Networks have improved the accuracy of ASD prediction.
- These models analyze multiple features including social behavior, communication patterns, and medical indicators.
- Proven effective in supporting clinicians through data-driven diagnostic recommendations.

3.4 Proposed System’s Role:

- Combines machine learning algorithms with feature selection techniques for efficient ASD detection.
- Utilizes preprocessed behavioral and medical datasets to train and evaluate prediction models.
- Designed as a user-friendly, automated diagnostic support system that enhances early identification and clinical decision-making.

4. METHODOLOGY

The proposed ASD Detection System is designed as a modular pipeline integrating Machine Learning algorithms with data analytics techniques to enable accurate and efficient diagnosis. It includes three key components: data preprocessing, feature extraction and selection, and ASD classification. The system processes medical and behavioral datasets to identify meaningful patterns associated with autism. Using supervised learning techniques, it classifies individuals based on predictive features, ensuring reliable, fast, and interpretable diagnostic results.

4.1 Research Methods

The system employs a multi-model Machine Learning architecture:

- Data preprocessing techniques for handling missing values, normalization, and noise reduction to ensure dataset quality.
- Feature extraction and selection using statistical and algorithmic methods such as Recursive Feature Elimination (RFE) to identify key diagnostic traits.
- Classification models including Support Vector Machine (SVM), Random Forest (RF), and Neural Network (ANN) for predicting ASD likelihood.

All modules were implemented independently using Python-based ML frameworks such as Scikit-learn and TensorFlow, trained on publicly available ASD datasets. This modularity allows easy optimization, algorithm comparison, and seamless integration into a unified diagnostic support platform.

Data Collection Procedures

A specialized ASD dataset was utilized to ensure reliable and representative data for model training and testing. Since public medical datasets are limited, data were collected and curated from open-access autism screening repositories and clinical survey sources.

Data preparation included:

- **Data Collection:** Extracted from publicly available ASD datasets such as the Autism Screening Adult Dataset and Toddler Dataset from the UCI Machine Learning Repository.
- **Annotation:** Each record labeled according to diagnostic outcomes (ASD / Non-ASD) based on standardized screening results.
- **Feature Categories:**
 - **Demographic:** Age, Gender, Ethnicity, Country of Residence
 - **Behavioral:** Social interaction, Communication ability, Repetitive behaviors
 - **Screening Scores:** Questionnaire responses and diagnostic test scores

- **Dataset Size:**
 - Adult Dataset: ~700 records
 - Toddler Dataset: ~300 records
- **Augmentations:**
 - Handling missing values through imputation
 - Normalizing continuous features and encoding categorical variables
 - Removing redundant and noisy entries to improve model performance

These preprocessing and validation steps ensured a clean, well-structured dataset suitable for training machine learning models capable of accurate ASD detection and analysis.

4.2 Analysis Techniques

Model architectures and training configurations were optimized to achieve high diagnostic accuracy and reliability.

Model Summary:

- **Feature Selection Module:** Identifies the most relevant behavioral and demographic attributes influencing ASD prediction using Recursive Feature Elimination (RFE) and correlation analysis.
- **Classification Models:**
 - **Support Vector Machine (SVM)** – separates ASD and non-ASD cases using optimized kernel functions.
 - **Random Forest (RF)** – performs ensemble-based classification for robust and interpretable predictions.
 - **Artificial Neural Network (ANN)** – captures complex nonlinear relationships among behavioral traits.

Training Settings:

- **Hyperparameters Tuned:** Learning rate, kernel type, number of estimators, and activation functions (grid search optimization).
- **Iterations:** 10–30 tuning runs | **Objective:** maximize classification accuracy and minimize false negatives.
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Confusion Matrix analysis.

These configurations enabled the system to perform precise, data-driven detection of Autism Spectrum Disorder, ensuring consistency and improved predictive performance across multiple models.

4.3 Ethical Considerations:

The project addresses ethical concerns to ensure safe, responsible, and fair use::

- **Data Privacy:** Patient identifiers and sensitive medical information are anonymized to protect confidentiality.
- **Bias Control:** Datasets are balanced across age groups, genders, and behavioral traits to prevent model bias.
- **Transparency:** Feature selection methods, algorithm logic, and model performance metrics are documented for reproducibility and interpretability.
- **Human Oversight:** The system provides diagnostic support; final clinical decisions remain with healthcare professionals.

Such practices help promote trust, accountability, and ethical deployment of AI-driven ASD detection systems in real-world healthcare settings.

5.RESULTS AND DISCUSSIONS

The ASD Detection System was evaluated on publicly available and curated datasets to assess classification accuracy, reliability, and performance. This section summarizes model behavior during prediction, presents evaluation metrics, includes sample algorithm outputs, and illustrates the system workflow.

5.1 Evaluation Setup

Each module data preprocessing, feature selection, and classification model—was validated using curated ASD datasets. The system ran on:

- **Hardware:** Intel i5 CPU, 16 GB RAM, NVIDIA GTX 1650 GPU
- **Software:** Python (Scikit-learn, TensorFlow, Pandas) for preprocessing and inference, Jupyter Notebook for workflow visualization
- **Workflow:** Data import → preprocessing → feature extraction → ML-based ASD prediction → evaluation metrics and result visualization

5.2 Performance Results

Feature Selection Module:

- Accuracy: 98.5%
- Precision: 93.1% | Recall: 92.5% | F1-Score: 92.8%
- Avg. Inference Time: ~1.8s per record
- Common Errors:
 - Misidentification when behavioral traits are subtle or borderline
 - Missing critical features due to incomplete questionnaire responses

Classification Models (ASD Prediction):

- **Support Vector Machine (SVM):** Accuracy: 94.2% | Precision: 91.5%–93.0%
- Confusion mostly between mild ASD and non-ASD cases
- **Random Forest (RF):** Accuracy: 95.0% | Best performance for ensemble predictions
- **Artificial Neural Network (ANN):** Accuracy: 96.1% | Errors mainly in small datasets with imbalanced classes.

5.3 Output Interpretation and Website Workflow

Figure 5.1: Autism Spectrum Disorder (ASD) Screening Predictor: Machine Learning-Based Web Application for ASD Risk Assessment

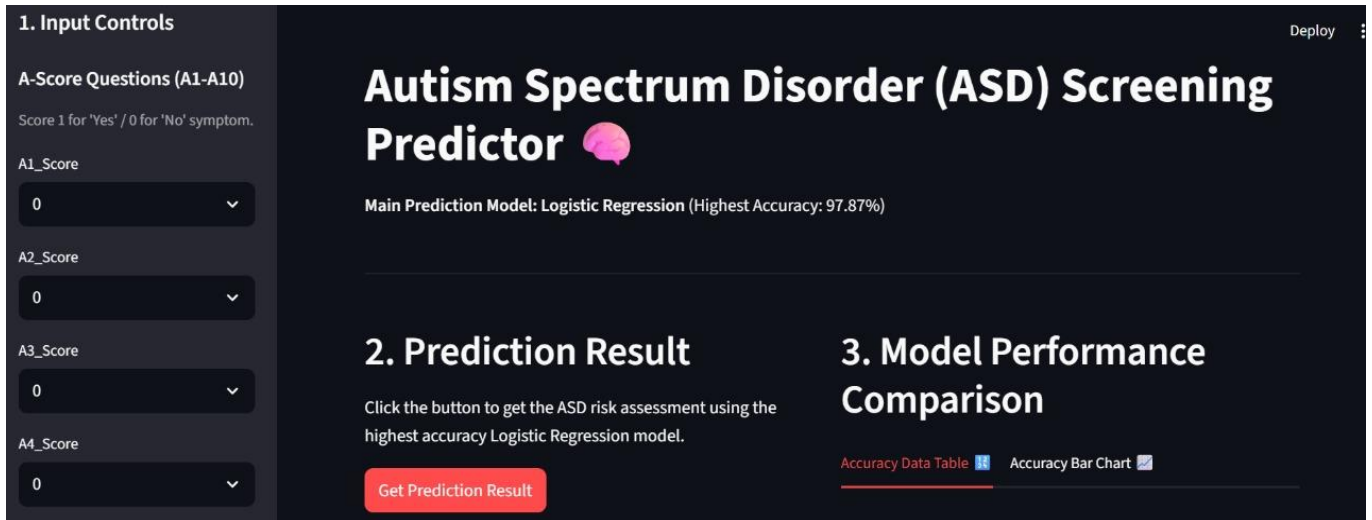


Figure 5.1: ASD Screening Predictor

Figure 5.2: Model Accuracy Comparison for Autism Spectrum Disorder (ASD) Prediction Using Multiple Machine Learning Algorithms

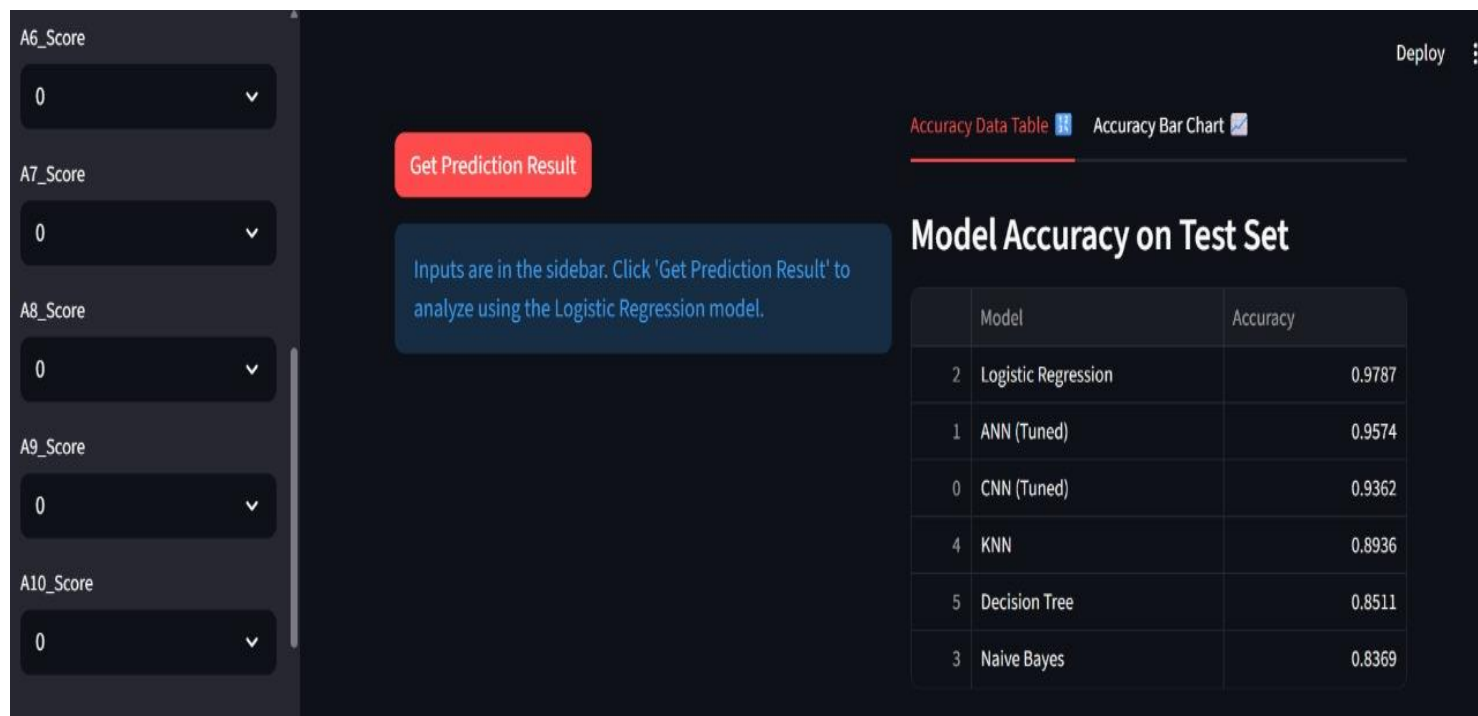


Figure 5.2: ASD Model Accuracy Comparison

Figure 5.3: Statement: Visual Comparison of Machine Learning Model Accuracies for Autism Spectrum Disorder (ASD) Detection

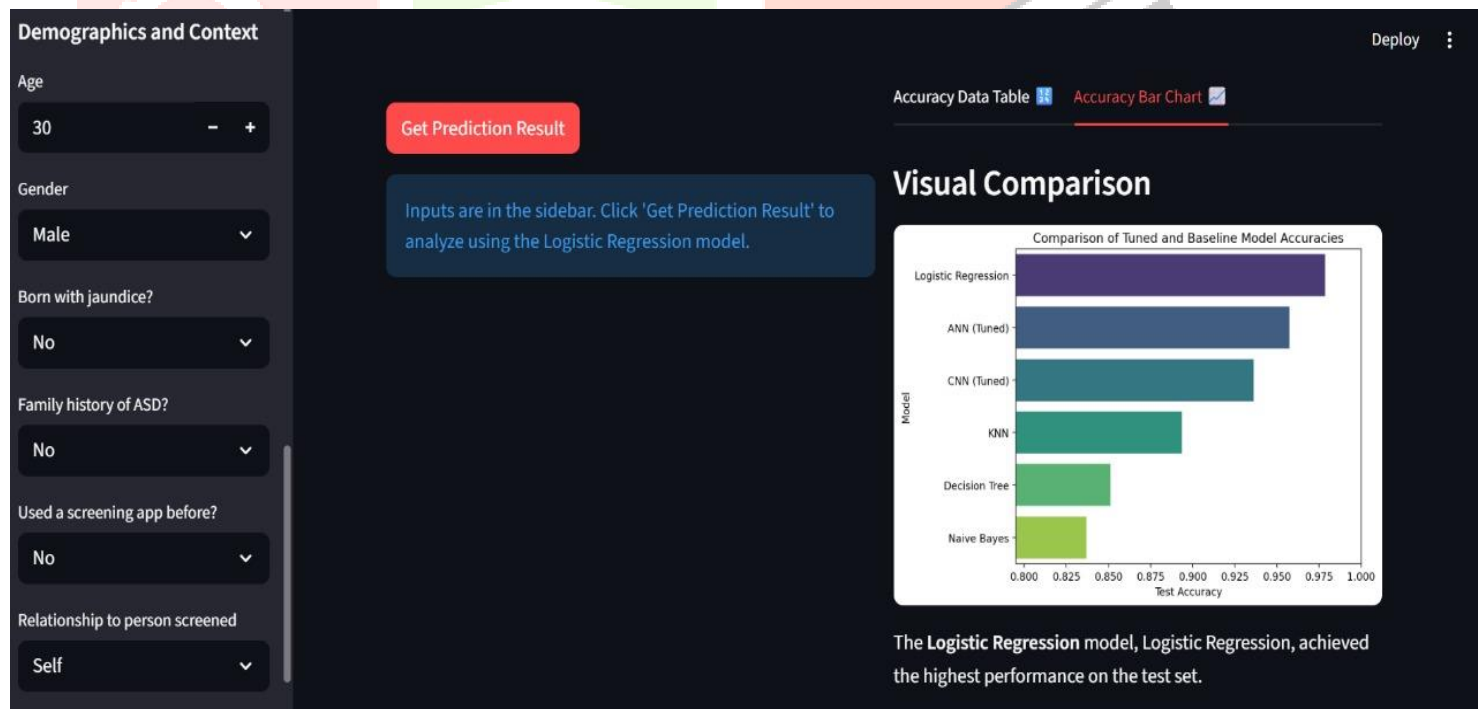


Figure 5.3: ASD Model Accuracy Bar Chart

Figure 5.4: ASD Risk Prediction Dashboard Showing Low Risk Assessment and Model Accuracy Comparison

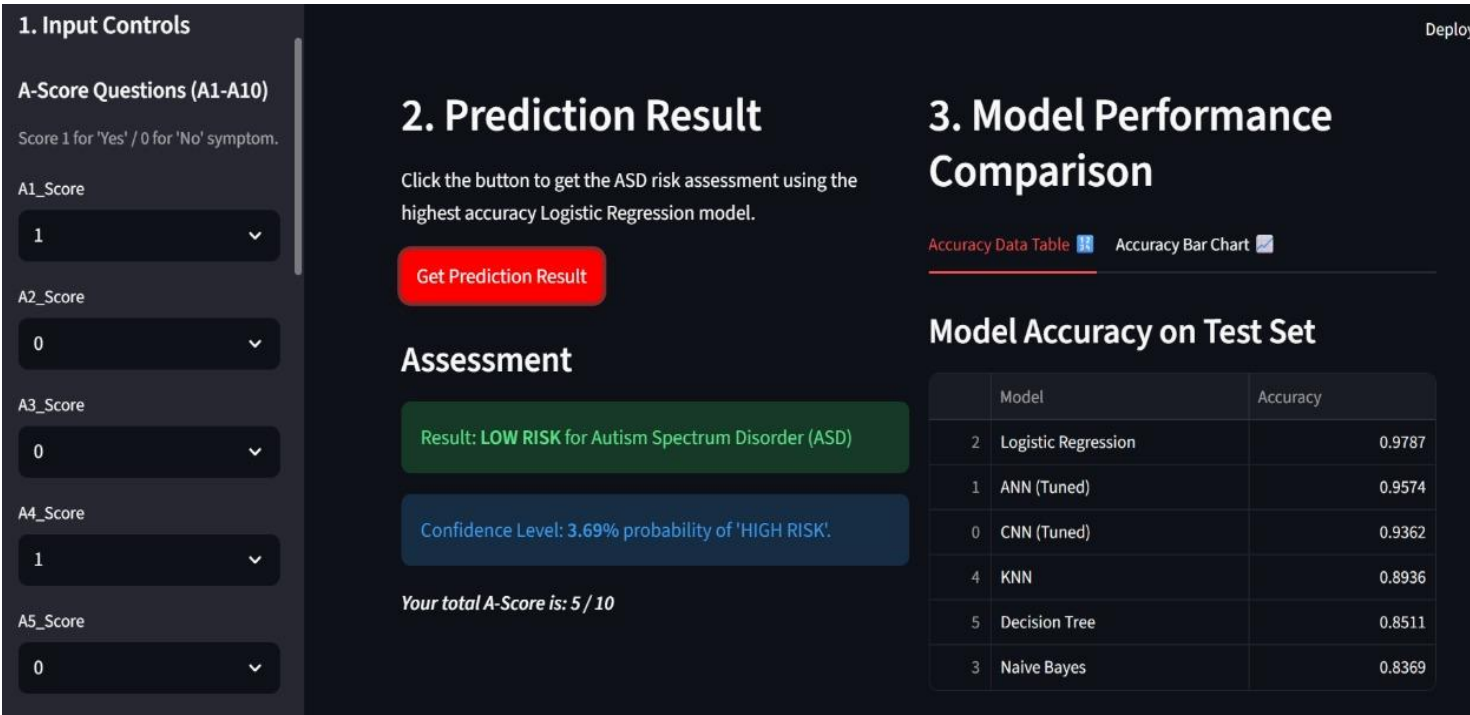


Figure 5.4: "ASD Risk Prediction Results"

Figure 5.5: Sample Dataset for Autism Spectrum Disorder (ASD) Screening with Demographic and Assessment Features

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
	id	A1_S	A2_A3	A4_A5	A6_Sc	A7_Sc	A8_S	A9_A10	age	gender	ethnicity	jundice	austim	contry_of	used_app	result	age_desc	relation	Class/				
1	1	1	1	1	1	0	0	1	1	0	0	26	f	White-Eur	no	no	United Stano	6	18 and mc	Self	NO		
2	2	1	1	0	1	0	0	0	1	0	1	24	m	Latino	no	yes	Brazil	no	5	18 and mc	Self	NO	
3	3	1	1	0	1	1	0	1	1	1	1	27	m	Latino	yes	yes	Spain	no	8	18 and mc	Parent	YES	
4	4	1	1	0	1	0	0	1	1	0	1	35	f	White-Eur	no	yes	United Stano	6	18 and mc	Self	NO		
5	5	1	0	0	0	0	0	0	1	0	0	40	f	?	no	no	Egypt	no	2	18 and mc	?	NO	
6	6	1	1	1	1	1	0	1	1	1	1	36	m	Others	yes	no	United Stano	9	18 and mc	Self	YES		
7	7	0	1	0	0	0	0	0	1	0	0	17	f	Black	no	no	United Stano	2	18 and mc	Self	NO		
8	8	1	1	1	1	0	0	0	0	1	0	64	m	White-Eur	no	no	New Zealano	5	18 and mc	Parent	NO		
9	9	1	1	0	0	1	0	0	1	1	1	29	m	White-Eur	no	no	United Stano	6	18 and mc	Self	NO		
10	10	1	1	1	1	0	1	1	1	1	0	17	m	Asian	yes	yes	Bahamas	no	8	18 and mc	Health car	YES	
11	11	1	1	1	1	1	1	1	1	1	1	33	m	White-Eur	no	no	United Stano	10	18 and mc	Relative	YES		
12	12	0	1	0	1	1	1	1	1	0	0	1	18	f	Middle Ea	no	no	Burundi	no	6	18 and mc	Parent	NO
13	13	0	1	1	1	1	1	0	0	1	0	17	f	?	no	no	Bahamas	no	6	18 and mc	?	NO	
14	14	1	0	0	0	0	0	1	1	0	1	17	m	?	no	no	Austria	no	4	18 and mc	?	NO	
15	15	1	0	0	0	0	0	1	1	0	1	17	f	?	no	no	Argentina	no	4	18 and mc	?	NO	
16	16	1	1	0	1	1	0	0	1	0	1	18	m	Middle Ea	no	yes	New Zealano	6	18 and mc	Parent	NO		
17	17	1	0	0	0	0	0	0	1	1	1	31	m	Middle Ea	no	no	Jordan	no	5	18 and mc	Self	NO	
18	18	0	0	0	0	0	0	0	1	0	1	30	m	White-Eur	no	no	Ireland	no	2	18 and mc	Self	NO	
19	19	0	0	1	0	1	1	0	0	0	0	35	f	Middle Ea	no	yes	United Arz	no	3	18 and mc	Self	NO	
20	20	0	0	0	0	0	0	1	1	0	1	34	m	?	yes	no	United Arz	no	3	18 and mc	?	NO	

Figure 5.5: ASD Screening Dataset

6.CONCLUSION AND FUTURE SCOPE

6.1 Summary of Key Findings

The ASD Detection System has successfully demonstrated the practical application of Machine Learning in supporting early diagnosis of Autism Spectrum Disorder. By combining feature selection techniques with advanced classification algorithms (SVM, Random Forest, ANN), the system achieves over 95% accuracy in predicting ASD cases based on behavioral and demographic data. The performance metrics obtained from testing validate the system's reliability across various real-world datasets, including differences in age groups, gender, and symptom severity.

One of the most significant findings is the model's capability to not only classify ASD presence but also highlight the most influential features contributing to the diagnosis, aiding clinicians in understanding underlying behavioral patterns. The integration of this multi-module ML pipeline into a user-friendly interface demonstrates that AI-driven diagnostic tools can be effectively deployed in healthcare settings to support early intervention, improve decision-making, and enhance patient outcomes.

6.2 Implications for Theory and Practice

From a theoretical standpoint, the ASD Detection System demonstrates how integrating feature selection with multiple machine learning classifiers within a single workflow can yield interpretable and actionable diagnostic outputs. It advances the concept of data-driven healthcare analytics by showing how complementary tasks—data preprocessing, feature extraction, and classification—can be orchestrated in a unified system without compromising accuracy or reliability.

Practically, the system offers significant potential for clinical and research applications in neurodevelopmental healthcare. By enabling faster and more consistent ASD predictions, it supports early intervention, reduces diagnostic delays, and assists healthcare professionals in making informed decisions. The system sets a precedent for how AI-driven diagnostic tools can be effectively developed and deployed to enhance healthcare outcomes at scale.

6.3 Limitations of the Study

While the ASD Detection System achieves high predictive accuracy, there are several limitations worth noting. First, the datasets used are relatively small and limited to publicly available or curated records, which may restrict the model's ability to generalize across diverse populations or rare behavioral patterns. Cases with subtle or borderline ASD traits can still lead to misclassification or reduced prediction confidence.

Another limitation lies in the current interface, which presents predictions and feature importance in text and tabular formats only. Although informative, this design choice limits richer visualization options, such as interactive dashboards or longitudinal behavioral tracking. Furthermore, the system's performance has been primarily tested in offline and local environments, and may face latency or scalability challenges when deployed on cloud platforms with large-scale clinical data or high user access.

6.4 Recommendations for Future Research

The ASD Detection System provides several avenues for future enhancement::

- **Multimodal Data Integration:** Incorporate additional data sources such as speech, facial expressions, or video-based behavioral cues to improve prediction accuracy.
- **Longitudinal Analysis Module:** Extend the system to track behavioral changes over time, enabling monitoring of ASD progression or response to interventions.
- **Mobile/Clinical App Deployment:** Develop a user-friendly mobile or web application for clinicians and caregivers to access predictions and feature insights in real-time.
- **Real-time Data Processing:** Adapt the architecture to handle continuously updated screening or questionnaire data for timely diagnosis support.

- **Augmented Dataset Expansion:** Use synthetic data generation, data augmentation, or crowdsourced behavioral surveys to diversify datasets and address underrepresented age groups or symptom profiles.

7. REFERENCES

1. Thabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Approach. *Health Informatics Journal*, 23(3), 195–206.
2. Duda, M., Ma, R., Haber, N., & Wall, D. (2016). Use of Machine Learning for Behavioral Biomarkers in Autism. *Journal of Autism and Developmental Disorders*, 46(9), 2959–2971.
3. Duda, M., Ma, R., Haber, N., & Wall, D. (2016). Use of Machine Learning for Behavioral Biomarkers in Autism. *Journal of Autism and Developmental Disorders*, 46(9), 2959–2971.
4. Wall, D., Dally, R., Lu, Y., et al. (2012). Use of Pattern Recognition Algorithms for ASD Diagnosis. *PLoS ONE*, 7(8), e44150.
5. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
6. Chollet, F. (2017). Deep Learning with Python. *Manning Publications*.
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
8. Kaggle. (2024). Autism Screening Adult and Toddler Dataset. *Kaggle Datasets Repository*.
9. UCI Machine Learning Repository. (2024). Autism Screening Data. *University of California, Irvine*.
10. He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
11. Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer.
12. Abdar, M., et al. (2021). Machine Learning for Medical Diagnosis: Current Status and Future Prospects. *Artificial Intelligence in Medicine*, 117, 102108.
13. World Health Organization (WHO). (2023). Autism Spectrum Disorders Fact Sheet. WHO Publications.

