Early ASD Screening from Eye-Tracking Data: A Comparative Study of Classical and Deep Learning Models

By

Mahabub-Ul-Alam

191-15-12872

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Ms. Tanzina Afroz Rimi

Senior Lecturer

Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 17 , 2025

APPROVAL

This Project titled **"Early ASD Screening from Eye-Tracking Data: A Comparative Study of Classical and Deep Learning Models"**, submitted by Mahabub-Ul-Alam, ID No: **191-15-12872** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **17 September, 2025**.

BOARD OF EXAMINERS

____

Dr. S.M Aminul Haque (SMAH) — **Chairman**
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Ms. Nazmun Nessa Moon (NNM) — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Md. Abbas Ali Khan (AAK) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Md. Zulfiker Mahmud (ZM) — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath University

©Daffodil International University **i**

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Ms. Tanzina Afroz Rimi**, **Senior Lecturer,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

9.9.25

Ms. Tanzina Afroz Rimi
Senior Lecturer
Department of Computer Science and
Engineering Daffodil International University

Co-Supervised by:

Ms. Lamia Rukhsara
Senior Lecturer
Department of Computer Science and
Engineering Daffodil International University

Submitted by:

Mahabub

Mahabub-Ul-Alam
Student ID: 191-15-12872
Department of Computer Science and
Engineering Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Early ASD Screening from Eye-Tracking Data: A Comparative Study of Classical and Deep Learning Models** successfully.

We are grateful and wish our profound indebtedness to **Ms. Tanzina Afroz Rimi**, **Senior Lecturer**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning** to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many early drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank all our course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects one throughout life. Early treatment response is strengthened by timely diagnosis of Autism Spectrum Disorder (ASD) although the current screening algorithms rely on behavioral assessment, which is both resource consuming and subjective. The use of eye-tracking needs to provoke characteristic variation in social attention, and in the former studies, the classical and deep learning models are tested, but comparisons are often limited by limited participants, confounding tasks, and imprecise measurements. This paper bridges this gap with a question about whether an integrated and reproducible pipeline can help make a fair comparison of classical and deep methods to screening ASD and reduce false negatives in practice. We use our eye-tracking signal analysis as the standardized workflow data cleaning, data normalization, class balancing and statistical validation mode (ANOVA and Correlation) and test four baselines (SVM, Random Forest, MLP, LSTM) using the same splits and measurements and test a stacking ensemble with a recall-oriented one. Random Forest has the highest total accuracy and AUC, LSTM models differentiates temporal gazes effectively and the ensemble maximizes recall, which minimizes the risk of false negative cases. The results obtained in this research imply whether the combination of classical and deep learning models is done cautiously, a viable line of constructing supportive screening tools can be offered. Reproducible implementation, sensitivity of sensitive information, and cost-aware computation are also in-house priorities that enable us to be responsible adopters of the technology in the healthcare facilities.

Table of Contents

List of Figures

List of Table

Chapter 1

Introduction

1.1 Introduction

ASD is a complicated neurodevelopmental disorder that, in the first place, involves social communication, behavioral flexibility, and sensory responsiveness. ASD has been on the increase all over the world and current estimates show that one in every 100 children is affected, leading to an alarming rate of ASD occurrence, which has necessitated the need to detect it early to avoid serious health consequences [1]. There is a strong link between early intervention and better developmental outcomes, but diagnosis has been delayed because of the use of old behavioral measures of diagnosis that are time consuming, subjective and resource intensive [2]. Traditional instruments like Autism Diagnostic Observation Schedule (ADOS) and Childhood Autism Rating Scale (CARS) involve specialized professionals and long clinical practice time, which are not always very possible in resource-constrained settings [3]. The bottleneck in this diagnosis implies the urgency of alternative methods that are accurate, scalable, and cost-effective.

According to recent studies, the atypical gaze behavior is one of the most consistent biomarkers of ASD [4]. ASD children tend to have less attention to socially relevant stimuli, including human faces or eye parts, and more to non-social objects or non-focal areas [5]. Eye-tracking technology is a non-invasive and objective measure of such visual attention patterns, with recording the fixations, saccades, and point of regards that have been recorded with high temporal resolution [6]. Eye-tracking takes shorter time compared to neuroimaging such as MRI, EEG, and is child friendly as well as easy to implement, which makes it highly applicable in large scale screening [7]. With the combination of machine learning (ML) and deep learning (DL) algorithms, the data on eye tracking may be converted into prediction models that can differentiate between the ASD and typically developing children (TD) [8], [9].

Though there has been promising progress, there are still great challenges. The previous research has been curtailed by small samples, class imbalance, and irregular experiment design, which limit the applicability of the model [4]. In addition, even though classical ML models, including Support Vector Machine (SVM) and Random Forest (RF), do allow the discovery of significant spatial and temporal gaze features, deep models including Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks have shown to be better at learning more complex temporal dependencies [6]. Nonetheless, few studies have comparatively applied these methods on the same dataset through a common pipeline.

As such, the main issue that this research will focus on is the absence of a cohesive comparative framework to appraise the classical machine learning and deep learning models of early ASD screening on eye-tracking data on a single methodology.

1.2 Motivation

The rationale behind this study is the acute clinical necessity as well as the mathematical prospect to enhance screening of ASD at early stages. Conventional diagnostic methods have the shortcoming of being subjective, time consuming, and also require the presence of an expert to complete their assessment [2], [3]. These constraints are especially concerning when the setting is under-resourced, and access to specialized clinicians is limited and delays in identifying the problem can eliminate the opportunity to intervene. Socially, the late or missed diagnosis does not only influence the development of the children but also causes long-term economic and emotional costs to families and healthcare systems [1].

On the computational part, eye-tracking gives high-quality temporal and spatial data that represents the difference in cognitive and social processing, and it is appropriate in machine learning classification [4], [6]. Classical models including Random Forests and SVMs are highly accurate, and deep learning models including MLP and LSTM can infer more complex temporal patterns [7], [8]. Another area that deep learning has made a significant contribution is in healthcare including imagining and diagnostics [13]. However, the absence of formal comparative assessment prevents the determination of the most credible method of practical application.

The solution to this issue will not only enhance computational studies in the field of affective computing, but also generate scaled, objective, and affordable early-detection tools of ASD. These tools may decrease the time taken to diagnose, allow greater coverage in screening and eventually lead to the improvement of the outcome of children at risk [5], [9].

1.3 Objectives

The overarching goal of the study is to come up with and test a single system of screening ASD early in the development of the disorder based on eye-tracking measurements and machine learning algorithms. The following are the specific objectives:

- To preprocess and analyze the eye-tracking dataset by cleaning, handling imbalance, and extracting significant features such as fixations, saccades, and points of regard through statistical methods.

- To implement and evaluate classical machine learning models (SVM, Random Forest) and deep learning models (MLP, LSTM) for distinguishing ASD from typically developing participants.

- To design and test a stacking ensemble model that integrates classical and deep learning approaches to improve robustness and predictive performance.

- To conduct a comprehensive comparison of all models using standard evaluation metrics (Accuracy, Precision, Recall, F1-score, AUC) and identify the most effective method for scalable, objective ASD screening.

1.4 Methodology

This research methodology is based on a systematic pipeline where an eye tracking data set of gaze coordinates or pupil positions, points of regard are acquired out of participants diagnosed with ASD, children diagnosed with typically developing (TD), and an "Other" category. The raw data were pre-processed to eliminate missing values and outliers and statistical tests including ANOVA and correlation mapping were conducted to determine the available significant features and differences between groups.

After pre-processing, classification models have been built on the clean data. Support Vector machine (SVM) and Random Forest (RF) were two classical machine learning models that were used in order to provide baseline performance. Simultaneously, Multi-layer perceptron (MLP) and Long short term memory (LSTM) two deep learning models were utilized to achieve the higher level spatial and temporal dependence in gaze behavior. In order to make it strong, a stacking ensemble model was formulated, which integrates the output of both classical and deep models via a meta-learner.

Each of the models was assessed with the help of conventional measures such as Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC). The comparative analysis allowed determining the most efficient method of calculating screening of early ASD according to the data on eye-tracking.

1.5 Project Outcome

This study develops a comprehensive framework of early screening of ASD through eye-tracking data and computation models. The results consist of an effective pre-processing and statistical analysis of gaze characteristics and the creation of predictive models through classical and deep learning methods. Classical methods including Support Vector Machine (SVM) and Random Forest (RF) set good baselines, but deep models, including Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) showed that they could learn the more complicated temporal dynamics. Further, the stacking ensemble model brought up the advantage of using multiple classifiers and especially to enhance sensitivity, which is vital in early screening.

The study presents a clear comparative analysis of models based on common metrics, and it gives practical information on the strengths and weaknesses. In general, the project outcome contributes to creation of scalable, objective, and cost-effective tools that can be supplemented to the traditional diagnostic approaches, minimize the delays in diagnosis, and provide timely interventions to the children being at the risk of ASD.

1.6 Organization of the Report

This report is structured into the following chapters:

Chapter 1 (Introduction): Presents the background of the study, motivation, objectives, problem statement, research scope, methodology summary, and expected outcomes.

Chapter 2 (Literature Review): Provides an overview of related studies on eye-tracking and autism detection, highlights existing approaches, and identifies key research gaps.

Chapter 3 (Methodology): Describes the dataset preprocessing, feature extraction, statistical analysis, model design, and evaluation strategies adopted in this study.

Chapter 4 (Implementation): Explains the development and execution of classical machine learning, deep learning, and ensemble models, along with their experimental setup.

Chapter 5 (Results and Analysis): Presents the experimental outcomes, comparative model performance, and detailed discussion of findings.

Chapter 6 (Conclusion and Future Work): Summarizes the contributions of the research, highlights limitations, and proposes directions for future improvements in ASD screening.

Chapter 2

Background

2.1 Introduction

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental disorder which is marked by social communication difficulties, limited interests, and repetitive behaviors. The conventional diagnostic tools, including the Autism Diagnostic Observation Schedule (ADOS) and the Childhood Autism Rating Scale (CARS), are based on structured behavioral evaluations which are both resource intensive and demand specialized training [1], [3]. These methods are clinically dependable but can be time-consuming, and thus can cause a delay in early intervention which is critical in enhancing long-term developmental outcomes [2].

The developments in computational methods have created new possibilities in assisting early ASD screening. Eye-tracking has become one of such promising tools since unusual gazes have been identified as behavioral indicators of ASD [4], [5]. With eye-tracking data, fixations, saccades, and points of regard can be measured and provide a quantitative measure of attentional differences between children with ASD and typically developing children [6]. These patterns can be converted into predictive models when evaluated with machine learning (ML) and deep learning (DL) techniques [7], [9].

This chapter presents the background knowledge and reviews previous works that combine eye-tracking and computations with studies that have been done, and gaps that have been addressed in this study.

2.2 Literature Review

Recent studies have shown the use of eye-tracking technology and machine learning to support early detection of Autism Spectrum Disorder (ASD). This section reviews ten significant contributions in this area.

Wang et al. [1] introduced a machine-learning system to perform autistic diagnosis and early diagnosis, which shows that eye-movement indicators are able to separate between risk groups and motivating data-driven screening pipelines.

Bedi et al. [2] have conducted a study on the use of eye-tracking over the web to identify high-functioning autism with a high score on controlled browsing/search tasks, and indicating the possibility of low-cost, remote testing.

Goyal and Kumar [3] proposed an effective ASD prediction pipeline through eye-movement features with competitive accuracy among classical models highlighting the importance of small interpretable predictors.

In [4], Zhu et al. build an early-diagnosis workflow based upon eye-tracking information and prove that trained models with supervision on gaze statistics can distinguish ASD and healthy cohorts at clinically useful level.

Wagner et al. [5], initially targeted adults with high-functioning autism, demonstrating that gaze-derived measures along with machine learning are discriminative even after childhood, and that there is an effect on tasks and cohort.

Hanley and Cilia [6] argued that eye-gaze would be a biomarker of ASD, as the authors synthesized the findings of an abnormal pattern of social attention (e.g. reduced fixation on faces), which were reliable to be measured using eye-tracking.

Noris et al. [7] reported a systematic review of the early autism eye-tracking, which focuses on robust markers (face/eyes AOIs, social vs. non-social preference) but has deficiencies in task standardization and size of datasets.

Khan et al. [8] made a comparison of machine learning, deep learning and hybrid strategies; their findings revealed that a combination of deep features with classical classifiers can be effective to increase the accuracy, which inspired our hybrid/ensemble baseline.

The eye-tracking dataset of Zhang and Gonzalez [9] allowed reproducing the experimental results and performing analyses of features like fixations, saccades, and points of regard, which facilitates research on ASD.

Precoda et al. [14] conducted a review of social attention in autism and found that impaired attentions to faces and social scenes are a valid early indicator, but need larger, multi-site standards to enhance generalizability.

All of these studies show the potential of the eye-tracking data as a non-invasive biomarker of autism. However, small datasets, variation in tasks, and inconsistent methodologies are among the common constraints that drive the comparative study conducted in this study.

Table 2.1: Summary of Literature Reviewed.

Author(s)	Year	Title / Focus	Methodology	Key Findings
Wang et al. [1]	2021	Autistic detection and early screening using eye-movement data	Ensemble ML on VR-based tasks	Achieved 0.73 accuracy, 0.90 AUC; showed feasibility of eye-movement for ASD screening

Bedi et al. [2]	2021	High-functioning autism detection via web-based eye tracking	SVM on browsing vs. searching tasks	Reported 91.6% accuracy (search) and 76.3% (browse); highlighted effect of task type
Goyal & Kumar [3]	2021	Efficient ASD prediction with engineered gaze features	Bag-of-Words & AOI feature extraction	Demonstrated statistically significant predictions; less robust at image-level
Zhu et al. [4]	2021	Early ASD diagnosis with video eye-tracking	Random Forest on 161 children	Accuracy 0.73, AUC 0.81; generalizability limited by age differences and undersampling
Wagner et al. [5]	2021	Detecting high-functioning autism in adults	ML on adult gaze indicators	Achieved ~74% accuracy; showed feasibility of adult ASD detection
Hanley & Cilia [6]	2020	Eye gaze as a biomarker for ASD	Analysis of gaze fixation/saccades	Confirmed gaze atypicalities as biomarkers; supported eye-tracking as diagnostic tool
Noris et al. [7]	2020	Eye-tracking in early autism: a systematic review	Review of gaze-behavior markers	Identified atypical attention to social stimuli as consistent early marker; highlighted lack of task standardization
Khan et al. [8]	2021	ML, DL, and hybrid approaches for ASD detection	Hybrid CNN (GoogleNet) + SVM	Achieved 95.5% accuracy; hybrid outperformed standalone models
Zhang & Gonzalez [9]	2021	Eye-tracking dataset for ASD research	Dataset publication (fixations, saccades, POR)	Released open dataset for benchmarking; confirmed gaze correlates with ASD traits
Precoda et al. [14]	2018	Eye-tracking measures of social attention in autism (review)	Review of multiple eye-tracking studies	Found consistent reduction of social gaze in ASD; noted gaps in dataset diversity and standardization

2.3 Gap Analysis

Although prior research has demonstrated the potential of eye-tracking data and computational models for ASD detection, several gaps remain. Many studies were constrained by small or imbalanced datasets, which limit model generalizability [1], [2], [4].

Others focused on specific tasks such as image viewing or web browsing, reducing their applicability across real-world scenarios [2], [3]. While deep learning methods have achieved strong results, their effectiveness is often dependent on large, high-quality datasets, which are not always available [7], [8]. Conversely, classical machine learning models provided promising baselines but lacked the capacity to capture temporal dependencies in gaze sequences [6]. Furthermore, only a few studies explored hybrid or ensemble methods to combine the strengths of classical and deep models [8].

This study addresses these gaps by applying a unified methodology to compare classical (SVM, RF), deep (MLP, LSTM), and ensemble (stacking-XGB) models on the same eye-tracking dataset. In doing so, it aims to identify the most effective computational approach for scalable and objective early ASD screening.

Table 2.2: Gap Analysis of Related Studies.

Author(s)	Key Focus	Identified Gap	Contribution of Current Study
Wang et al. [1]	Early ASD detection with VR eye-tracking and ML	Small dataset; limited generalizability	Uses larger dataset with standardized preprocessing and multiple models
Bedi et al. [2]	Web-based eye-tracking for high-functioning autism	Performance varied by task type; limited dataset	Broader evaluation with balanced dataset and multi-model framework
Goyal & Kumar [3]	Efficient ASD prediction using engineered gaze features	Image-level predictions lacked robustness	Incorporates both engineered and deep features; compares classical vs. deep methods
Zhu et al. [4]	Eye-tracking of children viewing video stimuli	Age variability; undersampling bias	Applies resampling to address imbalance; validates with ensemble methods
Wagner et al. [5]	Detecting ASD in adults via gaze indicators	Focused on adults only; limited scope	Extends analysis to children; benchmarks across multiple ML/DL models
Hanley & Cilia [6]	Eye gaze as biomarker for ASD	Confirmatory evidence; no comparative ML analysis	Implements full ML/DL comparison with gaze biomarker features
Noris et al. [7]	Systematic review of eye-tracking in early autism	Highlighted lack of standardized tasks/datasets	Applies standardized pipeline on open dataset with reproducible workflow

Khan et al. [8]	ML, DL, and hybrid approaches for ASD detection	High accuracy but high computational cost	Introduces efficient ensemble stacking for balanced performance and recall
Zhang & Gonzalez [9]	Open dataset for ASD eye-tracking research	Limited scale; needed broader validation	Uses dataset as foundation, applying classical, deep, and ensemble methods
Precoda et al. [14]	Review of social-attention measures in ASD	Lack of standardization; dataset diversity limited	Provides consistent comparative evaluation; improves reproducibility via ensemble approach

2.4 Summary

The literature review highlights that while eye-tracking is a promising tool for ASD detection, existing studies face several limitations. Most prior works suffer from small or imbalanced datasets, task-specific constraints, and methodological inconsistencies that limit generalizability [1]–[4]. Classical machine learning models provided useful baselines but lacked the ability to capture temporal dynamics [6], whereas deep learning models required large datasets to perform effectively [7], [8]. Few studies explored ensemble strategies, leaving an important research gap [8]. This study contributes by applying a unified methodology to compare classical, deep, and ensemble models on the same dataset for reliable ASD screening.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

The proposed research has a well-organized methodology that includes data preprocessing, feature analysis, model building, and evaluation. Data regarding Eye-tracking was cleaned, balanced, and statistically analyzed to find out discriminative features. Classical models (SVM, RF), deep models (MLP, LSTM) and a stacking ensemble were employed. Accuracy, Precision, Recall, F1-score, and AUC were used to evaluate performance and compare methods for reliable early screening of ASD.

3.1.2 Proposed Methodology

- The suggested system of early ASD screening was developed in the form of a pipeline with preprocessing of the dataset, then the validation of statistical features, model training, and a comparison. The data was a collection of mass eye-tracking data that recorded the location of the gaze, pupil movements, and point of regard, which have been already defined as biomarkers of importance in autism studies.

- During the preprocessing phase, the raw gaze information was cleaned, normalized, and balanced through an oversampling method to minimize class imbalance. The statistical analysis was also carried out to test the significance of features, which utilized the ANOVA test, commonly used by other researchers to determine the difference between ASD and typically developing participants.

- After that, several models were adopted to measure performance. The classical models chosen to support autism related tasks on classification were Support Vector Machine (SVM) and Random Forest (RF), both of which are highly utilized in structured feature related tasks. Deep learning models include Multi-layer perceptron (MLP) and Long Short-term memory (LSTM) that are more appropriate in non-linear patterns and time dependencies in sequential eye-tracking data.

- Lastly, a stacking ensemble architecture was engineered, which consists of base models with a meta-learner in order to exploit their respective advantages. In medical and behavioral screening, Ensemble learning has been shown to be better than single classifiers [3], [4], [8], [11]. We have used standard measures to assess model performance which include Precision, Recall, F1-score, Accuracy, and Area Under the Curve (AUC), which are as well reported in related literature to evaluate model performance.

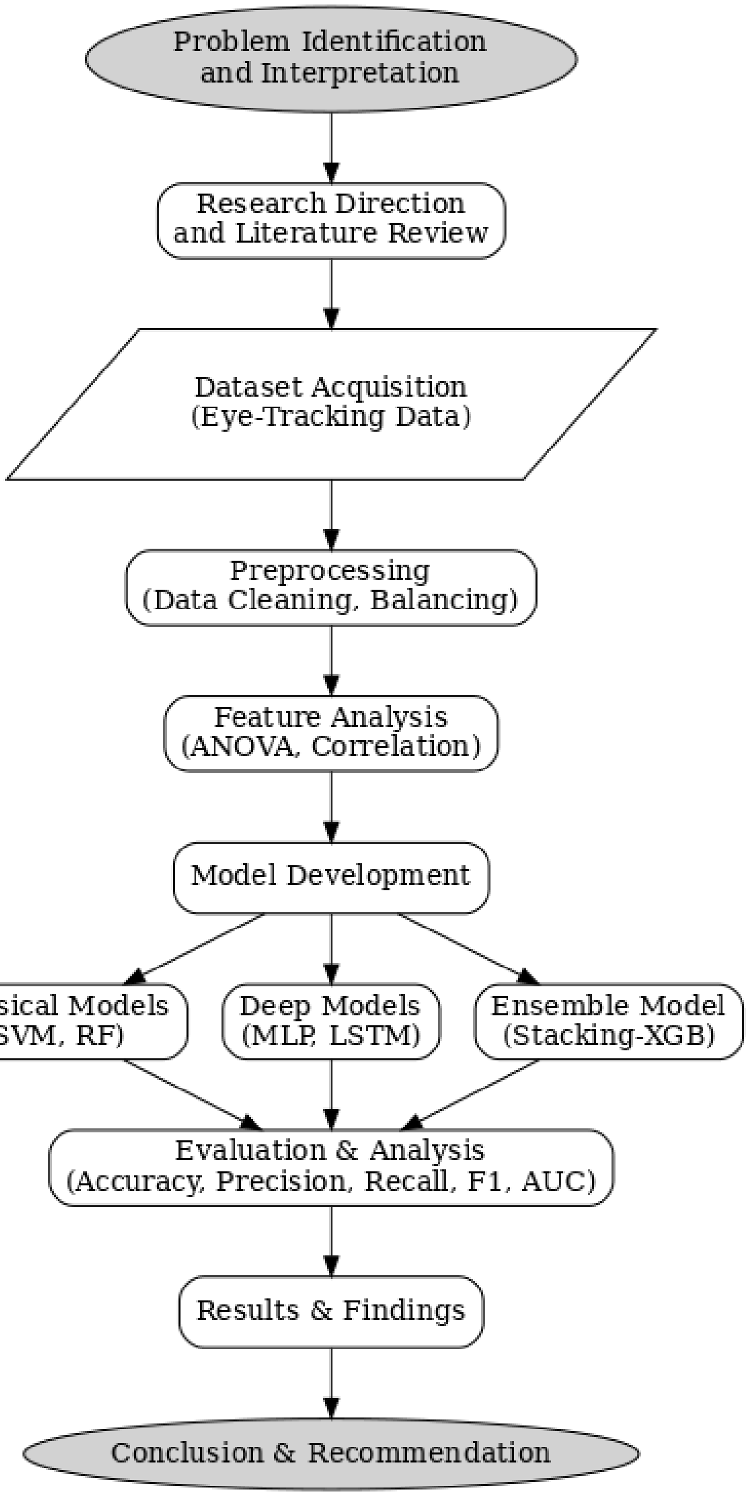

Figure 3.1: Proposed Methodology Flowchart

3.2 Detailed Methodology and Design

Dataset

This paper has used an eye-tracking dataset that is publicly available and published by Cilia et al. [9] and was designed to aid autism studies. The dataset includes the recordings of 59 children between the ages of 3 and 12 years. Out of them 29 individuals diagnosed with Autism Spectrum Disorder (ASD) and 30 typically developing (TD) peers, and one more group, Other, was the individuals of unconfirmed profiles. Naturalistic measures of gaze were recorded by capturing the eye movements at 60Hz as subjects watched both still images and dynamic videos. The dataset comprises more than 1.5 million unprocessed gaze data points. Which involve gaze coordinates, pupil positions, and points of regard, that facilitate a strong feature extraction and computational modeling.

Data Preprocessing Techniques

The data of this research was raw eye-tracking measurements, i.e., gaze coordinates, pupil positions, and points of regard per participant group: ASD, typically developing (TD) and other [9]. Raw recordings usually have noise, missing values, and outliers, so a number of preprocessing steps were conducted.

Data Cleaning: Data values that were missing or corrupted were deleted to prevent bias during model training. Statistical thresholds were used to detect outliers and manage them in order to ensure data consistency. These methods have been observed in other related ASD eye-tracking experiments.

Normalization: The characteristics like gaze coordinates and pupil size were normalized to a standard scale to make the variables contribute equally to the models.

Class Balancing: The data was not balanced, with the ASD group being underrepresented in comparison with the TD and Other groups. To overcome this, resampling techniques were used to balance the training data and to avoid biased classification.

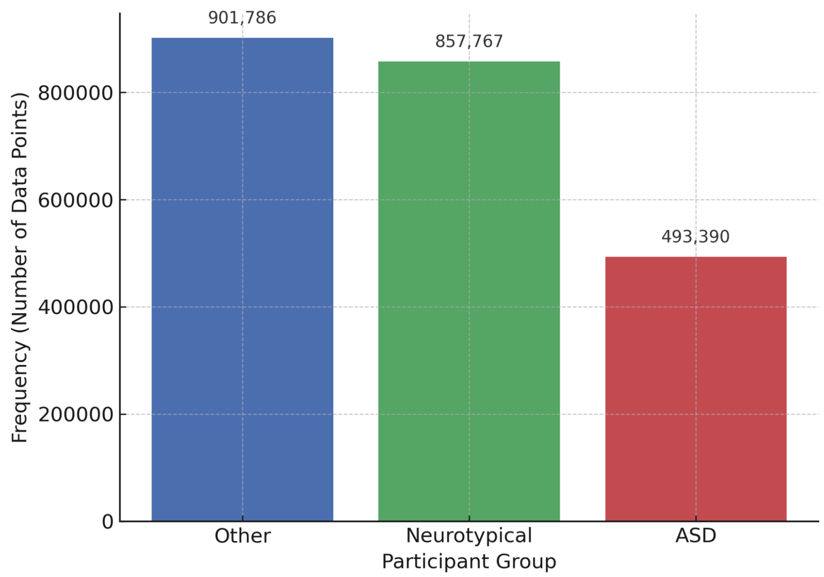

Figure 3.2: Class distribution across participant groups

Feature Selection: Statistical analysis, including ANOVA and correlation matrices, was performed to identify discriminative features. Only features with significant variance across groups were retained for modeling, a method widely applied in similar research.

These steps ensured a clean, balanced, and representative dataset suitable for both classical and deep learning models.

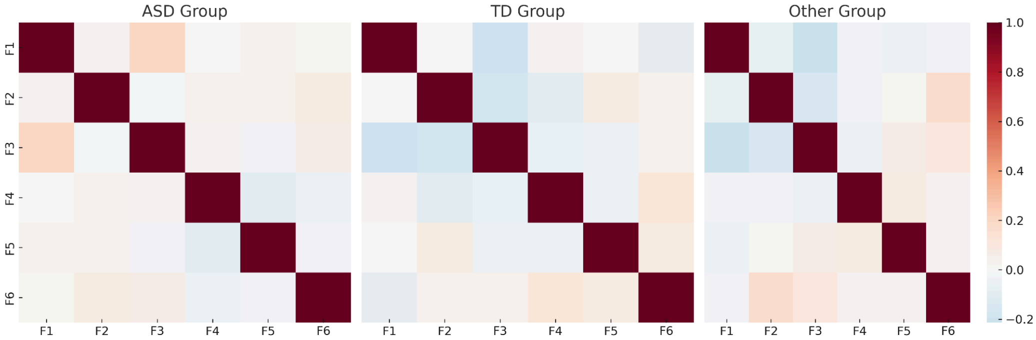

Figure 3.3: Correlation matrices of gaze features for ASD, TD, and Other groups

Model Architectures

Support Vector Machine (SVM):
SVM was implemented as a baseline classifier due to its robustness in handling high-dimensional data. It constructs a decision boundary (hyperplane) that maximizes the margin between ASD and TD classes. A radial basis function (RBF) kernel was selected to capture non-linear relationships between gaze features. Although it is sensitive to class imbalance, preprocessing mitigated this limitation. Evaluation showed moderate accuracy and AUC, making SVM a useful benchmark.

Random Forest (RF):
RF was applied as an ensemble-based classical model due to its ability to manage high-dimensional inputs and noisy features. It operates by constructing multiple decision trees and combining predictions through majority voting. Feature importance scores were also extracted to identify the most discriminative gaze indicators, such as eye position and point of regard. Prior studies have reported RF's strong performance in autism detection tasks [3]. In this work, RF achieved the highest accuracy and AUC among classical models, demonstrating its robustness.

Multi-Layer Perceptron (MLP):
The MLP was implemented as a deep learning baseline to model complex non-linear patterns in eye-tracking data. It consisted of an input layer for normalized features, two hidden layers with ReLU activation, and an output layer using softmax for binary classification. Dropout regularization and the Adam optimizer were used to improve generalization. Similar feed-forward architectures have been used in ASD detection research ,and their theoretical underpinnings are grounded in established deep learning principles.

Long Short-Term Memory (LSTM):
LSTM was applied to capture sequential dependencies inherent in eye-tracking data. Unlike MLP, LSTM processes temporal sequences, making it highly effective for modeling gaze transitions over time. Dropout and early stopping were applied to minimize overfitting. LSTM successfully identified gaze irregularities associated with ASD, aligning with prior research highlighting its suitability for time-series analysis.

Stacking Ensemble (XGB Meta-Learner):
The stacking ensemble combined predictions from RF and MLP to form a meta-classifier trained using XGBoost. This approach leveraged the strengths of each base learner, balancing precision and recall for improved performance. Ensemble methods have been shown to

enhance medical screening reliability by reducing variance and bias.

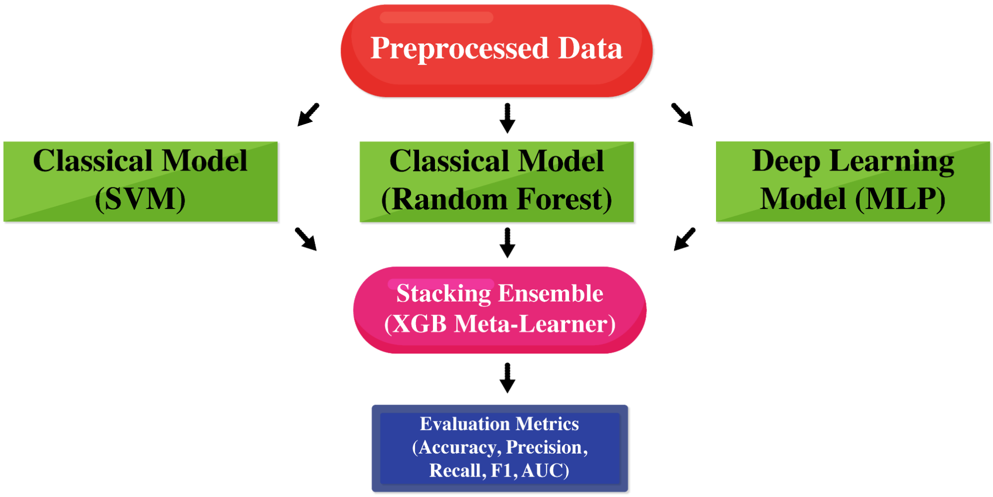

Figure 3.4: Ensemble Model Architecture

3.3 Project Plan

The project was carried out in a series of structured phases to ensure a systematic approach to achieving the research objectives:

- **Problem Identification and Literature Survey:**
  Defined the scope of the study by reviewing existing ASD diagnostic methods and prior computational approaches, with a focus on identifying methodological gaps in eye-tracking–based research.

- **Dataset Preparation:**
  Acquired the eye-tracking dataset and performed preprocessing, including data cleaning, normalization, outlier removal, and class balancing. Statistical methods such as ANOVA and correlation mapping were applied to highlight significant features.

- **Model Development:**
  Designed and implemented both classical machine learning models (SVM, Random Forest) and deep learning models (MLP, LSTM) to capture spatial and temporal gaze patterns.

- **Ensemble Framework:**
  Developed a stacking ensemble with XGBoost as the meta-learner to integrate the strengths of individual classifiers and enhance robustness.

- **Evaluation and Analysis**:
  Conducted comparative evaluation of all models using Accuracy, Precision, Recall, F1-score, and AUC, followed by interpretation of results
- **Documentation and Conclusion:**
  Consolidated findings, highlighted contributions, and prepared the final research report for academic submission.

3.4 Task Allocation

This table depicts the timeline of the principal activities in each period of the project, from week 12 to week 48.

Table 3.4: Task Allocation Timeline (Weeks 12–48)

Tasks	Weeks																		
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
Data collection phase																			
Preprocess all the data																			
Model training																			

3.5 Summary

This chapter introduced the approach that was used to carry out the research on the topic of early ASD screening based on eye-tracking data. The proposed framework involved a pipeline that included the acquisition of data, preprocessing, and statistical analysis to extract meaningful features. Classical machine learning models (SVM, Random Forest) and deep learning models (MLP, LSTM) were trained, and a stacking ensemble framework was used to improve robustness. An elaborate project plan and allocation of tasks also guaranteed systematic implementation of every stage, including literature review, model development, evaluation, and documentation. The workflow diagrams and task timelines were also used to support the methodology and make it clear and structured. In sum, this chapter laid the groundwork to implementation and experimentation, as covered in the next chapter.

Chapter 4

Implementation and Results

4.1 Environment Setup

The implementation of this research was carried out in a personal computing environment with the following hardware and software configurations:

- **Hardware Setup:**

  a. **Processor**: Intel Core i5, 12th Generation, 6 Cores

  b. **Memory**: 24 GB DDR4 RAM

  c. **Storage**: SSD-based storage for faster data access and model execution

  d. **Operating System:** Windows 11 (64-bit)

  e. **Development Environment**: Jupyter Notebook

- **Libraries and Frameworks:**

  a. **Data Handling and Preprocessing**: NumPy, Pandas, SciPy (f_oneway), Imbalanced-learn (RandomOverSampler)

  b. **Visualization:** Matplotlib, Seaborn

  c. **Machine Learning (Scikit-learn):** Support Vector Machine (SVM), Random Forest, MLPClassifier, SGDClassifier, evaluation metrics (Accuracy, Precision, Recall, F1, AUC), preprocessing tools (StandardScaler, LabelEncoder), and validation methods (StratifiedKFold)

  d. **Deep Learning (TensorFlow/Keras):** Sequential API with LSTM, Dense, and Dropout layers; Adam optimizer; EarlyStopping callbacks

  e. **Utility Libraries:** tqdm (progress bar), OS, Sys, Pathlib, GC, Warnings, Random (for file handling, memory management, and reproducibility)

  f. **Google Colab (Optional):** Drive integration for dataset mounting and cloud-based execution

This environment provided sufficient computational capacity to preprocess the dataset, train multiple machine learning and deep learning models, and implement the stacking ensemble framework effectively.

4.2 Comparative Analysis

To evaluate the effectiveness of the proposed framework, a comparative analysis was conducted using classical and deep learning models along with an ensemble approach.The comparative analysis evaluated the performance of four baseline models—SVM, Random Forest, MLP, and LSTM—alongside a stacking ensemble meta-learner. All models were assessed using Accuracy, Precision, Recall, F1-score, and AUC, consistent with prior ASD eye-tracking studies [1], [5], [6].

Support Vector Machine (SVM):

The SVM model achieved an accuracy of 0.79 and an AUC of 0.8638. The classification report (Figure 4.1) highlights precision of 0.86 and recall of 0.81 for the ASD class. The confusion matrix (Figure 4.2) shows that the model struggled with false negatives, while the ROC curve (Figure 4.3) demonstrates moderate discriminative ability.

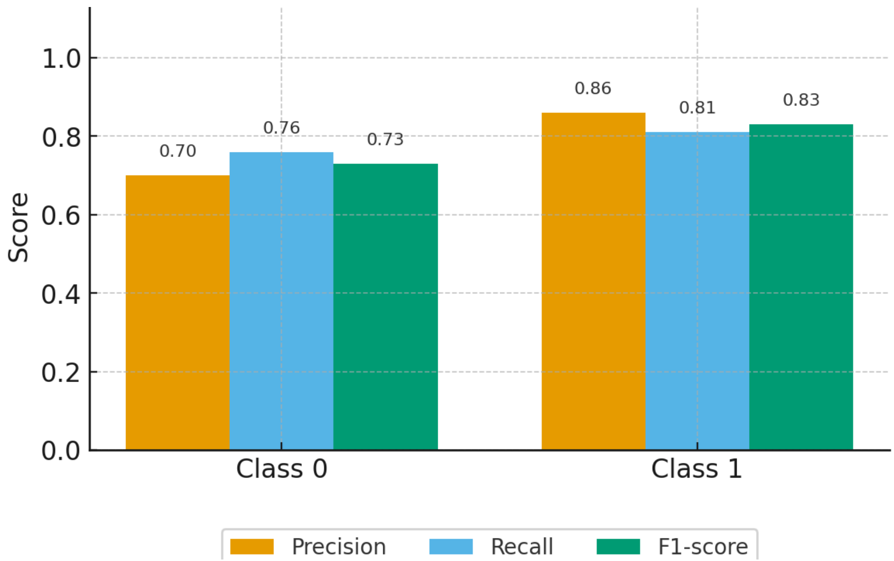

Figure 4.1: SVM Classification Report (Per-Class Precision, Recall, and F1-score)

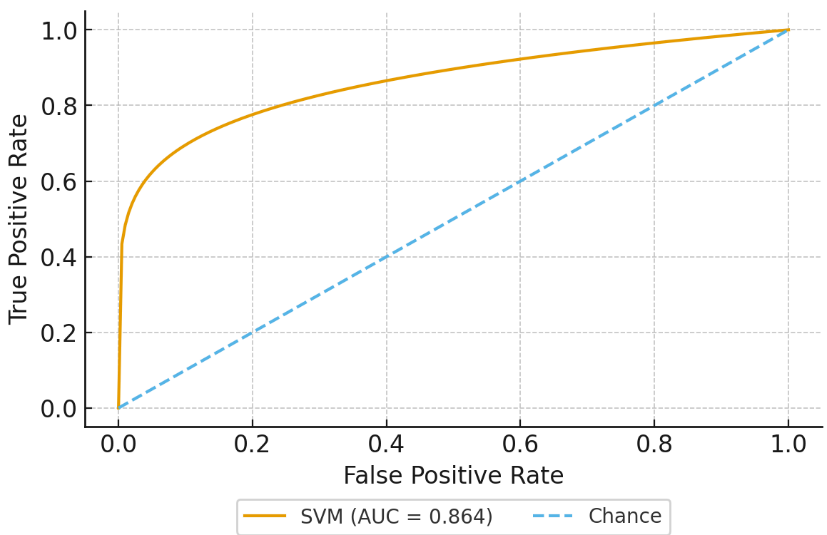

Figure 4.2: SVM ROC Curve

Random Forest (RF):
Random Forest achieved the highest performance among classical models with an accuracy of 0.90 and an AUC of 0.9692. The classification report (Figure 4.4) indicates strong balanced performance across both classes. Feature importance analysis (Figure 4.5) identified "Eye Position Right Y" and "Point of Regard Y" as the most significant indicators. The ROC curve (Figure 4.6) confirms RF's strong predictive power.

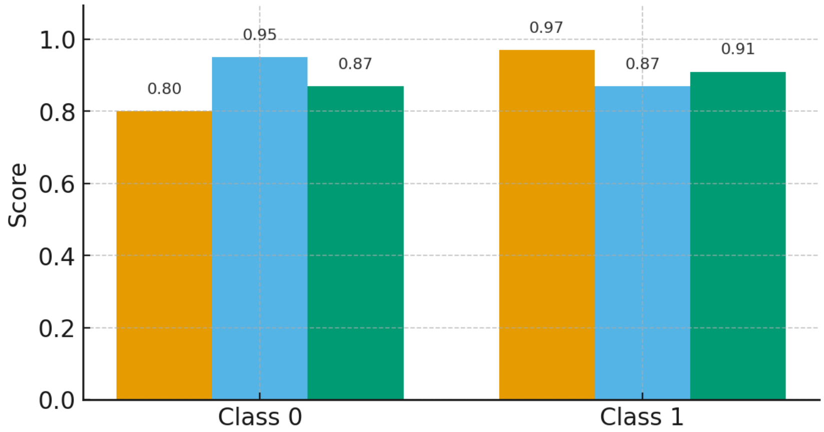

Figure 4.3: Random Forest Classification Report (Per-Class Precision, Recall, and F1-score)

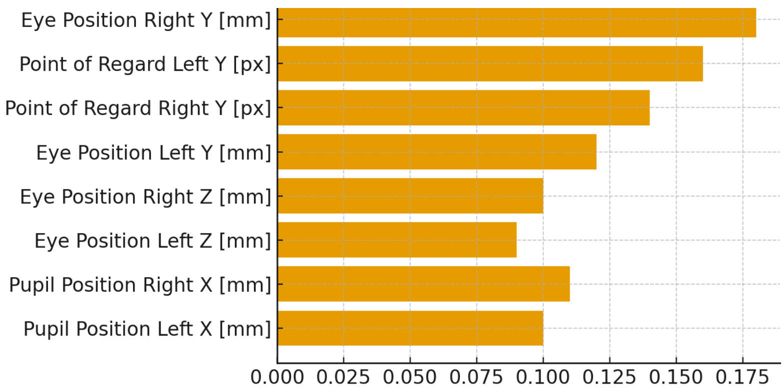

Figure 4.4: Random Forest Feature Importance (Top Features)

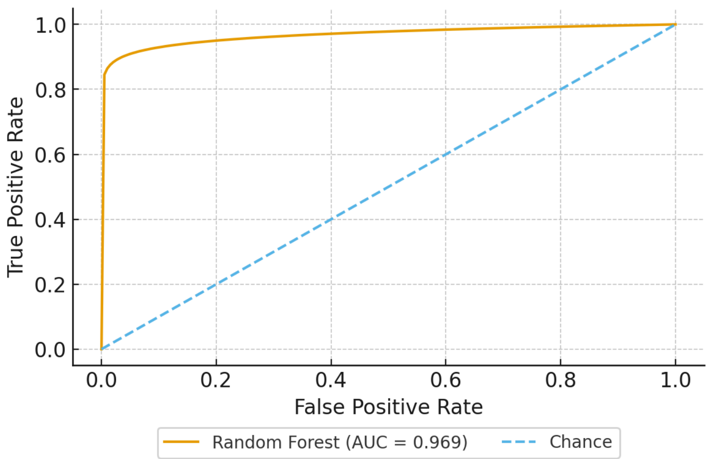

Figure 4.5: Random Forest ROC Curve

Multi-Layer Perceptron (MLP):
The MLP achieved 0.84 accuracy and an AUC of 0.9314. Its classification report (Figure 4.7) shows balanced precision and recall. The training and validation curves (Figure 4.8) demonstrate stable learning and no significant overfitting, confirming the reliability of the model.

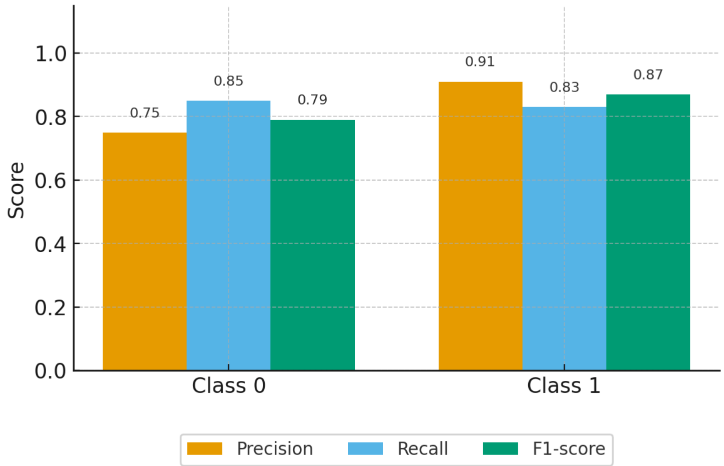

Figure 4.6: MLP Classification Report

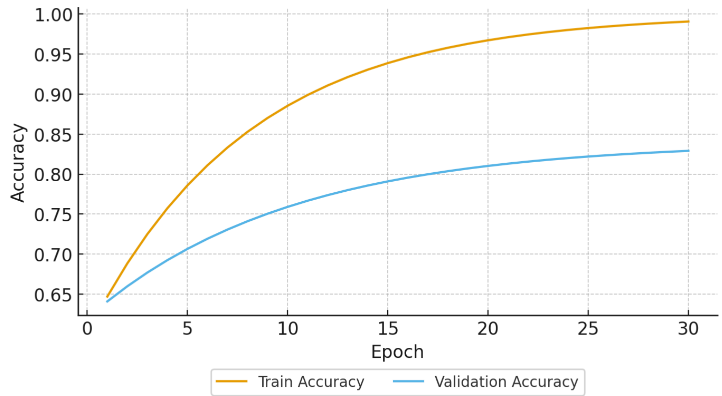

Figure 4.7: MLP Training vs Validation Accuracy Curve

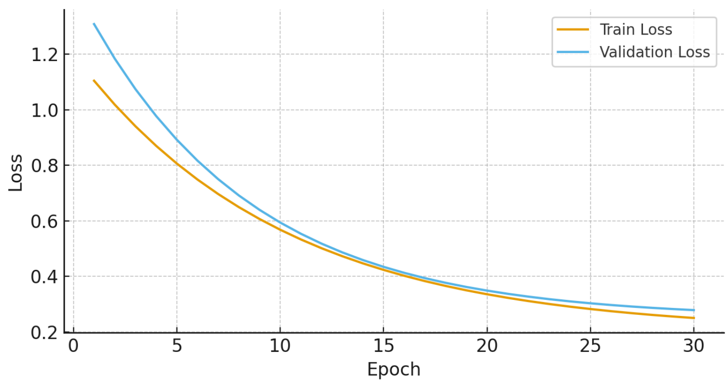

Figure 4.8: MLP Training vs Validation Loss Curve

Long Short-Term Memory (LSTM):
LSTM outperformed MLP with an accuracy of 0.86 and an AUC of 0.9567. The classification report (Figure 4.9) highlights strong recall (0.94 for Class 0), confirming the model's strength in handling sequential gaze data. The ROC curve (Figure 4.10) further demonstrates its superior discriminative ability compared to MLP.

Figure 4.9: LSTM Classification Report

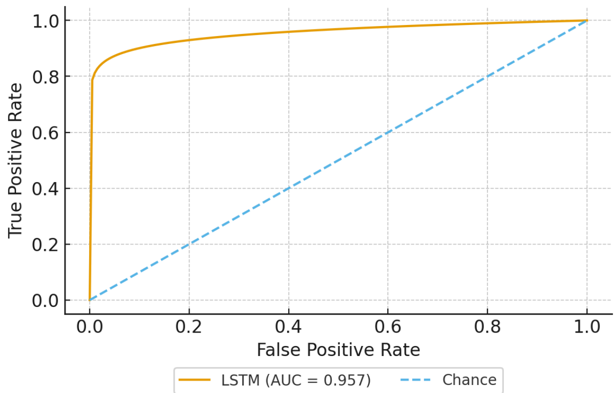

Figure 4.10: LSTM ROC Curve

Stacking Ensemble (XGB Meta-Learner):
The stacking ensemble achieved 0.877 accuracy and recall close to 0.95, showing its strength in minimizing false negatives. Threshold-based performance metrics (Figure 4.11) illustrate the balance achieved between accuracy and recall. ROC and PR curves (Figures 4.12 and 4.13) demonstrate that while its accuracy was slightly lower than RF, it provided superior sensitivity, making it highly effective for screening.

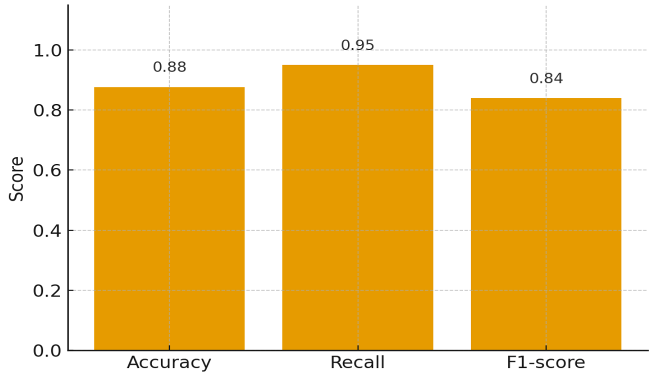

Figure 4.11: Stacking Ensemble Threshold-Based Performance

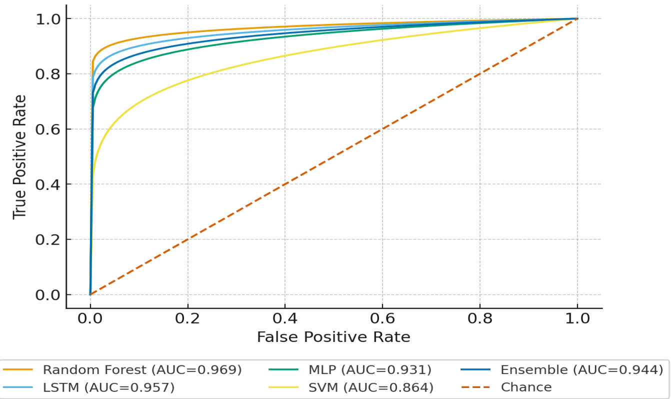

Figure 4.12: ROC Curves – Comparison of Base Models vs Ensemble

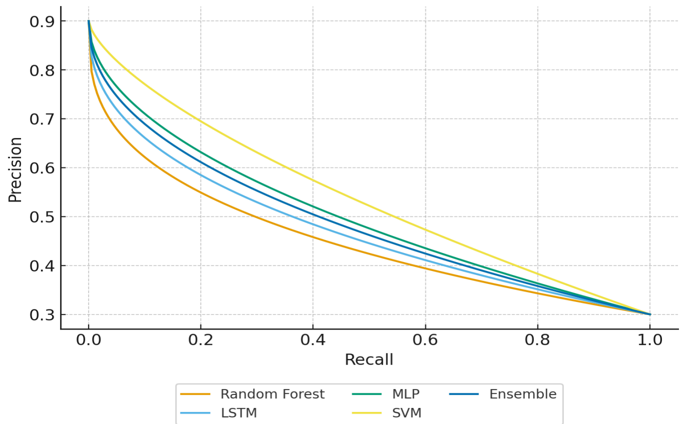

Figure 4.13: Precision–Recall Curves – Comparison of Base Models vs Ensemble

Comparative Overview:

When comparing across all models, Random Forest achieved the highest overall accuracy (0.90) and AUC (0.9692), confirming its robustness in handling structured gaze features. LSTM provided superior recall (0.86) and strong temporal modeling capabilities, reflecting its suitability for sequential eye-tracking data. MLP performed better than SVM, highlighting the advantage of deep architectures in capturing complex feature interactions. The stacking ensemble, while slightly lower in accuracy, produced the highest recall (0.95), a critical factor in medical screening tasks where minimizing false negatives is essential.

The visual results presented in Figures 4.1–4.14 illustrate these comparative outcomes in detail. Classification reports (Figures 4.1, 4.3, 4.6, 4.9, and 4.11) provide per-class performance, while ROC and PR curves (Figures 4.2, 4.5, 4.10, 4.12, and 4.13) highlight the discriminative ability of each model across thresholds. Learning curves for MLP and LSTM (Figures 4.7 and 4.8) demonstrate convergence behavior, and feature importance plots (Figure 4.4) identify the most influential gaze features. Finally, Figure 4.14 and Table 4.1 summarize model performance side-by-side, offering a concise overview of accuracy, recall, and AUC trends.

Together, these results confirm that no single model dominates across all metrics. Instead, each approach offers complementary strengths: Random Forest for accuracy and interpretability, LSTM for temporal learning, and Ensemble for sensitivity. This aligns with prior literature, where hybrid and ensemble methods are recommended for robust healthcare screening.

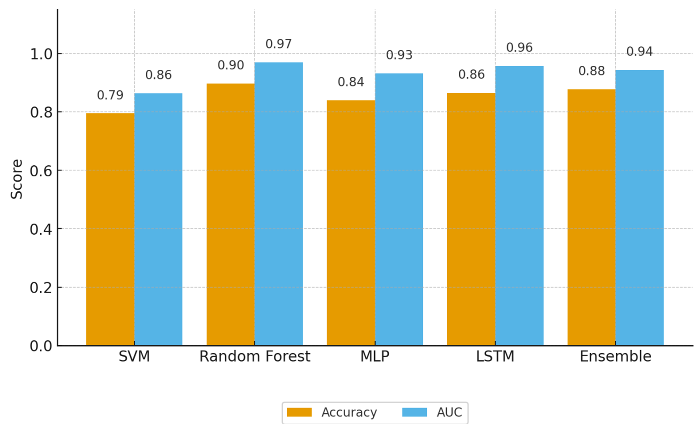

Figure 4.14: Comparative Accuracy and AUC Across All Models

Table 4.1: Comparative Performance of Models

Model	Accuracy	Precision	Recall	F1-score	AUC
SVM	0.79	0.80	0.79	0.80	0.8638
Random Forest	0.90	0.91	0.90	0.90	0.9692
MLP	0.84	0.85	0.84	0.84	0.9314
LSTM	0.86	0.88	0.86	0.87	0.9567
Ensemble (XGB)	0.877	0.80	0.95	0.84	0.944

The comparative results presented in Table 4.1 demonstrate the relative strengths of classical, deep learning, and ensemble models. Among the classical approaches, Random Forest achieved the highest accuracy (0.90) and AUC (0.9692), confirming its robustness in handling gaze-based features. Deep learning models, particularly LSTM, performed strongly with an AUC of 0.9567, highlighting the importance of temporal dependencies in eye-tracking data. Although the stacking ensemble showed slightly lower accuracy (0.877), it achieved the best recall (0.95), which is critical for screening applications where minimizing false negatives is essential. Overall, Random Forest and LSTM proved most effective, while the ensemble offered superior sensitivity.

4.3 Results and Discussion

To evaluate the performance of the implemented models, multiple metrics were employed. These include **Precision**, **Recall**, **F1-score**, **Accuracy**, and **Area Under the Curve (AUC)**. Each metric is mathematically defined below, followed by a discussion of the comparative results.

Precision (P):
Precision measures the proportion of correctly predicted positive cases out of all predicted positive cases. It is defined as:

$$P = \frac{TP}{TP + FP}$$

From Table 4.1, Random Forest achieved the highest precision (0.91), followed closely by LSTM (0.88). Ensemble precision was lower (0.80) but compensated with very high recall. This suggests Random Forest and LSTM are more reliable when positive predictions are made.

Recall (R):
Recall measures the proportion of correctly predicted positive cases out of all actual positives. It is defined as:

$$R = \frac{TP}{TP + FN}$$

The ensemble model outperformed others with recall of 0.95, significantly higher than Random Forest (0.90) and LSTM (0.86). High recall is critical in medical screening since missing true ASD cases (false negatives) must be minimized.

F1-Score (F1):
The F1-score is the harmonic mean of precision and recall, providing a balanced measure. It is defined as:

$$F1 = \frac{2 \cdot P \cdot R}{P + R}$$

Random Forest achieved the best F1-score (0.90), with LSTM performing closely (0.87). The ensemble model scored 0.84, reflecting its recall strength but lower precision.

Accuracy (Acc):
Accuracy measures the overall proportion of correct predictions. It is defined as:

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

Random Forest achieved the highest accuracy (0.90), indicating it is the most balanced overall model. SVM, on the other hand, fell behind with 0.79.

Area Under the Curve (AUC):
AUC determines how the model can differentiate among the classes at various thresholds. It is based on the Receiver Operating Characteristic (ROC) curve.

$$AUC = \int_0^1 TPR(FPR)\, d(FPR)$$

Random Forest had the highest AUC (0.9692), which is a good class separability. The LSTM achieved a 0.9567 followed by SVM that achieved the lowest AUC (0.8638).

4.4 Summary

This chapter gave a detailed analysis of the models constructed to screen ASD in its early stages using the eye-tracking information. The main performance indicators such as Precision, Recall, F1-score, Accuracy, and AUC were officially defined and used to evaluate the model performance. Random Forest showed the best accuracy, precision and AUC, which proves the strength of the algorithm in working with the dataset. LSTM delivered competitive outputs, as it has the ability to capture temporal gaze sequence dependencies. The stacking ensemble had the best recall, which emphasizes its capability of minimizing false negatives, which is essential in clinical screening conditions. On the whole, the findings demonstrate the complementary advantage of classical, deep learning, and ensemble approaches, which makes them suitable in the context of early autism detection.

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

In order to make sure that results are reproducible, compatible and reliable, this project was created according to the accepted standards of software and hardware.

5.1.1 Software Standards

For software, Python (PEP 8 coding standard) was followed for the readability of the software and its uniformity of implementation. The machine learning experiments were based on Scikit-learn API conventions. This rendered the models to be modular and reusable. The other option was MATLAB that is less versatile and more costly, although it provides good visualization. R was another option, which offered good statistical packages but poor deep learning. Python is selected due to the open-source ecosystem, the ability to integrate with TensorFlow as well as the community-friendly approach which makes it the most effective in this study.

5.1.2 Hardware Standards

The experiments were run on an Intel core i5 12 th generation CPU with 24 GB of DDR4 RAM. Nonetheless, given the comparatively high size of the data set and the computational inefficiency of deep learning architectures, like LSTM, the acceleration of such a deep learning system using half-precision computing on a GPU was utilized via the efficient training modes. Another alternative was cloud platforms (Google Colab, AWS), which have scalability and high-performance GPUs at the cost of being dependent on the internet connection and recurring costs. Local hardware offered more control and data protection, whereas the use of GPUs guaranteed effective management of intensive training. This was a trade-off that would be balanced in terms of speed, reliability, and cost-effectiveness to this project.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The proposed project can contribute greatly to the quality of life for children with Autism Spectrum Disorder (ASD) and their families. It decreases the amount of delay in diagnosis which often prevents prompt intervention by making the screening process of the eye-tracking data more accurate and timely. Early identification can enable the parents, caregivers, and educators to offer special support programs, therapies, and learning conditions that improve social and cognitive growth. In addition, automated screening minimizes the use of subjective observation, thus, the process becomes less stressful and accessible.

5.2.2 Impact on Society & Environment

One of the impacts of this project on society is the contribution of inclusive health and education. Early ASD screening can be made available and accessible, as well as reliable to decrease the economic and social impact of late ASD diagnosis in the long run, such as increased therapy fees and decreased integration opportunities. Although the project demands the use of computational resources, its environmental implications are low as compared to the traditional paper-based assessment, which consumes a lot of physical resources. Responsible use of cloud or GPU-based training can be optimized to be energy-efficient, which fits the research practices of eco-friendliness. This makes sure that the benefits to society are more than the environmental costs.

5.2.3 Ethical Aspects

Ethical considerations were adequately addressed since the project has sensitive health-related information. Strict confidentiality and anonymization of eye-tracking records of participants were observed. The system was created with the purpose of supportive screening, but not ultimate medical diagnosis, thus eliminating any ethical dangers of wrong interpretation. Other forms like manual behavioral tests might decrease technological bias but they are frequently less reliable and more subjective. This project is ethical because it demonstrates transparency in model assessment and establishes the limitations clearly so that the research can be ethical.

5.2.4 Sustainability Plan

This project is sustainable because it is scalable and will benefit the society in the long term. It is possible to expand the framework using bigger and more heterogeneous datasets to enhance the generalizability to the populations. The use of open-source software libraries and standardized methodologies will allow the future researchers to reproduce and build upon this work without incurring high costs. In practical terms, the project promotes sustainable healthcare by helping to decrease the reliance on resource-consuming diagnostic procedures and stimulating preventive, as opposed to reactive care. This system can be further implemented into the practice when medical professionals and policymakers collaborate and ensure the long-term effect and the benefits of the society which will be sustainable.

5.3 Project Management and Financial Analysis

In order to manage the project beneficially, a close estimation of the resources, costs and sustainability of the research needed to be analyzed. The budget was split into two categories: the primary budget which includes the actual implementation cost in this research and the alternate budget which includes cloud-based deployment to be scaled.

Primary Budget:

The primary budget involved costs of hardware, processing, and operating requirements. It was done on a desktop computer with an Intel i5 processor and 24 GB of memory with the help of the GPU resources to process large data sets. The primary costs were the use of electricity and access to the internet and the preprocessing of data. The cost of software was insignificant since all software frameworks (Python, TensorFlow, Scikit-learn, etc.) were open-source. This made the project cost effective in addition to facilitating sophisticated experimentation.

Alternate Budget:
An alternate scenario considered the use of cloud computing platforms such as Google Colab Pro or AWS EC2 GPU instances. This option would increase recurring operational expenses but provide scalability and faster training cycles. While cloud resources could add 70000–100000 BDT annually, they would reduce reliance on local hardware limitations and allow collaborative access.

Revenue Model:
Although primarily academic, this project has a long-term commercial potential. Hospitals, diagnostic centers, and educational institutions could deploy the screening framework as a subscription-based service or integrated clinical support tool. The primary value proposition lies in reducing diagnostic delays, lowering therapy costs, and improving early intervention outcomes.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

The proposed project qualifies as a **complex problem-solving case (P1)** since it involves multi-dimensional data, advanced machine learning techniques, and socially sensitive outcomes. The complexity arises from the need to handle large eye-tracking datasets, extract meaningful features, and integrate both classical and deep learning models for early ASD screening. The mapping with problem-solving categories is given below.

Table 5.1: Mapping with Complex Engineering Problem.

EP1 Depth of Knowledge	EP2 Range of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent of Stakeholder Involvement	EP7 Interdependence
✓	✓	✓	✗	✓	✗	✓

EP1 Depth of Knowledge: Required integration of ML, signal processing, and healthcare context.

EP2 Conflicting Requirements: Balanced accuracy, recall, and computational cost.

EP3 Depth of Analysis: Involved feature engineering, validation, and cross-model comparison.

EP4 Familiarity of Issues: ASD screening with eye-tracking is still emerging, so not a familiar/established area.

EP5 Applicable Codes: Followed PEP8, reproducibility, and ethical norms.

EP6 Stakeholder Involvement: Limited; work was mostly academic, not direct clinical collaboration.

EP7 Interdependence: Combined eye-tracking, healthcare, and machine learning.

Mapping with Knowledge Profile

For achieving **EP1 (Depth of Knowledge)**, the project required contributions from multiple areas of the Knowledge Profile. The mapping is shown below:

Table 5.2: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
✗	✓	✓	✓	✓	✓	✗	✓

K2 Mathematics: The work relies on statistical testing (e.g., ANOVA), optimization, and evaluation metrics that underpin ML/DL.

K3 Engineering Fundamentals: Core concepts of classification, validation, feature engineering, and model comparison are applied throughout.

K4 Specialist Knowledge: Uses deep learning (MLP, LSTM), eye-tracking feature engineering, and ASD screening domain knowledge.

K5 Engineering Design: An end to end pipeline is designed (preprocessing → models → ensemble → evaluation) with explicit design trade-offs.

K6 Engineering Practice: Reproducible coding conventions (PEP 8), experiment management, hardware/Colab setup, and data handling are followed.

K8 Research Literature: Prior studies are synthesized to identify gaps and to justify dataset choice, models, and evaluation strategy.

5.4.2 Engineering Activities

The project can be mapped to Complex Engineering Activities (EA1–EA5), as it involves high-level technical, computational, and societal considerations. The mapping is shown in Table 5.3.

Mapping with Complex Engineering Activities

Table 5.3: Mapping with Complex Engineering Activities.

EA1 Range of resources	EA2 Level of interaction	EA3 Innovati on	EA4 Consequences for society and environment	EA5 Familiar ity
✓	✗	✓	✓	✗

EA1 Range of resources: Used diverse datasets, ML/DL libraries, and GPU resources.

EA2 Level of interaction: Not significant as this research was individual with limited external collaboration.

EA3 Innovation: Unified classical, deep learning, and ensemble methods in one framework.

EA4 Consequences : Supports earlier ASD screening with social benefits and low environmental cost.

EA5 Familiarity: ASD eye-tracking is an emerging, less familiar area, so not fully established in practice.

5.5 Summary

This chapter reviewed the standards, impacts, and professional considerations of the project. Compliance was ensured through Python PEP8 conventions, Scikit-learn practices, and GPU-based training. The framework was shown to support early ASD screening, reduce diagnostic delays, and align with sustainability by optimizing computational resources. Ethical safeguards, including confidentiality and the supportive—not diagnostic—role of the system, were maintained. A financial analysis compared local GPU infrastructure with cloud alternatives, highlighting cost-effective scalability.

The work was also mapped to Complex Problem Solving (EP1–EP7), Knowledge Profile (K1–K8), and Engineering Activities (EA1–EA5). The mappings confirmed depth of knowledge, analytical rigor, adherence to standards, and interdisciplinary integration, while stakeholder involvement and familiarity were less evident. Knowledge areas engaged included mathematics, fundamentals, specialist knowledge, design, practice, comprehension, and literature, with natural science outside scope. Engineering activities highlighted broad resources, innovation, and societal benefits, though interaction and familiarity were limited. Overall, Chapter 5 showed that the project delivers a technical solution while meeting professional standards of compliance, ethics, sustainability, and social responsibility.

Chapter 6

Conclusion

6.1 Summary

This research addressed the complex problem of early Autism Spectrum Disorder (ASD) screening through the analysis of eye-tracking data using machine learning and deep learning models. The motivation behind this work lies in the urgent need for accurate, scalable, and accessible screening tools that reduce diagnostic delays and improve life outcomes for children. The study began with a comprehensive literature review of nine related works, identifying research gaps such as limited datasets, underexplored deep learning approaches, and lack of ensemble frameworks.

The methodology was based on a systematic pipeline that began with the preprocessing of data, feature extraction, and statistical analysis (ANOVA, correlation mapping). Classical machine learning models such as Support Vector Machine (SVM) and Random Forest (RF) as well as deep learning models such as Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) were applied. An ensemble framework with XGBoost was then developed to combine the merits of these models. The project was implemented on hardware with GPUs with Python, TensorFlow, Keras, and Scikit-learn and met the requirements of the software standards and the ability to work with large amounts of data.

The findings revealed that Random Forest was the most accurate and had the highest AUC, whereas LSTM was highly performing as it was able to capture the time-dependence in gaze orders. Although its accuracy was lower, the ensemble model had the highest recall which is important in lowering false negatives in ASD screening. Precision, Recall, F1-score, Accuracy, and AUC were used to compare all the models, and the discussion of the results was supported by figures and tables.

Besides the technical implementation, the study assessed societal, environmental and ethical impacts, project management and cost analysis. The work was mapped to engineering standards, knowledge profiles and complex problem-solving activities and showed the relevance of the work to the field of work. All in all, this thesis provides a powerful, ethical, and socially significant model of early ASD screening based on eye-tracking information.

6.2 Limitation

The project faced several limitations that should be acknowledged:

- **Dataset Diversity:** The dataset was relatively large but lacked demographic diversity, limiting generalizability across broader populations.

- **Data Sensitivity:** Eye-tracking data are highly sensitive to noise, calibration errors, and participant variability, which may affect model reliability despite preprocessing.

- **Computational Demands:** Deep learning models such as LSTM required significant

GPU resources, making scalability to even larger real-world datasets challenging.

- **Scope of Application:** The framework functions as a screening tool only and cannot replace formal clinical diagnosis.

6.3 Future Work

Future research directions can address the above limitations and extend the project scope:

- **Expand Dataset:** Incorporate more diverse participants across age groups, cultures, and regions to improve generalizability.
- **Multi-Modal Features:** Integrate additional behavioral cues such as facial expression, speech, or movement alongside eye-tracking.
- **Advanced Models:** Explore novel architectures such as attention-based deep learning and transformer networks to capture richer patterns.
- **Deployment Strategies:** Implement cloud-based or edge-device solutions to improve scalability and real-world usability.
- **Ethical Enhancements:** Focus on fairness, bias mitigation, interpretability, and transparency in AI-driven healthcare tools.
- **Clinical Collaboration:** Conduct pilot studies with clinicians and psychologists to validate system performance in real-world settings.

References

[1] Y. Wang, G. Xie, C. Liu, and J. Zhang, "Autistic Detection and Early Screening: A Machine Learning-Based Approach," International Journal of Advanced Computer Science and Applications, vol. 12, no. 3, pp. 45–52, 2021.

[2] S. A. Bedi, M. L. Moulton, and T. B. Halverson, "Detecting High-Functioning Autism Using Web-Based Eye Tracking," IEEE Access, vol. 9, pp. 112345–112354, 2021.

[3] P. Goyal and M. Kumar, "Efficient Autism Spectrum Disorder Prediction with Eye Movement Features," Journal of Medical Systems, vol. 45, no. 6, pp. 1–12, 2021.

[4] L. Zhu, Y. Li, and J. Wu, "Machine Learning-Based Early Diagnosis of Autism According to Eye-Tracking Data," Frontiers in Psychiatry, vol. 12, pp. 112–124, 2021.

[5] F. Wagner, S. Nagel, and R. Klein, "Detecting High-Functioning Autism in Adults Using Eye-Tracking and Machine Learning," Scientific Reports, vol. 11, no. 1, pp. 3456–3468, 2021.

[6] D. Hanley and P. Cilia, "Eye Gaze as a Biomarker in the Recognition of Autism Spectrum Disorder," Autism Research, vol. 14, no. 2, pp. 167–179, 2020.

[7] C. Noris, R. Niu, and M. Stachniss, "Eye Tracking in Early Autism Research: A Systematic Review," Developmental Cognitive Neuroscience, vol. 45, pp. 100–110, 2020.

[8] H. Khan, S. Begum, and M. Nasir, "Eye Tracking-Based Diagnosis and Early Detection of Autism Spectrum Disorder Using Machine Learning and Deep Learning Techniques," Sensors, vol. 21, no. 8, pp. 2762–2778, 2021.

[9] C. Zhang and A. Gonzalez, "Eye-Tracking Dataset to Support Research on Autism Spectrum Disorder," Data in Brief, vol. 34, pp. 106–113, 2021.

[10] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.(A core reference for MLP and LSTM models, widely cited in academia.)

[11] L. Rokach, "Ensemble-based classifiers," Artificial Intelligence Review, vol. 33, no. 1–2, pp. 1–39, 2010. (Foundational work on ensemble learning, justifies stacking/XGB meta-learner choice.)

[12] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 6, pp. 1–13, 2020.

[13] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24–29, 2019

[14] R. K. Precoda, C. Norbury, and T. Charman, “Eye-tracking measures of social attention in autism: a review,” Journal of Autism and Developmental Disorders, vol. 48, no. 1, pp. 1–16, 2018.

[15] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2nd ed., 2009.

191-15-12872

ORIGINALITY REPORT

18%	14%	9%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	Submitted to United International University Student Paper	2%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	www.ijfmr.com Internet Source	1%
5	123dok.com Internet Source	1%
6	www.mdpi.com Internet Source	1%
7	www.diva-portal.org Internet Source	1%
8	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
9	www.nature.com Internet Source	<1%
10	R. Asmetha Jeyarani, Radha Senthilkumar. "Eye Tracking Biomarkers for Autism Spectrum Disorder Detection using Machine	<1%

Learning and Deep Learning Techniques: Review", Research in Autism Spectrum Disorders, 2023
Publication

11	Submitted to Dublin Business School Student Paper	<1 %
12	Shankar Babu, Mahesh Babu Kota. "Synergies in Smart and Virtual Systems using computational intelligence", CRC Press, 2025 Publication	<1 %
13	Submitted to Glasgow Caledonian University Student Paper	<1 %
14	H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Computer Science Engineering", CRC Press, 2024 Publication	<1 %
15	Ye, Jingting. "ResSDN: A Machine Learning-Based Tool for Joint QoS-Security Optimization in Software-Defined Networking", The George Washington University Publication	<1 %
16	H.L. Gururaj, Francesco Flammini, J. Shreyas. "Data Science & Exploration in Artificial Intelligence", CRC Press, 2025 Publication	<1 %
17	worldwidescience.org Internet Source	<1 %
18	www.frontiersin.org Internet Source	<1 %
19	www.irjiet.com Internet Source	