Autism Spectrum Disorder Classification Based on ML and DL Techniques

Nabaa Jabar Darwish, Ammar I. Shihab[*]
Department of Computer Science, College of Science, University of Baghdad

Abstract

This study presents a novel approach to classifying Autism Spectrum Disorder (ASD) using eye-tracking data and advanced machine-learning techniques. The study employed a comprehensive dataset of eye movements from 28 children, 14 with ASD and 14 typically developing (TD) controls, collected while viewing 300 natural scene images. This methodology encompassed traditional machine learning algorithms (Logistic Regression, Random Forest, and Gradient Boosting) and deep learning models (Convolutional Neural Network and a hybrid CNN-LSTM architecture). The hybrid CNN-LSTM model achieved the highest accuracy of 99.89%, outperforming other approaches and comparable studies in literature. Notably, even our simpler Logistic Regression model attained 99.37% accuracy, demonstrating the robust discriminative power of eye-tracking data for ASD classification. Our results suggest that integrating spatial and temporal patterns in eye movements significantly enhances ASD detection accuracy. This study contributes to the growing body of research on automated, data-driven approaches to ASD diagnosis, potentially facilitating earlier and more accurate identification of ASD in clinical settings.

Keywords: Autism Spectrum Disorder (ASD), Eye-tracking, Machine Learning, Deep Learning, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Classification.

تصنيف اضطراب طيف التوحد استنادًا إلى تقنيات التعلم الآلي والتعلم العميق

نبأ جبار ، عمار ابراهيم شهاب*
قسم الحاسبات، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصه

تقدم هذه الدراسة نهجًا جديدًا لتصنيف اضطراب طيف التوحد (ASD) باستخدام بيانات تتبع العين وتقنيات التعلم الآلي المتقدمة. تم استخدام مجموعة بيانات شاملة لحركات العين من 28 طفلًا (14 مصابًا باضطراب طيف التوحد و14 من الضوابط الذين يعانون من اضطراب طيف التوحد عادةً)، تم جمعها أثناء عرض 300 صورة لمشهد طبيعي. شملت منهجيتنا كل من خوارزميات التعلم الآلي التقليدية (الانحدار اللوجستي والغابة العشوائية وتعزيز التدرج) ونماذج التعلم العميق) الشبكة العصبية التلافيفية وهندسة-CNN LSTMالهجينة .(حقق نموذج CNN-LSTM الهجين أعلى دقة بنسبة 99.89%، متفوقًا على الأساليب الأخرى والدراسات المماثلة في الأدبيات. والجدير بالذكر أن حتى نموذج الانحدار اللوجستي الأبسط لدينا حقق دقة بنسبة 99.37%، مما يدل على القوة التمييزية القوية لبيانات تتبع العين لتصنيف اضطراب طيف التوحد. تشير نتائجنا إلى أن دمج الأنماط المكانية والزمانية في حركات العين يعزز بشكل كبير من دقة

*Email: ammar.a@sc.uobaghdad.edu.iq

اكتشاف اضطراب طيف التوحد. تساهم هذه الدراسة في مجموعة متنامية من الأبحاث حول الأساليب الآلية

القائمة على البيانات لتشخيص اضطراب طيف التوحد، مما قد يسهل التعرف المبكر والأكثر دقة على

اضطراب طيف التوحد في البيئات السريرية.

1. Introduction

Background Autism Spectrum Disorder (ASD) is a complex, heterogeneous neurodevelopmental disorder that affects social communication and the presence of repetitive behaviors with restricted interests. In the United States, in 2024, about 1 in every 33 children are diagnosed with ASD (Children and Teens, Centers for Disease Control and Prevention), demonstrating an urgent need to improve diagnostic and intervention approaches. Early diagnosis and intervention are essential for achieving desired developmental outcomes and improving overall quality of life [1]. Nevertheless, conventional diagnostic approaches can be slow and rely on expertise that is not available in all settings, which also may introduce subjectivity into the diagnoses, leading to possible delays. The majority of those with ASD have a combination of symptoms and severity, which are characteristics influenced by other neurological factors and environmental influences [2]. Progress in diagnostic technologies and analytic approaches is core to our understanding of the behavioral and cognitive underpinnings of ASD. Figure (1) shows the Disorder ASD pervasiveness rates from 2012 to 2024 [3].
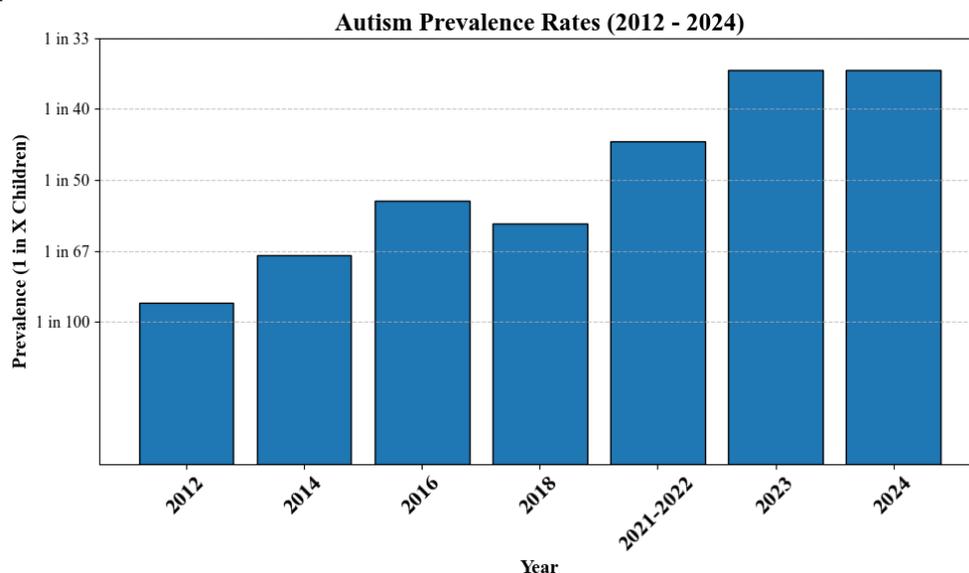

Figure 1 : The Autism Spectrum Disorder (ASD) prevalence rates from 2012 to 2024 [3].

Recent interventions in artificial intelligence, particularly machine learning and deep learning, have demonstrated great potential for detecting autism spectrum disorders across a range of healthcare applications. AI technology has the potential to provide more objective, reliable, and efficient diagnostic tools than traditional clinical judgment alone while reducing the requirement for human subjective assessment and enhancing early detection and intervention [4].

Using eye-tracking technology to examine social and attentional patterns in individuals with autism is a promising area of research in this regard. Eye-tracking data provides a non-invasive and objective view of visual attention, which is known to be influenced by neural processes. These studies have provided evidence that individuals with autism exhibit unique eye-movement patterns compared to individuals without autism, and these unique eye-movement patterns could be analyzed for early detection of children with autism [5].

This study presents a comprehensive evaluation framework for ASD classification using eye-tracking data, combining traditional machine learning and advanced deep learning approaches. Our key contributions are:

1. Development of a systematic data preparation pipeline that transforms raw eye-tracking data through:

- Converting text-based eye-tracking records to structured CSV format
- Aggregating and merging ASD and TD datasets
- Implementing preprocessing steps, including data cleaning, one-hot encoding, and feature scaling
- Establishing an 80-20 train-test split for robust model validation

2. Development of a systematic comparison between classical ML algorithms (logistic regression, gradient boosting, and random forest classifiers) and sophisticated DL architectures (CNN and hybrid CNN-LSTM models) for ASD detection [6].

3. Implementation of a robust evaluation methodology using multiple performance metrics (accuracy, precision, recall, and F1-Score) to provide a thorough assessment of each model's diagnostic capabilities.

4. Creation of a comparative analysis framework that identifies the most effective approach for ASD classification through eye-tracking data, establishing a foundation for developing reliable AI-based diagnostic tools.

2. Literature Review

Autism Spectrum Disorder (ASD) is a complex neuro-biological condition associated with different range of symptoms and degrees. There has been significant research over the last several decades that have been motivated to understand ASD better, how it impacts children and adults with its diagnosis, and ways in which clinicians can help manage these patients. This section presents a review of key findings that have shaped the current understanding of ASD.

- Romuald et al [7]. Focused on 8-to10-year-old children categorized into autism and typical classes. Eye-tracking data, including saccadic movement information, was collected while the children watched prerecorded movies of joint attention offerings. The data is split into 75% training and 25% test sets. The neural network design consists of two LSTM hidden layers with 20 nodes each, 7 input nodes, and 3 output nodes (class 1, class 2, and undecided). The network achieved an average of 90% confidence, with the best result at 98% (fitness 0.006), while Sensitivity was 75%, and specificity was 1.

- Ibrahim Abdulrab Ahmed et al [8]. Focused on effective learning and early autism diagnosis using eye-tracking and artificial intelligence (AI). They employed three methods: the first utilized neural networks (FFNNs and ANNs) with a hybrid feature classification technique, achieving 99.8% accuracy. The second method involved deep feature map extraction using pre-trained CNN models (GoogleNet: 93.6% accuracy, ResNet-18: 97.6% accuracy). The third approach combined GoogleNet/ResNet-18 with SVM, achieving excellent results. Generally, their paper showed the potential of AI-based eye-tracking for accurately diagnosing autism in children, paving the way for improved learning strategies and early intervention.

- Fawaz Waselallah Alsaade et al [9]. Developed a novel approach to identifying distinctive facial traits in children with the disorder using social media data and biomedical pictures. The system presented in the proposed research is based on DL and uses a flask framework, a convolutional neural network, and transfer learning. 2,940 face photos from Kaggle were classified using three pre-trained models (Xception, VGG19, and NASNETMobile). Utilizing evaluation parameters including accuracy, specificity, and sensitivity, Xception outperformed VGG19 (80%) and NASNETMobile (78%), with

Xception earning the greatest accuracy at 91%. Through the analysis of face features, this promising study may assist communities and psychiatrists in the early detection of ASD.

• Praveena K N et al [10]. Based on the observer's fixation maps, this study seeks to differentiate between usual developments (TD) and (ASD).300 training photos were collected from 14 TD and 14 ASD children. 14 fixation maps are not present in all photos. The suggested model was trained using images 1 through 110. Images 211–300 were utilized to test and validate the model. From 300 ASD and 300 TD fixation maps, scaled down to 224x224 and transformed to grayscale, features were extracted using a CNN architecture. On the validation dataset, the suggested model achieved 75.23% accuracy.

• Mujeeb Rahman Kanhirakadavath [11]. Eye-tracking data and machine learning were used to screen autism with high accuracy. The process involved image preprocessing, converting RGB to grayscale, and scaling images. Four different ML techniques were tried, including three conventional methods and one DNN-based approach using the original ETSP Image Dataset and an augmented dataset. The DNN-based model achieved a cross-validated AUC of 78% on the initial dataset and 97% on the enhanced dataset. With just 59 trials in the mini-batch dataset, the DNN model achieved 72.88% accuracy. The results suggest that the proposed DNN model and ETSP images can aid in a quick, precise, and reliable diagnosis of ASD when correlated with other screening tools.

• Amrutha S M et al [12]. Enhanced the autism diagnosis by automating it through classification techniques. Using the datasets collected from online kaggle.com. Preprocessing begins with removing the record containing the missing value. The second step is to eliminate attributes that do not have benefits in classification. Represents the total number of autism disorder patients and normal patients in dataset. There are 189 normal patient records and 515 autism patient records. They use different Machine Learning Algorithms such as Logistic Regression(LR) Algorithm, Naïve Bayes(NB) Algorithm, Decision Tree(DT) Algorithm, K-Nearest Neighbor(KNN) Algorithms to predict Autism Spectrum Disorder. They found that the Decision Tree achieved higher accuracy (100%), and KNN (97.73%), Naive Bayes (96.59%), and Logistic Regression (99.37%) Natural Language Processing (NLP) were used. Simulation results showed that the Naive Bayes classifier, KNN, Logistic Regression, and Decision tree give the best results in terms of prediction accuracy and execution time.

• Mohammad-Parsa Hosseini et al [13]. Proposed a deep learning approach using the MobileNet architecture coupled with two dense layers for feature extraction and image classification. The study utilized a dataset from Kaggle consisting of 3,014 images, partitioned into training, validation, and test subsets prior to model training. The introduced deep convolutional neural network architecture attained a high classification accuracy of 94.6% in differentiating between typically developing children and those potentially exhibiting autism spectrum disorder. The dataset comprised an equal number of images of children with and without ASD, ensuring balanced class representation. Through this work, the authors demonstrate the capability of deep learning techniques for automated image-based screening and diagnosis of neurodevelopmental disorders from photographic data. Further validation on expanded real-world datasets would be valuable future work to translate these methods into clinical practice.

• Suman Raj et al [14]. Presented various machine learning methods, including Naïve Bayes, Support Vector Machine, Logistic Regression, K-Nearest Neighbors, Neural Networks, and Convolutional Neural Networks, for Autism Spectrum Disorder (ASD) prediction across age groups. Three non-clinical ASD datasets were rigorously evaluated: the child, adult, and adolescent datasets. Convolutional Neural Networks (CNN) achieved the highest accuracy rates of 99.53%, 98.30%, and 96.88% for ASD screening in adults, children, and adolescents, respectively, after meticulous application of techniques and handling missing values.

- Kaushik Vakadkar et al [15]. The study presents various types of machine learning techniques and uses large and open source ASD datasets to build an accurate model that contains categorical, continuous, and binary attributes. Initially, the dataset had 1054 instances along with 18 attributes (including class variables). The researchers started by preprocessing the dataset to remove noise and eliminate missing values and outliers. And applied various ML techniques - Support Vector Machines, Random Forest Classifier, Naïve Bayes, Logistic Regression, and K-Nearest Neighbors. Among the assessed models, Logistic Regression (LR) manifests the utmost accuracy within the dataset at 97.15%, and the (RFC) achieved the least accuracy with 81.52%.

- Md Delowar Hossain et al [16]. The research aimed to enhance autism diagnosis by automating it through classification techniques. Using ASD datasets of Toddlers, Child, Adolescents, and Adults, Preprocessing begins with removing the record containing the missing value. The second step is to eliminate attributes that do not have benefits in classification. Firstly, they build the machine learning training model. They used k-fold cross-validation to train the dataset, and for evaluation purposes, they applied 25 classification techniques. They found that 4 out of 25 classifiers, such as SMO, Logistic Regression, Multi Class Classifier, and Multilayer perceptron (MLP), outperform others in terms of accuracy. SMO SVM offers higher accuracy and lower errors in ASD detection.

- Talib et al [17]. Developed a fuzzy decision-making framework for prioritizing autism patients with moderate emergency levels. Using a dataset of 538 autism patients, they focused on 432 moderate cases. They implemented a multi-phase framework combining the FWZIC method for criteria weighting and the VIKOR method for patient ranking. The study evaluated 19 criteria, with verbal communication receiving the highest weight (0.07298), followed by laughing for no reason (0.06666), nodding (0.06528), and patient movement at home (0.06330). Four sensitivity analysis scenarios were conducted using different weight allocations (60%:40%, 40%:60%, 70%:30%, and 30%:70%) to validate the framework's effectiveness. The integrated approach successfully prioritized patients based on emergency levels, though the authors recommended further validation on diverse datasets and extension to other severity levels. The framework demonstrated practical utility for healthcare professionals in optimizing resource allocation and intervention timing for ASD patients.

3. Dataset

This study uses a unique dataset to investigate the visual attention characteristics in children with Autism Spectrum Disorder (ASD) relative to typically developing (TD) children [18]. This open-access resource seeks to aid research in investigating and modeling visual attention in ASD. This includes 300 photos of natural scenes, chosen from a wide range to provide varied visual stimuli. In this study, 28 children (14 with ASD and the remaining typically developing controls) were selected to represent a sample based on specific diagnostic criteria. The dataset contains fixation maps and scan paths, revealing how long participants looked at each piece of the images. The eye tracker that was used in this research study is the Tobii T120, which provides information about the location and duration of a fixation as well as pathing data (scan paths) where participants browsed through selected scene photos. The sequence and duration of each fixation are available, allowing for a very detailed temporal inspection of visual attention. This dataset was created under the hypothesis that an ASD-specific visual attention model could be developed to assess gaze behavior in children with TD and ASD. Insights derived from eye-tracking data can help in building diagnostic tools that automatically detect an individual as ASD based on their gaze patterns and pave the way for scalable early diagnosis and intervention, which has four columns: index (idx), x-coordinate (x), y-coordinate (y) and duration the data is described as follows:

- **Index (idx):** This is a unique sequential number for each data point, the first one starting at 0.

- **X-coordinate (x):** Position on the horizontal scale of the fixation point in pixels.
- **Y-coordinate (y):** The fixation point is placed on the screen in pixels based on its vertical position.
- **Duration:** Represents in milliseconds the time duration for which the participant kept his gaze fixed at position (x, y)

3. The Proposed System

ASD classification based on eye-tracking data is proposed using a multistage approach. The raw text dataset is first converted into CSV format and aggregated by merging ASD and TD data. It includes data cleaning, one-hot encoding for categorical variables, and feature scaling in the preprocessing stage. Both machine learning and deep learning are used. The training and testing datasets are then used to train and evaluate the conventional machine learning models, which include Random Forest, Gradient Boosting, and Logistic Regression. In contrast, the deep learning approach involves feature transformation into a format that can be fed into convolutional neural networks and hybrid CNN-LSTM configurations. These deep learning models are further trained using techniques to avoid overfitting, such as early stopping and reducing the learning rate. The system assesses all models by utilizing metrics including accuracy, precision, recall, and F1-score, in addition to examining learning curves and feature significance. These will classify the data into ASD or TD. This shows the robustness of the system in handling diverse datasets and methodologies.
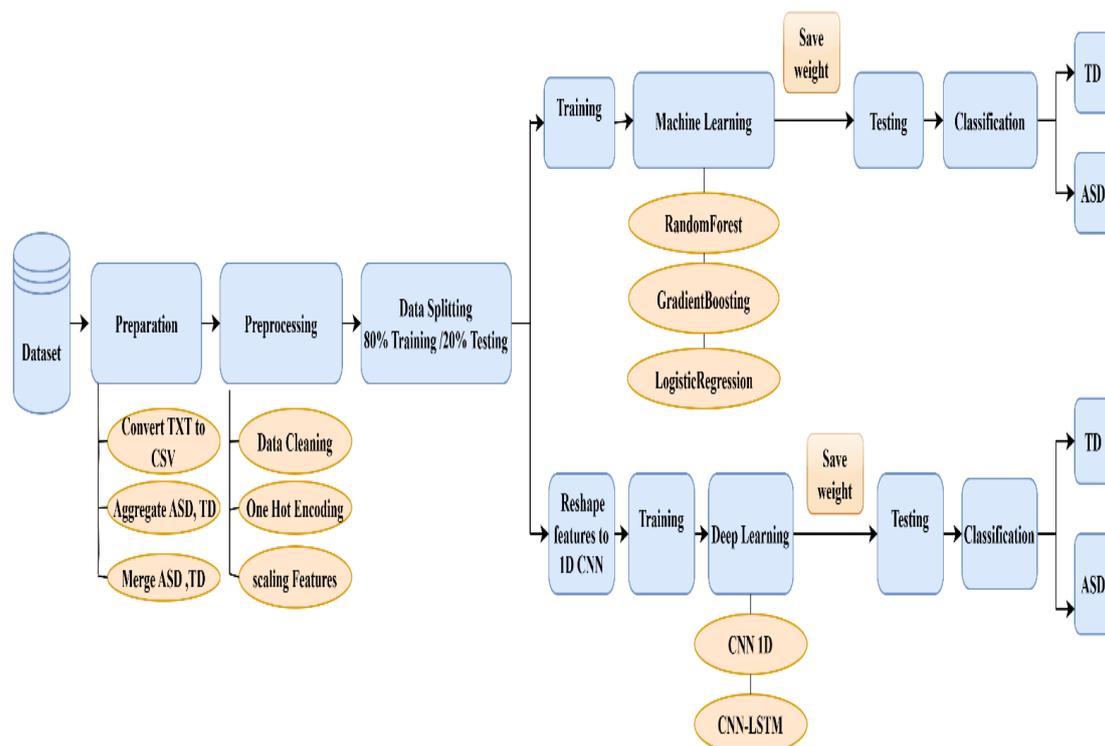
Figure 2: the proposed system.

3.1 Data Preparation

Training the proposed system on raw data, "A Dataset of Eye Movements for Children with Autism Spectrum Disorder," which is available online on Zenodo, requires a lot of data manipulation and preparation. This dataset comprises eye movement data from children with autism spectrum disorder (ASD) and typically developing (TD) controls, collected while viewing 300 landscape photographs. The Tobii T120 eye tracker data contain fixation maps and scan paths containing important information such as the index (idx), fixation length, and

gaze point coordinates (x, y). Several steps were used to prepare the data before preprocessing and training:

3.1.1 Data Organization
The scan paths are separated into two main folders. The first folder contains 300 scan paths from children with ASD, stored as a text document, and the second folder contains 300 scan paths from children with ASD, also stored as a text document. Each scan path file includes eye-tracking information such as index, x, y coordinates, and fixation duration.

3.1.2 Data Format Conversion
All files were converted from text documents to CSV format for easier data processing. This conversion makes data processing and analysis easier in subsequent steps.

3.1.3 Data Merging
Data merging occurred in two stages: First, all TD scan paths are combined into one CSV file, and tags and file names are added to maintain data integrity and traceability. In the second step, all ASD scan paths were merged into another sheet, including class and file names.

3.1.4 Final Dataset Creation
The final step involves collecting all the eye-tracking data into a unified comprehensive dataset of TD and ASD. This unified dataset is used to analyze the system and train the proposed model. This data preparation process aims to transform the raw eye tracking data into a structured and analyzable format. The resulting dataset contains rich eye movement patterns from children with ASD and ASD, allowing for intelligent analysis of visual attention differences and the design of potential diagnostic tools.

filename	Idx	x	y	duration	label
0	TD_scanpath_1	0	349	541	649	TD
1	TD_scanpath_1	1	221	370	649	TD
2	TD_scanpath_1	2	214	439	491	TD
3	TD_scanpath_1	3	365	385	441	TD
4	TD_scanpath_1	4	247	455	217	TD

```
Idx, x, y, duration
0,379,808,500
1,123,808,500
2,369,407,175
3,397,286,416
4,571,622,208
5,430,277,316
6,429,203,400
7,387,288,450
8,497,721,300
```

Figure 3: A screenshot of the dataset collected from Tobii T120 before and after preparation.

Began with a complete dataset compiled from 600 text files (300 ASD and 300 TD) and each file contains from (100 to 150) scanpath records for the same child. After conversion to CSV format and combination, it yielded a total of 63,627 eye-tracking records used for training and testing.

3.2 Data Preprocessing
Preprocessing is important in the machine learning and deep learning pipeline, where the data is prepared for model building and training. Data preprocessing is essential to ensure the prediction process quality and accuracy. The first step involves removing outliers and cleaning the data to address missing values and any inconsistencies within the dataset. Median imputation effectively addressed missing values across all four features, including filename, coordinates (x, y), and duration, this maintaining data integrity while ensuring

completeness. The second step involves encoding the categorical features using one-hot encoding, which converts them into a format suitable for our machine-learning models. Datasets that include (ASD), as well as individuals with (TD), are then collected and combined to create a comprehensive dataset. Data normalization is then applied to normalize values measured at different scales into a common metric. These steps are essential for obtaining improved results for most machine-learning applications. The final stage is to split the dataset into 80% training and 20% testing sets before the classification phase to ensure proper model evaluation on unseen data [19].

Table 1: Shows the preprocessing steps.

Step	Discretion
Data cleaning	Median addressed missing values across all four features
One-hot encoding	Converting categorical variables into a format suitable for machine learning algorithm
Feature scaling	Normalizing numerical features to ensure all variables are on a similar scale

3.3 Classification

This paper presents a comprehensive approach to developing a model capable of predicting and classifying children with autism spectrum disorder using eye movement data by employing a number of traditional machine learning algorithms and deep learning techniques to achieve this goal [20].

3.3.1 Machine Learning Models

This study utilizes and compares the performance of three traditional machine learning algorithms: Gradient Boosting Classifier, Logistic Regression, and Random Forest Classifier, based on the strength of these models and their ability to handle complex, high-dimensional data. The models are described in detail below:

3.3.1.1 Random Forest Classifier

Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the class that is the mode of the classes (classification) of the individual trees [21]. The algorithm of using Random Forest Classifier:

Algorithm 1: Random Forest Classifier

FOR i = 1 to n_trees DO
  1. Bootstrap Sampling:
    - Create Di by randomly sampling n samples with replacement from D
  2. Tree Construction:
    Initialize empty decision tree Ti
    WHILE nodes can be split DO
      a. Randomly select m_features from m available features
      b. Find best split point among m_features
      c. Split node into left and right child nodes
      d. Continue until min_samples or leaf node is reached
    END WHILE
  3. Add Ti to forest F
END FOR

3.3.1.2 Gradient Boosting Classifier

Gradient Boosting is another ensemble technique that builds trees sequentially, where each new tree corrects the errors of the previous ones [22]. The algorithm of using Gradient Boosting Classifier:

Algorithm 2: Gradient Boosting Classifier

1. Initialize model with a constant value:
 $F_0(x) = \text{argmin} \sum L(y_i, \gamma)$
 where L is the loss function
 FOR m = 1 to n_iterations DO
   1. Compute pseudo-residuals:
     $r_{im} = -[\partial L(y_i, F(x_i))/\partial F(x_i)]$
     where $F(x_i)$ is the current prediction
   2. Fit a weak learner (decision tree) $h_m(x)$ to pseudo-residuals
   3. Compute multiplier γ_m:
     $\gamma_m = \text{argmin} \sum L(y_i, F_{m-1}(x_i) + \gamma * h_m(x_i))$
   4. Update model:
     $F_m(x) = F_{m-1}(x) + \eta * \gamma_m * h_m(x)$
 END FOR

3.3.1.3 Logistic Regression Classifier

Despite its name, Logistic Regression is a linear classification model rather than regression. It estimates the probability of an instance belonging to a particular class [23]. The algorithm for using the Logistic Regression Classifier:

Algorithm 3: Logistic Regression Classifier

1. Initialize parameters:
   $\theta = [\theta_0, \theta_1, ..., \theta_m]$ with small random value// where θ_0 is bias term
 REPEAT
   1. Calculate hypothesis (sigmoid function):
     $h(x) = 1 / (1 + e^{\wedge}(-\theta^T x))$
   2. Compute cost function:
     $J(\theta) = -(1/n) \sum [y * \log(h(x)) + (1-y) * \log(1-h(x))]$
   3. Update parameters:
     FOR j = 0 to m DO
       $\theta_j = \theta_j - \alpha * (1/n) * \sum (h(x) - y) * x_j$
     END FOR
 UNTIL (change in $J(\theta) < \varepsilon$) OR (iterations > max_iter)

3.3.1.4 Model Selection Rationale Analysis

These models were selected for their complementary strengths:

1. Random Forest provides robust performance across various data types and can capture non-linear relationships.

2. Gradient Boosting often achieves high accuracy and can handle complex patterns in the data.

3. Logistic Regression, while simpler, provides easily interpretable results and often performs well on high-dimensional data.

By comparing these diverse models, we aimed to gain a comprehensive understanding of the predictive power of our eye-tracking data for ASD classification. The variation in their

underlying algorithms allows us to explore different aspects of the data and potentially uncover insights that might be missed by using a single approach [24].

3.3.2 Deep Learning Models

This study investigated deep learning methods for ASD classification using eye-tracking data and standard machine-learning techniques. Constructed a hybrid CNN-LSTM model and a Convolutional Neural Network (CNN) as separate deep learning architectures. These models were selected because of their capacity to extract hierarchical features from raw data automatically and identify patterns in eye movement trajectories that are both temporally and spatially related [25].

The binary cross-entropy loss and Adam optimizer were used to train both models. Further methods to avoid overfitting, including early stopping and learning rate reduction, were also used. While the hybrid CNN-LSTM model captures both temporal and spatial patterns in eye movements, the CNN architecture is highly effective at extracting pertinent features from raw data. Based on our findings, both models demonstrated high accuracy, with the hybrid model demonstrating superior performance due to its ability to model temporal dynamics. Table (2) below shows the layers of each deep learning model used in the study:

Table 2: layers of each deep learning model used.

Layer Number	Layer Type	CNN Model	Hybrid CNN-LSTM Model
1	Input	1D Input Layer	1D Input Layer
2	Convolutional	Conv1D (64 filters, kernel size 3)	Conv1D (64 filters, kernel size 3)
3	Activation	ReLU	ReLU
4	Max Pooling	MaxPooling1D (pool size 2)	MaxPooling1D (pool size 2)
5	Convolutional	Conv1D (64 filters, kernel size 3)	Conv1D (64 filters, kernel size 3)
6	Activation	ReLU	ReLU
7	Max Pooling	MaxPooling1D (pool size 2)	MaxPooling1D (pool size 2)
8	Flatten	Flatten	Flatten
9	Recurrent	-	LSTM (50 units)
10	Dense	Dense (50 units)	Dense (50 units)
11	Activation	ReLU	ReLU
12	Dropout	Dropout	Dropout
13	Output	Dense (2 units)	Dense (2 units)
14	Activation	Softmax	Softmax

This table includes a "Layer Number" column to show the sequence of layers in each model. The CNN model has 13 layers in total, while the Hybrid CNN-LSTM model has 14 layers, with the additional LSTM layer being the key difference.

4. Results

In this study, we built and analyzed three different ways of identifying ASD using eye-tracking data: standard machine learning models, a Convolutional Neural Network (CNN), and a hybrid CNN-LSTM [24]. Each strategy yielded encouraging results, with deep learning models performing particularly well.

4.1 Traditional Machine Learning Models

This work evaluated three widely used machine learning algorithms: Random Forest Classifier, Gradient Boosting Classifier, and Logistic Regression. These models were trained and tested on the preprocessed eye-tracking dataset.

4.1.1 Logistic Regression: demonstrated the highest accuracy at 99.37%, outperforming both Random Forest and Gradient Boosting. This is a surprising result, as Logistic Regression is generally considered a simpler model than ensemble methods like Random Forest and Gradient Boosting.

4.1.2 Random Forest: also performed exceptionally well, achieving an accuracy of 98.30%. This high accuracy suggests that the ensemble of decision trees could capture complex patterns in the eye-tracking data effectively.

4.1.3 Gradient Boosting: underperformed compared to the other two models, with an accuracy of 73.75%. This result is contrary to what is often seen in practice, where Gradient Boosting tends to perform very well on various tasks.

Table 3: Performance of Traditional Machine Learning Models

Model	Accuracy
Logistic Regression	99.37%
Random Forest	98.30%
Gradient Boosting	73.75%

4.2 Convolutional Neural Network (CNN)

The CNN model demonstrated exceptional performance, achieving an accuracy of 99.70% on the test set. This result surpasses the performance of the traditional machine learning models, indicating that the CNN model was able to learn highly discriminative features from the raw eye-tracking data automatically. The CNN's ability to capture spatial relationships in the data appears to be particularly valuable for ASD classification. The high accuracy suggests that there are distinct patterns in the eye movements of individuals with ASD that convolutional filters can effectively detect.

4.3 Hybrid CNN-LSTM Model

The hybrid CNN-LSTM model further improved upon the standalone CNN's performance, achieving an accuracy of 99.89%. This marginal but significant improvement suggests that incorporating temporal information through the LSTM layer provides additional discriminative power for ASD classification. The hybrid model's superior performance indicates that both spatial patterns (captured by the CNN layers) and temporal sequences (captured by the LSTM layer) in eye movements are important for accurate ASD classification

Table 4 : Performance Comparison of All Models

Model	Accuracy
Logistic Regression	99.37%
Random Forest	98.30%
Gradient Boosting	73.75%
CNN	99.70%
Hybrid CNN-LSTM	99.89%

4.4 Comparison of Approaches

The results shown in Figure 5 indicate that all models performed exceptionally well. However, the deep learning approaches, particularly the hybrid CNN-LSTM model, demonstrated superior performance. The traditional machine learning models, while highly accurate, were outperformed by the deep learning models. This suggests that the ability of CNNs and LSTMs to automatically learn hierarchical features and temporal dependencies provides an advantage in capturing the nuanced patterns in eye-tracking data that distinguish ASD from TD individuals.
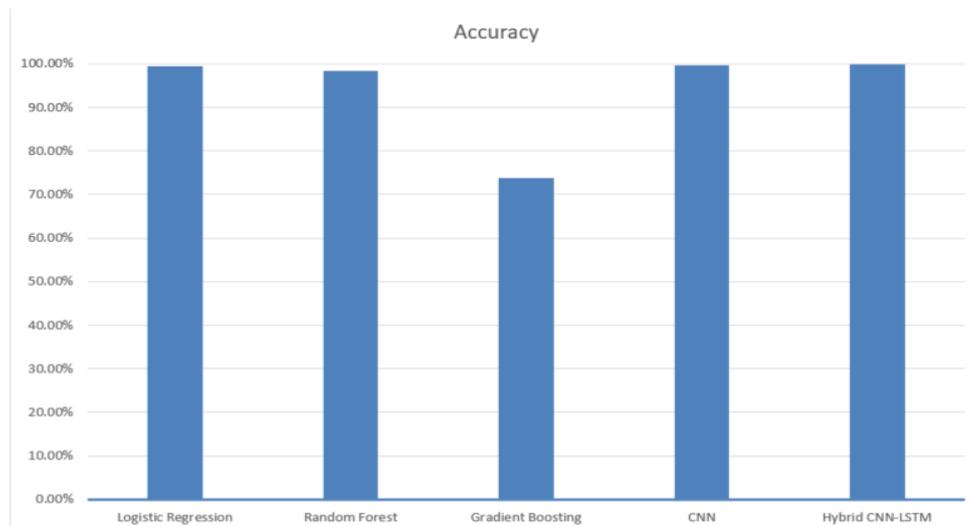
Figure 5 : a visual comparison of the accuracy achieved by each model

4.5 Model Evaluation Metrics

To provide a more comprehensive evaluation of our models, we calculated additional metrics, including precision, recall, and F1-score for each approach. Table (5) presents these metrics:

Table 5 : Detailed Evaluation Metrics for All Models

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	99.37%	0.972	0.971	0.971
Random Forest	98.30%	0.983	0.983	0.983
Gradient Boosting	73.75%	0.71	0.78	0.70
CNN	99.70%	0.997	0.997	0.9970
Hybrid CNN-LSTM	99.89%	0.999	0.999	0.998

4.6 Feature Importance

We analyzed feature importance for the traditional machine learning models to understand which aspects of the eye-tracking data were most influential in classification decisions.
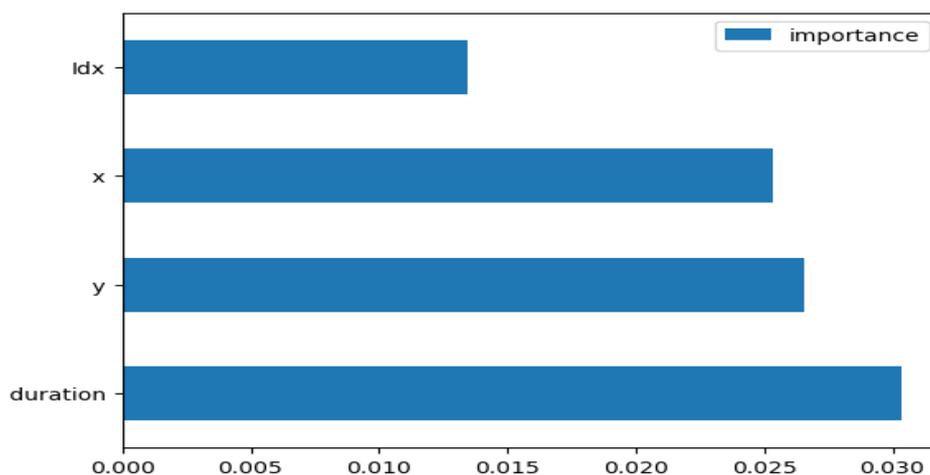
Figure 6 : Top 4 Most Important Features.

Feature importance analysis revealed the relative impact of input variables on the model's ASD detection. The features ranked by importance are: 1-Duration, 2-Y coordinate, 3-X coordinate, 4- index (Idx). Duration emerged as the most critical feature, suggesting fixation length strongly distinguishes between ASD and typically developing children. This aligns with research indicating atypical gaze patterns in ASD. The next most important features were Y and X coordinates representing spatial gaze information. Their prominence underscores the value of eye-tracking data in ASD detection. While less important, the Index still contributed to the model's decisions. This analysis validates eye-tracking data's usefulness in ASD detection and guides future research. It highlights the need to focus on fixation duration and spatial distribution of attention when developing diagnostic tools or interventions for ASD.

4.7 Learning Curves

To assess model generalization and potential overfitting, this study analyzed the Area Under the Curve curves (AUC) for traditional machine learning models and learning curves for deep learning models.

4.7.1 ROC Curves for Traditional Machine Learning Models

The ROC curves in Figure (7) reveal:
- Random Forest and Logistic Regression: Both achieved perfect discrimination (AUC = 1.00), indicating excellent performance. This suggests highly effective decision boundaries and strong separability of classes in the feature space.
- Gradient Boosting: Good performance with AUC of 0.83, aligning with its lower accuracy than other models. This model, while effective, didn't capture the class distinctions as well as the others.
- The near-perfect performance of Random Forest and Logistic Regression, while impressive, raises concerns about potential overfitting, especially if the dataset is not fully representative of the broader population.
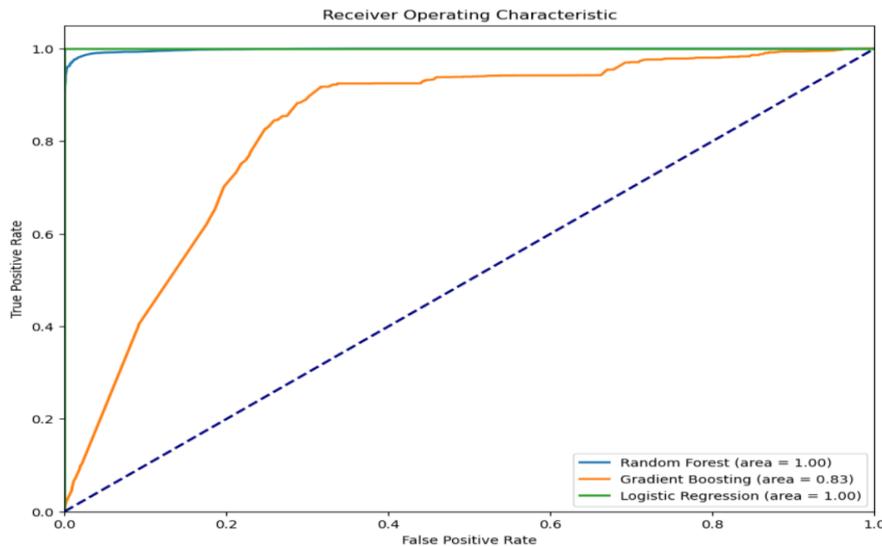
Figure 7 : the ROC curves for our traditional machine learning models: Random Forest, Gradient Boosting, and Logistic Regression.

4.7.2 Learning Curves for Deep Learning Models
 The CNN Model in Figure (8) shows:
- Accuracy: Rapid increase in the first few epochs, stabilizing around 99-100% for training and test sets.
- Loss: Sharp decrease initially, stabilizing at very low values for both sets.
- Close alignment of training and test metrics suggests good generalization.
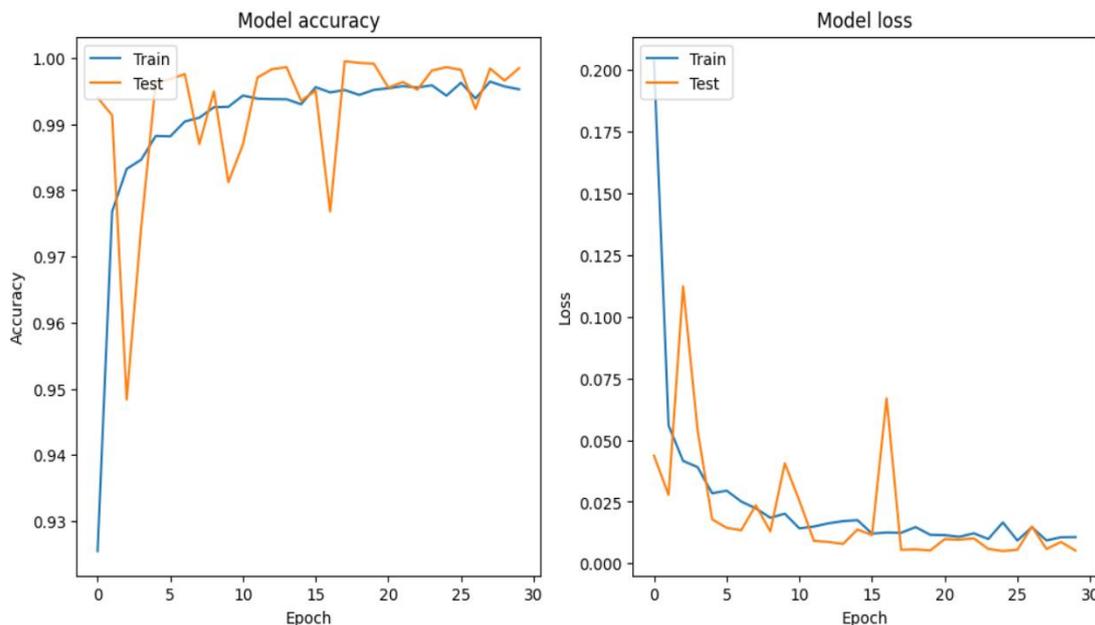
Figure 8 : CNN learning curves

The Hybrid CNN-LSTM Model in Figure (9):
- Accuracy: Similar pattern to CNN, with some fluctuations in test accuracy, likely due to model complexity and batch randomness.
- Loss: Consistent decrease for both sets, with some fluctuations in test loss.
- Overall trend shows good convergence without clear signs of overfitting.

Figure 9 : Hybrid CNN-LSTM learning curves

Both models demonstrate excellent learning patterns, with high accuracies and low losses on training and test sets. The stability in later epochs indicates successful learning of discriminative features from eye-tracking data.

4.8 Comparison with Other Studies

To contextualize our findings within the broader field of ASD classification using eye-tracking data, this study is compared with several recent studies in literature. Figure (10) provides a visual comparison of the classification accuracies achieved by our models and those reported in other studies.
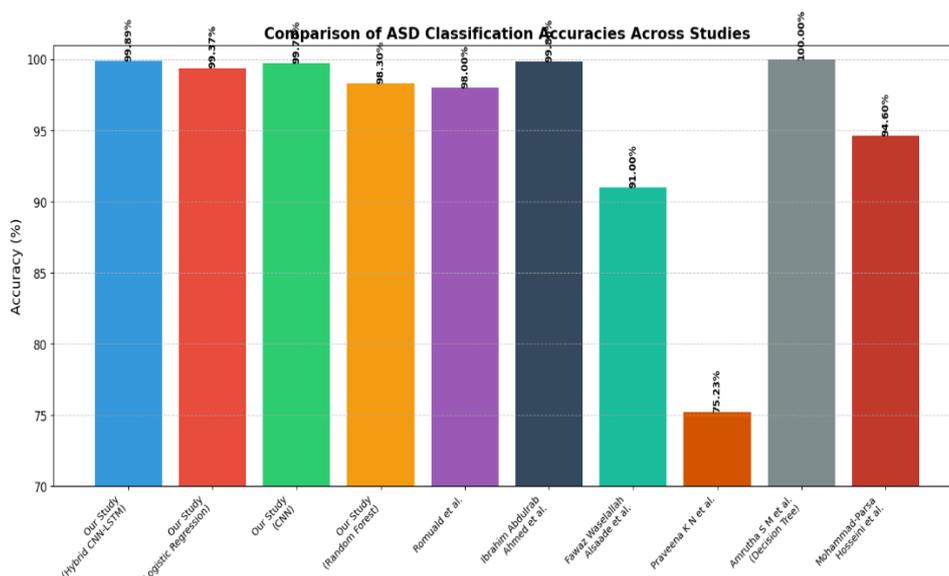
Figure 10 : Comparison of ASD Classification Accuracies across Studies.

The hybrid CNN-LSTM model achieved an accuracy of 99.89%, comparable to the highest accuracies reported in the literature. Notably, our model slightly outperformed the neural network approach of Ibrahim Abdulrab Ahmed et al [8]. Which achieved 99.8%

accuracy. Our logistic regression model (99.37% accuracy) also performed exceptionally well, matching, or exceeding the performance of many deep learning approaches in other studies.

It's important to note that Amrutha S M et al [11]. Reported 100% accuracy using a Decision Tree model. While this result is impressive, it's crucial to consider potential overfitting, especially given the perfect accuracy. While not achieving 100% accuracy, our models demonstrate consistently high performance across different architectures, suggesting robust and generalizable results.

Some studies, such as Praveena K N et al [10]. and Fawaz Waselallah Alsaade et al [9] reported lower accuracy (75.23% and 91% respectively). This variability in results across studies highlights the challenges in ASD classification and the potential impact of factors such as dataset characteristics, preprocessing techniques, and model architecture. Similarly, Talib et al [17] fuzzy decision-making framework achieved varying accuracies across different weight allocation scenarios (60%:40%, 40%:60%, 70%:30%, and 30%:70%). Their approach focused more on patient prioritization than pure classification, demonstrating how different methodologies can address various aspects of ASD assessment. This variability in results across studies highlights the challenges in ASD classification and the potential impact of factors such as dataset characteristics, preprocessing techniques, and model architecture.

The study results are particularly encouraging as they demonstrate high accuracy across multiple model types, from traditional machine learning approaches to advanced deep learning architectures. This consistency suggests that the eye-tracking dataset used in this research contains strong, reliable signals for ASD classification. These variations in performance can largely be attributed to differences in data preparation approaches and model architecture. Our hybrid approach, combined with comprehensive data preparation, demonstrates the importance of sophisticated model architecture and careful data handling in achieving reliable ASD classification. Table (6) compares our study's results comprehensively with those reported in recent literature on ASD classification using eye-tracking data.

Table 6 : Comparison of ASD Classification Accuracies across Studies

Study	Method	Accuracy
Our Study (best model)	Hybrid CNN-LSTM	99.89
Romuald et al. [7]	Neural Network with LSTM	98.00
Ibrahim Abdulrab Ahmed et al[8].	Neural Networks with Hybrid Feature Classification	99.80
Fawaz Waselallah Alsaade et al[9].	CNN with Transfer Learning (Xception)	91.00
Praveena K N et al[10].	CNN	75.23
Mujeeb Rahman Kanhirakadavath[11]	DNN-based model	97.00
Amrutha S M et al[12].	Decision Tree	100.00
Amrutha S M et al[12].	Logistic Regression	99.37
Mohammad-Parsa Hosseini et al[13].	Deep CNN (MobileNet)	94.60

4.9 Model Prediction Using Unseen Data

To validate the model's performance and effectiveness in real-world applications, the trained model was tested using data it had not been trained on before. New eye-tracking data was used, which was a new scanpath of an individual with Autism Spectrum Disorder (ASD). This data was continuously processed in the same way as the data the model was trained on to obtain the best prediction result. The results were in figure (11) as follows:

- RandomForest classifier: Consistently predicted ASD (0) for all cases in the new dataset, indicating that the child is likely to have autism.
- Gradient Boosting classifier: Predicted a combination of ASD (0) and TD (1), with a slight majority of TD predictions. This means that the child may not have autism.
- Logistic Regression classifier: Classified all eye movement data as ASD (0), indicating that the patient is likely to have autism.

```
RandomForest classifier predictions for the new data:
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

Gradient Boosting classifier predictions for the new data:
[1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 0 1 0 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1
 1 1 1 1 0 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 1
 0 0 0 0 1 0 1 1 1 1 1 0 1 1 1 0 1]

Logistic Regression classifier predictions for the new data:
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```

Figure 11: Test the model on new unseen data for ASD child.

5. Conclusion and Future Work

This study demonstrates the significant potential of machine learning and deep learning techniques in classifying ASD using eye-tracking data. Through a comprehensive approach that employed both traditional and advanced models, we achieved exceptional results in distinguishing between individuals with ASD and typically developing controls.

The hybrid CNN-LSTM model achieved the highest accuracy of 99.89%, showcasing the power of integrating spatial and temporal features in eye movement analysis. Notably, even our simpler logistic regression model attained 99.37% accuracy, highlighting the robust discriminative power of eye-tracking data for ASD classification. These findings contribute significantly to the growing research on automated, data-driven approaches to ASD diagnosis.

The implications of this research are far-reaching, suggesting that automated analysis of eye-tracking data could serve as a valuable tool for ASD screening and diagnosis. This approach offers potential advantages, including increased objectivity, earlier detection, and improved scalability of screening processes. However, limitations must be acknowledged. The very high accuracies achieved warrant careful consideration of potential overfitting. Additionally, the real-world applicability of these models in clinical settings requires further investigation.

Future research should focus on external validation using diverse datasets, longitudinal studies to understand developmental trajectories, and exploring multimodal approaches combining eye-tracking with other biomarkers. Key challenges include standardizing data collection across clinical settings and developing real-time processing capabilities. While further research is necessary, our results provide strong evidence for the utility of machine

learning in ASD classification using eye-tracking data. In conclusion, this study represents a significant step forward in the intersection of machine learning, eye-tracking technology, and ASD research, potentially paving the way for innovative neurodevelopmental disorder diagnosis and management approaches.

References

[1] A. K. Nasir, W. Strong-Bak, and M. Bernard, "Diagnostic Evaluation of Autism Spectrum Disorder in Pediatric Primary Care," *J. Prim. Care Community Heal.*, vol. 15, 2024, doi: 10.1177/21501319241247997.

[2] K. M. Halboot, A. K. Melconian, and J. A.- Saffar, "Profile of Some Cytokines in Sera of Children with Autism Syndrome," (2023). *Iraqi Journal of Science*, 56(1A), 125-131.

[3] M. J. Maenner *et al.*, "Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020," *MMWR Surveill. Summ.*, vol. 72, no. 2, 2023, doi: 10.15585/mmwr.ss7202a1.

[4] A. Sundas, S. Badotra, S. Rani, and R. Gyaang, "Evaluation of Autism Spectrum Disorder Based on the Healthcare by Using Artificial Intelligence Strategies," *J. Sensors*, 2023, doi: 10.1155/2023/5382375.

[5] A. I. Shihab, F. A. Dawood, and A. H. Kashmar, "Data Analysis and Classification of Autism Spectrum Disorder Using Principal Component Analysis," *Adv. Bioinformatics*, 2020, doi: 10.1155/2020/3407907.

[6] M. Alsaidi, N. Obeid, N. Al-Madi, H. Hiary, and I. Aljarah, "A Convolutional Deep Neural Network Approach to Predict Autism Spectrum Disorder Based on Eye-Tracking Scan Paths," *Inf.*, vol. 15, no. 3, pp. 1–23, 2024, doi: 10.3390/info15030133.

[7] R. C. B, F. Cilia, G. Dequen, J. Bosche, J. Guerin, and L. Vandromme, "Detection Thanks to Eye-Tracking and Neural Network-Based Approach," vol. 1, pp. 75–81, doi: 10.1007/978-3-319-76213-5.

[8] I. A. Ahmed *et al.*, "Eye Tracking-Based Diagnosis and Early Detection of Autism Spectrum Disorder Using Machine Learning and Deep Learning Techniques," *Electron.*, vol. 11, no. 4, 2022, doi: 10.3390/electronics11040530.

[9] F. W. Alsaade and M. S. Alzahrani, "Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms," *Comput. Intell. Neurosci.*, 2022, doi: 10.1155/2022/8709145.

[10] K. N. Praveena and R. Mahalakshmi, "Classification of Autism Spectrum Disorder and Typically Developed Children for Eye Gaze Image Dataset using Convolutional Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 3, pp. 362–370, 2022, doi: 10.14569/IJACSA.2022.0130345.

[11] M. R. Kanhirakadavath and M. S. M. Chandran, "Investigation of Eye-Tracking Scan Path as a Biomarker for Autism Screening Using Machine Learning Algorithms," *Diagnostics*, vol. 12, no. 2, pp. 1–19, 2022, doi: 10.3390/diagnostics12020518.

[12] S. Vinitha Sri and A. Mohana Priya, "Autism Spectrum Disorder Detection Using Machine Learning Methods," *Gedrag Organ. Rev.*, vol. 34, no. 01, pp. 1252–1254, 2021, doi: 10.37896/gor34.01/036.

[13] M. P. Hosseini, M. Beary, A. Hadsell, R. Messersmith, and H. Soltanian-Zadeh, "Deep Learning for Autism Diagnosis and Facial Analysis in Children," *Front. Comput. Neurosci.*, vol. 15, no. January, pp. 1–7, 2022, doi: 10.3389/fncom.2021.789998.

[14] S. Raj and S. Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 994–1004, 2020, doi: 10.1016/j.procs.2020.03.399.

[15] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques," *SN Comput. Sci.*, vol. 2, no. 5, pp. 1–9, 2021, doi: 10.1007/s42979-021-00776-5.

[16] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning techniques," *Heal. Inf. Sci. Syst.*, vol. 9, no. 1, 2021, doi: 10.1007/s13755-021-00145-9.

[17] H. Mohammed Talib, A. S. Albahri, and T. O. C. EDOH, "Fuzzy Decision-Making Framework for Sensitively Prioritizing Autism Patients with Moderate Emergency Level," *Appl. Data Sci. Anal.*, pp. 16–41, 2023, doi: 10.58496/adsa/2023/002.

[18] W. Zhou, M. Yang, J. Tang, J. Wang, and B. Hu, "Gaze Patterns in Children with Autism Spectrum Disorder to Emotional Faces: Scanpath and Similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 865–874, 2024, doi: 10.1109/TNSRE.2024.3361935.

[19] K. Goyle, "DataAssist : A Machine Learning Approach to Data Cleaning and Preparation."

[20] K. Nermend, W. A. G. A. Hussain, and A. I. Shihab, "Classification and monitoring of autism using svm and vmcm," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 14, pp. 4379–4389, 2018.

[21] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020, doi: 10.1109/JSTARS.2020.3026724.

[22] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of XGBoost," pp. 1–20, 2019, doi: 10.1007/s10462-020-09896-5.

[23] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, "Logistic Regression in Clinical Studies," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 112, no. 2, pp. 271–277, 2022, doi: 10.1016/j.ijrobp.2021.08.007.

[24] A. Halbouni, T. S. Gunawan, and S. Member, "CNN-LSTM : Hybrid Deep Neural Network for Network Intrusion Detection System," *IEEE Access*, vol. 10, no. August, pp. 99837–99849, 2022, doi: 10.1109/ACCESS.2022.3206425.

[25] N. A. K. Hussein and B. Al-Sarray, "Deep Learning and Machine Learning via a Genetic Algorithm to Classify Breast Cancer DNA Data," *Iraqi J. Sci.*, vol. 63, no. 7, pp. 3153–3168, 2022, doi: 10.24996/ijs.2022.63.7.36.