Applying Unsupervised and Supervised Machine Learning Methods to Saccadic Eye Movement Data to Differentiate Between ASD and TD Individuals

Justin Lisoway[1] · Katy Tarrit[1]

[1]Information and Computer Sciences Department, University of Hawai'i at Manoa, Honolulu, HI, United States

***Abstract*—This study aims to find eye movement differences between individuals with Autism Spectrum Disorder and their Typically Developing peers using advanced Machine Learning techniques. Utilizing a dataset previously recorded using the video-based eye tracker EyeLink 1000 (SR Research, Ottawa, Canada), we compare the performance of supervised and unsupervised ML models in identifying unique gaze patterns that can potentially distinguish individuals with ASD from TD. Our findings reveal that supervised models trained on specific labeled eye movement data achieve moderate accuracy in classification, while unsupervised models fail to uncover distinct groupings based on gaze metrics. This supports ML techniques potential to only detect ASD-related differences with labeled data. The hybrid approach produced the best results, highlighting the importance and value of manual feature extraction and deep learning. These results suggest that the use of ML in combination with eye movement data can enhance ASD screening and potentially aid in diagnosis, offering insights into supervised classification effectiveness.**

***Index Terms*—Autism spectrum disorder (ASD), typically developing (TD), eye movement data, saccadic eye movements, machine learning, deep learning, supervised learning, unsupervised learning, convolutional neural networks (CNNs), long-short term memory (LTSMs), k-mean clustering, DBSCAN, agglomerative clustering.**

I. Introduction

A. Background

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by differences in social interaction, communication, and repetitive behaviors. In recent years, the potential of eye-tracking technology to identify unique gaze patterns associated with ASD has gained considerable interest. Eye movement differences, such as saccadic adaptation rates and variability, have been explored as potential biomarkers for ASD, yet the findings remain inconsistent across studies. While some research highlights significant differences in gaze metrics between ASD and typically developing (TD) individuals, others fail to replicate these results.

Supervised machine learning methods that leverage labeled data may identify patterns that distinguish ASD from TD individuals, while unsupervised learning offers the possibility of discovering latent groupings without predefined labels. Despite the promise of these techniques, few studies have directly compared their effectiveness on the same dataset.

This study addresses these gaps by applying supervised and unsupervised machine learning models to eye-tracking data collected from ASD and TD individuals. We aim to determine whether distinct eye movement patterns exist between the groups and can be detected by unsupervised and supervised methods. We also aim to assess the effectiveness of a hybrid approach that combines feature engineering with a deep learning model. By doing so, this research contributes to the ongoing exploration of eye-tracking data as a potential diagnostic tool for ASD and seeks to clarify the roles of manual feature extraction and deep learning in achieving accurate classification.

B. Contribution

In the literature, there is disagreement on specific eye-movement features that are consistently different between ASD and TD individuals. Features such as the saccade adaptation rate [1][2], saccade variability [2][3], and saccadic accuracy [3] were shown in some studies to be different between ASD and TD groups, but other studies showed no evidence of this [4]. This study will apply ML methods to analyze any differences in those features and further determine which (if any) are consistently different enough in our data to predict ASD or TD from only eye-movement data. This study accomplishes this by creating novel unsupervised and supervised ML models.

The literature surrounding the use of ML models in this domain is also incomplete. Many papers apply and discuss the results of different categories of supervised learning approaches. Limited research was found that had applied an unsupervised method, and there is little to no available literature comparing the effectiveness of either approach on the same dataset. This paper addresses this gap in research by comparing the effectiveness of unsupervised and supervised methods on the same dataset.

Although the surrounding literature does compare many models, it is unclear whether manual feature extraction or deep learning plays a more important role in the contribution to accuracy. Studies have shown success in implementing a model trained on specific features and other studies have shown success in implementing a deep learning approach. Thus, this study implements both approaches independently, and then implements a hybrid approach, combining feature extraction and deep learning, to evaluate the extent to which manually extracted features versus deep learning models are able to differentiate ASD from TD based on eye movement data. Considering the success with both approaches in the literature, it may be that the relevant features are already extracted in feature engineering and that those same features are also being detected and incorporated into deep learning models. Or, the deep learning approaches may be learning features that are currently unknown and may be unrelated to those extracted manually. This paper will help uncover that nuance.

C. Motivation

The literature has identified potential eye movement differences that may serve as biomarkers for ASD. Leveraging machine learning methods on this saccadic eye movement data could help uncover patterns in this specific type of eye movement that differ among ASD individuals, thereby improving screening efficiency and accuracy of ASD prediction models.

D. Goals & Research Questions (RQs)

RQ1: Can supervised and unsupervised machine learning models help in differentiating eye movements of TD individuals from individuals with ASD?

This research question aims to compare the performance of both supervised and unsupervised ML models in identifying patterns within the saccadic eye movement data of ASD and TD individuals. On the unsupervised side, we want to explore whether a model can naturally group participants based on differences in their eye movements, revealing patterns that may not be immediately apparent to humans. On the supervised side, the goal is to see if a model can effectively learn features that distinguish ASD from TD participants when provided with labeled data. By comparing both approaches, we can evaluate whether a predefined feature learning process (supervised) or a more organic discovery of groupings (unsupervised) performs better in categorizing ASD and TD participants.

RQ2: Can a hybrid model approach combining feature engineering and end-to-end deep learning improve the detection of ASD-related differences in eye-tracking data compared to ML models that use either feature engineering or deep learning alone?

This research question analyzes the extent to which manually extracted features in saccadic eye movement data are being learned by the deep learning approach.

If the hybrid approach performs similarly to the model trained only on manually extracted features and is better than the deep learning approach, it can be inferred that the level of accuracy in the hybrid approach is coming from the extracted features. If the hybrid approach performs similarly to the deep learning approach and better than the model trained only on manually extracted features, it can be inferred that the accuracy in the hybrid approach is coming from unknown features that the deep learning procedure is learning. Lastly, if the hybrid approach performs better than both other approaches, then it can be inferred that both the manually extracted features and unknown learned features play an important role in distinguishing between ASD and TD eye movements.

II. Related Works

There are multiple related studies preceding this research. The following related works provide a more insightful understanding of this research field, including any established features that are different between ASD and TD groups and the current state of ML research on this topic as it relates to our research questions. The table in Appendix A summarizes the key findings of each related work and their respective research questions.

Several studies have focused on differentiating saccadic eye movement characteristics between individuals with Autism Spectrum Disorder (ASD) and Typically Developing (TD) controls. These studies are generally looking into features in eye movement data that may distinguish the two groups, which may provide insight into features that should be extracted for supervised machine learning.

Johnson et al. [1] observed that individuals with high-functioning autism (HFA) exhibited a slower adaptation of saccade gain and longer corrective saccade latencies, suggesting differences in visual feedback utilization and motor learning. This aligns with findings by Mosconi et al. [2], who reported that ASD individuals have slower saccade adaptation and increased amplitude variability, indicating impaired cerebellar-dependent learning mechanisms and broader motor control deficits. Moreover, Schmitt et al. [3] identified that saccades in ASD individuals were characterized by reduced accuracy, greater trial-to-trial variability, and altered dynamics, such as prolonged acceleration to peak velocity. All three of these studies seem to be aligned in their conclusions that there are inherent differences in the saccades of ASD individuals compared to TD individuals.

In contrast, Tarrit et al. [4] found no significant differences in saccadic adaptation rates between ASD and TD groups. Although individuals with ASD exhibited slightly hypermetric saccades, the overall rate of adaptation and variability in saccade amplitude were comparable to those of TD individuals, suggesting that some eye movement characteristics may not differ markedly between the groups.

The application of machine learning (ML) and deep learning (DL) techniques has also been a major focus for ASD classification using eye-tracking data since 2019-2020, when more advanced ML models emerged. Here, we examine what technical approaches have shown success in the literature.

Several machine learning approaches have been applied to differentiate between ASD and TD participants using eye-tracking data. Supervised learning methods, such as Support Vector Machines (SVM), Random Forest, Long-Short Term Memory (LSTM), and Convolutional Neural Networks (CNN) have been widely used due to their effectiveness in classification tasks.

Asmetha Jeyarani and Senthilkumar [5] reviewed the use of ML models like SVM, Random Forest, and K-Nearest

Neighbors (KNN), as well as DL models like CNN and LSTM networks. These models have shown considerable accuracy in distinguishing ASD from TD individuals, with SVM being the most commonly used and achieving high accuracy in some studies. According to this study, deep learning models have proven particularly effective in handling complex eye-tracking data and temporal sequences.

Cilia et al. [6] transformed eye-tracking scan paths into visual representations to utilize CNNs for image classification, achieving a high accuracy of around 90% in distinguishing ASD from non-ASD participants. Additionally, Ahmed et al. [7] demonstrated that Artificial Neural Networks (ANNs) and Feedforward Neural Networks (FFNNs) achieved the highest accuracy (99.8%) when classifying ASD using eye-tracking data, outperforming pre-trained CNN models and hybrid approaches. They also found that combining feature extraction techniques like Local Binary Pattern (LBP) and Grey Level Co-occurrence Matrix (GLCM) enhanced performance.

Ahmed and Jadhav [8] applied a deep learning model based on CNNs to classify ASD and TD children using eye-tracking scan path data. With a dataset of 547 images from 59 participants, they achieved an impressive accuracy of 98%. Their work underscores the effectiveness of CNN architectures in processing visual and temporal patterns inherent in eye-tracking data for ASD detection. Furthermore, the use of data augmentation techniques such as rotation and shearing helped to reduce over-fitting and improve model robustness, especially given the small dataset size.

Gaspar et al. [9] proposed a novel Giza Pyramids Construction (GPC)-optimized Kernel Extreme Learning Machine (KELM) methodology, which demonstrated superior performance (98.8% accuracy) compared to traditional ML techniques. Furthermore, NLP-based transformations of eye-tracking data have been explored by Elbattah et al. [10], who transformed saccade and fixation sequences into textual strings for sequence-based classification models, with ConvNet models outperforming LSTM models.

Yoo et al. [11] further showed the potential of integrating eye-tracking data with ML models for classifying neurodevelopmental disorders, achieving 76.3% accuracy in identifying ADHD using a soft voting model that combined extra tree and random forest classifiers. The integration of eye-tracking data alone produced robust results, further validating its potential applicability to ASD classification.

Supervised learning approaches have generally shown substantial levels of accuracy across many studies, with varying datasets and methodologies. Some papers focus on learning very specific features (blink rate, gaze duration, peak velocity, etc), and others apply an entirely black-boxed-based deep learning approach. Both of these methods have shown accuracy; however, there is a gap in analyzing the potential overlap of these approaches. Since there is no explainability in the deep learning approach of these studies, there is no way of knowing to what extent the deep learning approach may be learning the same manually identified features as the former approach.

On the other hand, unsupervised learning techniques have been less extensively studied but show some potential in this domain. Elbattah et al. [12] applied unsupervised clustering techniques to eye-tracking scan path images and identified patterns that differentiated ASD from TD participants. Their findings revealed that ASD participants exhibited higher gaze velocity and acceleration, and clustering techniques were effective in grouping participants based on these characteristics.

Although unsupervised methods appear to reveal hidden patterns in the data in one study, their implementation is not as well studied and their performance in this study did not reach the same level as supervised models, suggesting a need for further refinement to achieve comparable accuracy or an underlying inability of unlabeled data to be as effective in this domain of machine learning as labeled data. Furthermore, there is no available study that aims to compare both approaches on the same dataset.

Summary: The peer-reviewed studies in literature that examine features without applying machine learning do not all agree on whether eye movement dynamics alone are a good predictor of ADHD. This mixture in the literature suggests that more research must be done in this area to better understand which differences in saccadic eye movement (if any) exist and are consistent enough between ASD and TD individuals to predict ASD. Also, a wide range of machine learning and deep learning techniques, including Long-Short Term Memory (LSTMs), Convolutional Neural Networks (CNNs), and hybrid feature extraction methods, have shown high accuracy in classifying ASD based on eye-tracking data. Unsupervised methods, such as clustering, have yielded promising insights, but they are less common and have underperformed compared to supervised approaches. Research question 1 aims to compare supervised and unsupervised methods on the same dataset to determine their relative efficacy in ASD classification. There is also limited research on the explainability of any deep learning approach in the literature. Research question 2 aims to analyze to what extent there might be overlap in the features being learned by a deep learning approach compared to a manual-feature-extraction-based machine learning approach.

III. Methodology

A. Dataset

1) Description: The raw data for this paper was made available from Tarrit et al. [4] and can be found here: Dataset

The dataset includes 109 eye-tracking recordings (68 ASD, 41 TD), recorded at 500 Hz during trials. The data is comprised of gaze coordinates (GazeX, GazeY), timestamps, and pupil size. Only adaptation trials were analyzed. This data was recorded from study participants from Tarrit et al. [4] while they performed the following task while their eye movements were recorded. Participants were asked to fixate on a cross that appeared on a screen and were instructed to always fixate on the dots that would appear later on. After a time interval, the cross would disappear and a dot would appear simultaneously somewhere else on the screen. In the short period of time that the participants were undergoing their initial saccade to the first dot, it would disappear, and a new dot would appear somewhere else on the screen simultaneously. The participant's saccade would then be altered to direct their gaze to the new dot. Finally, the dot would disappear and the initial fixation cross would reappear, marking the start of the next trial.

The data contains 109 raw .edf files that contain the data directly outputted by the EyeLink 1000 during each participant's trials. The data was recorded in 500 Hz and contains the timestamp, X position of gaze (GazeX), Y position of gaze (GazeY), and Pupil size at each moment in time throughout the entire experiment for each participant. There were 68 control trials followed by 240 adaptation trials followed by another 68 control trials for each participant (a total of 376 trials per participant). That is, each single raw .edf file contained the data from all 376 trials for a single participant in sequence. There were 41 TD participants and 68 ASD participants, which make up the 109 raw .edf files in the data.

2) Sample: The following image shows a sample of the raw data that can be collected by the EyeLink 1000. Each row represents a moment in time with the X and Y coordinates of the gaze along with the pupil size being recorded. A record is made every 0.002 seconds (500 Hz).

Timestamp	Gaze X	Gaze Y	Pupil Size	
148054	958.3	545.7	218.0	...
148056	958.5	545.4	217.0	...
148058	958.6	545.2	217.0	...
148060	958.8	545.0	217.0	...
148062	958.8	545.1	218.0	...
148064	958.8	545.4	218.0	...
148066	958.8	545.7	218.0	...
148068	958.8	546.2	218.0	...
148070	958.8	546.6	219.0	...
148072	958.9	547.0	219.0	...
148074	959.0	547.2	219.0	...
148076	959.1	547.1	219.0	...

Fig. 1. Taken from sr [13], the raw data from the EyeLink 1000 collects data in this format.

3) Event Messages: The raw data also contains corresponding messages from the EyeLink 1000 that contain timestamps and corresponding event messages that mark specific events in the experiment. This message data was tokenized and incorporated into the learning process. Some of the possible messages are:

TABLE I: EyeLink 1000 Messages

Message	Description
ControlTrial1	Start of a Control Trial type 1
ControlTrial2	Start of a Control Trial type 2
ControlTrial3	Start of a Control Trial type 3
ControlTrial4	Start of a Control Trial type 4
Trial1	Start of an Adaptation Trial type 1
Trial2	Start of an Adaptation Trial type 1
Trial3	Start of an Adaptation Trial type 1
Trial4	Start of an Adaptation Trial type 1
EndTrial	Indicates the end of any trial

B. General Preprocessing

A general preprocessing methodology was applied to all the raw .edf files. Then, more case-specific preprocessing was further applied as required for the unsupervised and supervised approaches. Another layer of specific preprocessing was also applied based on the specific supervised approach (feature-based, deep learning, or hybrid).

Each raw .edf file was first pre-processed by an event extraction script in MATLAB that matched events to the messages in the raw data files. This script also processed the .edf files into .asc files that reformatted the raw data and automatically extracted the data of each event. This produced event data that showed the peak velocity, duration, start X, start Y, end X, end Y, and amplitude of each saccade detected by the EyeLink 1000, but also retained the raw data.

These subsets of the data were then separated, one containing only the features extracted by the EyeLink 1000, and the other containing all the extracted numerical data. This numerical data is technically different from the original raw .edf files, although it contains all the same information in a different file format and will therefore be known as the raw data from this point on.

The control trials were removed from each subset of the data, and all 240 adaptation trials for each participant were exported to a .csv file. To decrease the density and variability of trial data, the .csv file was filtered to remove everything except the Trial 1 Adaptation Trials. This allowed the machine learning process to focus on learning features that varied between ASD and TD participants while they completed homogeneous tasks.

At this stage, each row in the .csv file represented a single trial from a single participant. The columns indicated the participant type, participant number, and either the timestamp followed by the feature data or the timestamp followed by the raw data depending on the data subset, for the given trial.

The last stage of preprocessing (done in Python) expanded each row of data to contain all the data from every trial for each participant, not just a single trial per row. The data was then further pre-processed slightly differently for each machine learning method technique. The data at this point still contains all of the relevant information from the original raw .edf files and will continue to be known as the raw data.

C. Unsupervised Learning

In Python, the labels of the data were removed, and the raw data was copied into a Pandas data frame. Given the lack of literature on this topic, 6 iterative unsupervised learning approaches were applied as follows.

An overview of each of the preprocessing steps and each of the 6 iterative unsupervised methods is shown in Figure 2.

1) Unsupervised Method 1: The raw data was scaled based on the event features detected by the EyeLink 1000, then a K-means clustering algorithm was applied. These features included saccade peak velocity, saccade duration, saccade amplitude, and saccade starting and ending coordinates (AKA Features 1).

2) Unsupervised Method 2: The raw data was scaled in the same way as method 1. Features 1 were used and Principal Component Analysis (PCA) was then applied to condense the high-dimensionality data with n=2 as a parameter. Three different unsupervised methods were applied; K-means clustering (n=2), DBSCAN (min_samples = 5), and agglomerative clustering (n=2).

3) Unsupervised Method 3: Different specific features were extracted for the algorithms. Howard et al. [14] showed that ASD individuals may shift their gaze more often than TD individuals and they may spend less time fixating on a gaze than TD individuals. In the raw eye tracking data, this is shown as a higher number of events detected by the EyeLink 1000 as well as a shorter average fixation duration for ASD individuals versus TD individuals.

Due to the reviewed study, the features extracted for this method included the number of events across all trials per participant and the mean duration of a fixation event across all fixation events in adaptation trials for that participant. The mean duration was calculated after removing the highest and lowest 3% of duration data points, which corrected for erroneously long- or short-duration events. This set of features is AKA Features 2.

From there, the same steps as in Unsupervised Method 2 were applied; the features were scaled, PCA was applied, and the same three unsupervised algorithms were applied (K-means, DBSCAN, and agglomerative).

4) Unsupervised Method 4: Different features were extracted. Only the average peak velocity of a saccade across all events per participant was used. Sadria et al. [15] found that TD individuals had a higher peak velocity in their saccades than ASD individuals.

The average peak velocity was calculated for each participant after removing each participant's respective highest and lowest 1% of peak velocities, which corrected for erroneous outliers. This is AKA Features 3. The features were then scaled as in methods 1-3 and PCA (n=2) was applied. The same 3 unsupervised learning techniques as in methods 2-3 were applied (K-means, DBSCAN, and agglomerative).

5) Unsupervised Method 5: Method 5 repeated method 4, without applying PCA. Features 3 were scaled and the same unsupervised algorithms were applied.

6) Unsupervised Method 6: Features 2 and Features 3 were combined. Due to the superior results from excluding PCA in method 5, PCA was also excluded in method 6. Thus, the mean peak velocity (excluding the highest and lowest 1%), the mean duration (excluding the highest and lowest 3%), and the number of events per participant were the features scaled and then given as input to the same 3 unsupervised models (K-means, DBSCAN, and agglomerative).

This final unsupervised iteration yielded the best results, which are shown in Table II for each of the 3 clustering algorithms.

D. Supervised Learning

For supervised learning, three general data preprocessing ideas were applied: feature-based, deep learning, and a hybrid combination of both.

1) Feature-Based: A set of specific features were extracted from the event data from the Eyetracker II. These features included the peak velocity, duration, reaction time, change in GazeX, and change in GazeY for each trial. On the participant level, the number of events was also added as a feature. This represents a combination of Features 2 and Features 3 from the unsupervised method that performed the best.

These features and their participant-type labels were then split into an 80/20 train/test split. The machine learning model outlined in Figure 4 was then applied to this labeled feature data and its results are outlined in Table III (Feature row).

2) Deep Learning: The second preprocessing method aimed to give the model as much raw data as possible while balancing available computational resources. No specific features were extracted. The original .edf files that had underwent initial preprocessing steps were organized sequentially. That is, a single row in the Pandas data frame contained all raw data points from the Trial 1 adaptation trials sequentially for each participant.

All of the labeled raw data was split into an 80/20 train/test split. The machine learning model outlined in Figure 4 was then applied to this dataset and its results are outlined in Table III (Deep row).

3) Hybrid Approach: The model from the feature-based approach was used as a starting point for training on all the raw data. That is the CNN/LSTM hybrid model was pre-trained on feature sets Features 2 and Features 3, and then trained further on all of the raw data discussed in the deep learning method.

An overview of this approach is outlined in Figure 5 and its results are outlined in Table III (Hybrid Row).

IV. Results

To answer both of our research questions regarding the predictability of Autism Spectrum Disorder given eye tracking data, the results from applying the methodology described in this paper are analyzed here.

For each unsupervised learning method, Table II shows a row for each unsupervised learning model and a row demonstrating what a perfect result would be. In the first outer column, we demonstrate each model's ability to cluster ASD participants and in the second column we see its ability to cluster TD participants. There is no real distinction between group 1 and group 2, meaning that the perfect example could have also put 100% of the participants with ASD in group

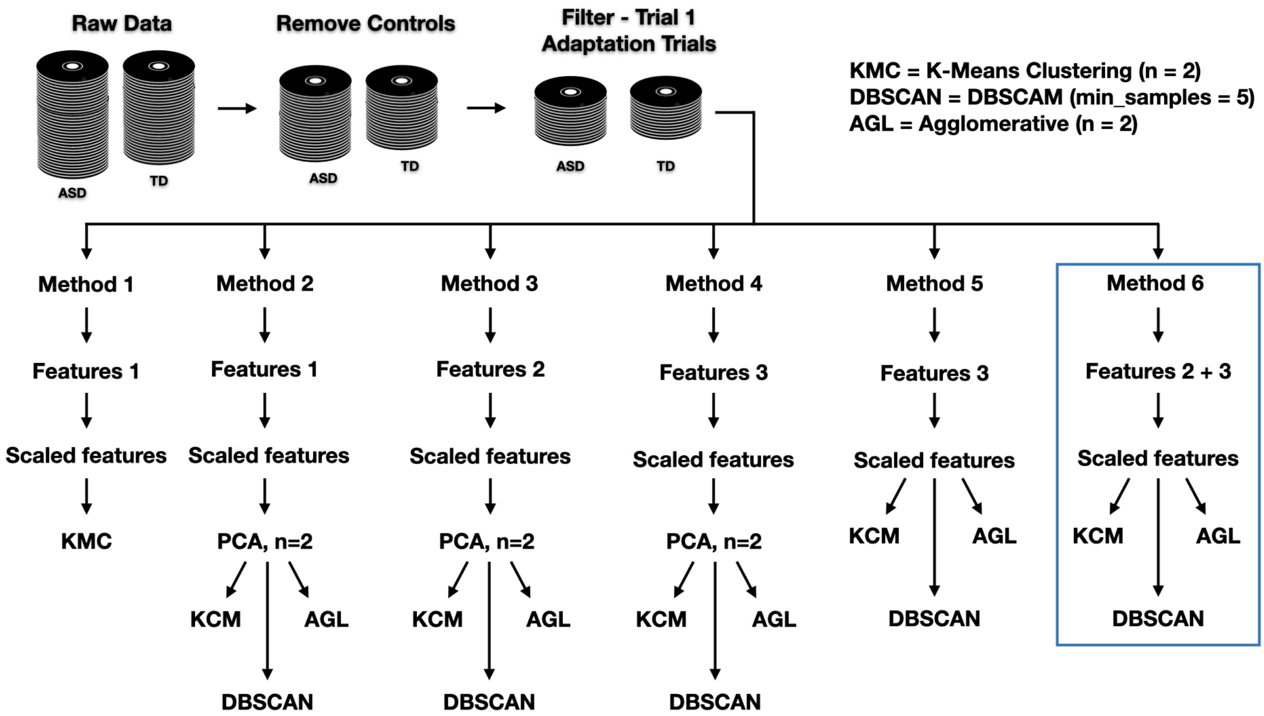

Fig. 2. Unsupervised Learning Methods Overview

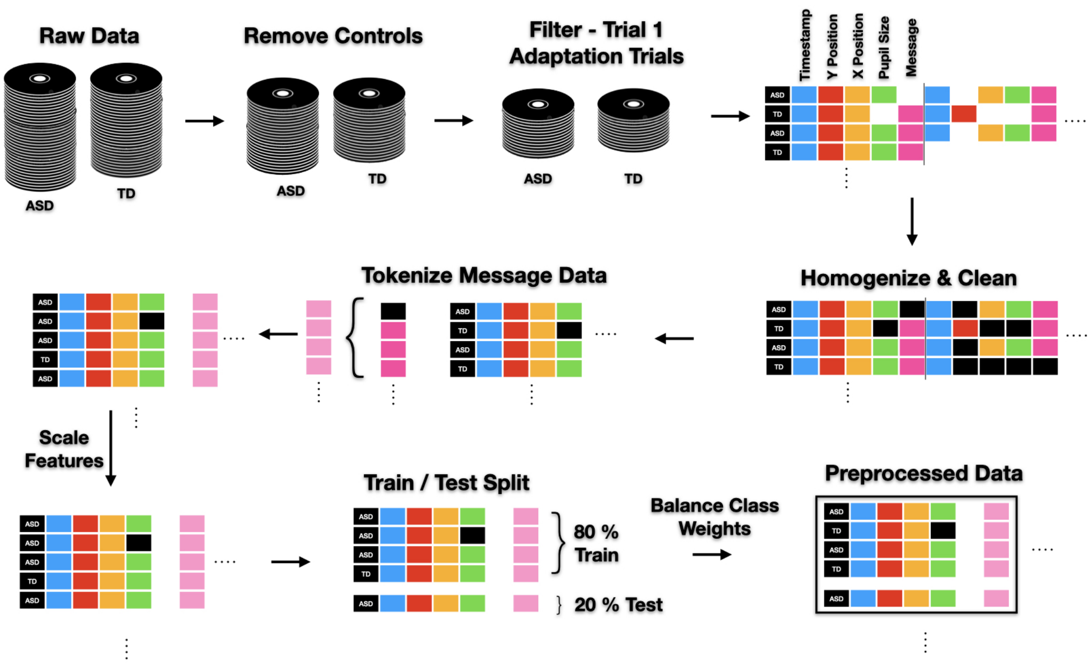

Fig. 3. Preprocessing Method for Supervised Learning

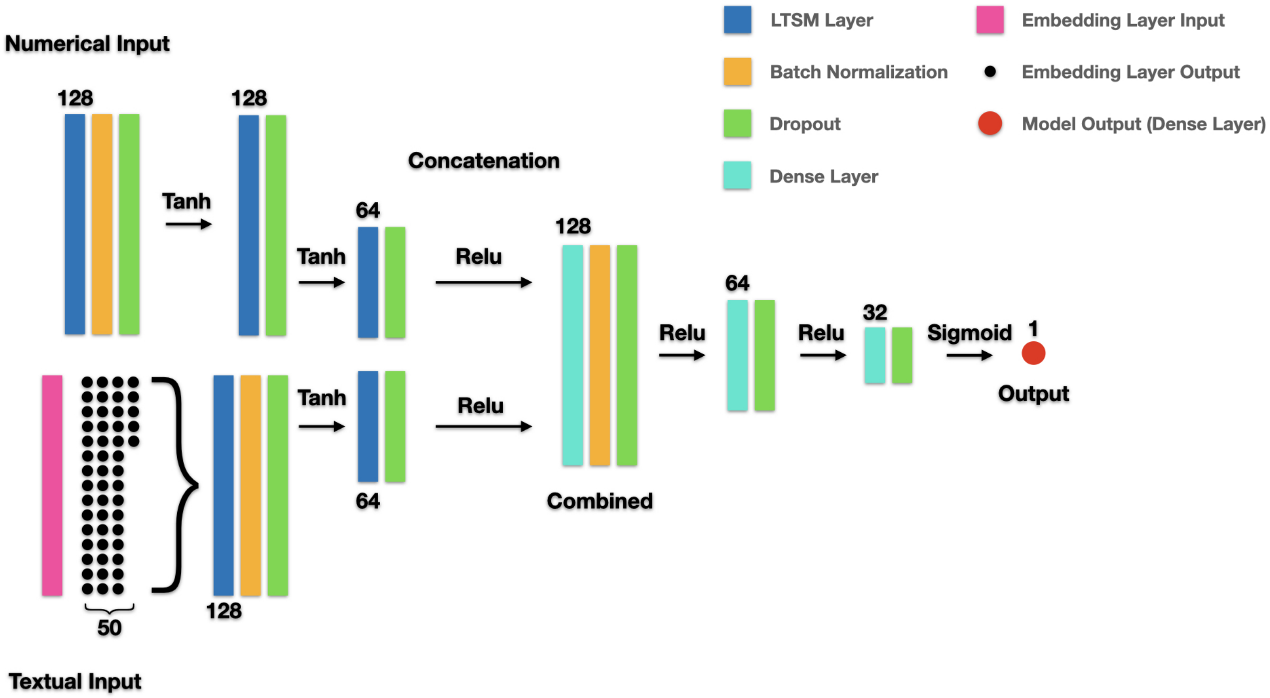

Fig. 4. Supervised Learning Model Overview

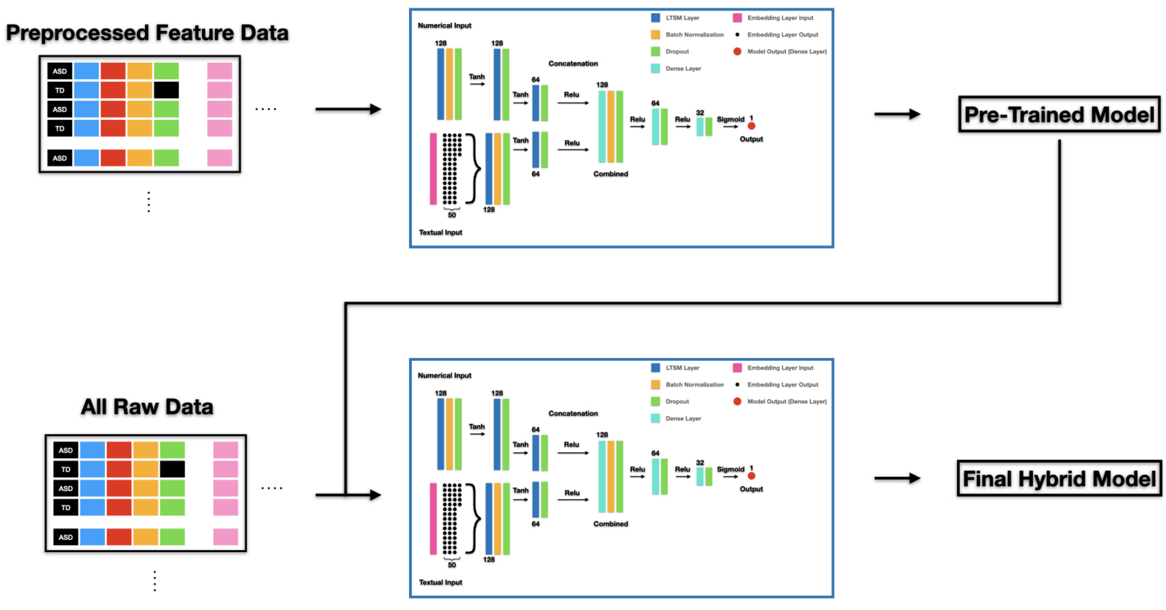

Fig. 5. Hybrid Method Overview

2 and 100% of the TD participants in group 1, and it would still be considered a perfect model, because the ASD and TD participants are totally separated in the clustering.

For each supervised learning method, accuracy, precision, recall, and F1 were recorded and are shown in the results in Table III.

The accuracy measures the proportion of all predictions that each model got correct. It is the number of correctly classified instances (both true ASD and true TD) divided by the total number of instances. Although it is a standard metric and provides a simple snapshot of overall performance, accuracy can be misleading if the dataset is imbalanced. This dataset is somewhat imbalanced, with 62.4% ASD participants. Thus, the All ASD comparison row is shown in Table III, exhibits what each metric would produce if the model simply guessed ASD every time (62.4% accuracy, but 100% precision and recall). Without accuracy, a model guessing ASD on every occasion would appear perfect due to the way precision and recall are calculated, thus accuracy is included as an important metric.

Precision is also included because it focuses on the quality of positive predictions. It is the ratio of true positives to all instances predicted as positive (true positives + false positives, where a positive is an ASD prediction). High precision shows that when the model predicts a positive class, it is very likely correct. Precision is useful when the cost of false positives is high; however, in the use case of a potential ASD diagnosis tool, our aim is not to maximize precision. However, it is still an important metric to show how many TD participants are being falsely labeled as ASD by the model, and can be compared with accuracy to better understand how the model is classifying participants.

Recall measures the model's ability to capture all the positive instances. It is the ratio of true positives to all actual positives (true positives + false negatives). A high recall means that the model rarely misses positive cases. This is likely the most important metric in a prescreening diagnosis tool, which is why it is included in this study. In this use case, this metric should be optimized for, which would result in minimal ASD participants being misdiagnosed as TD.

Lastly, the F1 score is the harmonic mean of precision and recall. It combines both into a single metric, providing a balance between them. This is particularly helpful when you need a single measure to compare models and when neither precision nor recall alone sufficiently captures the success of the model, especially in situations with class imbalance. Thus, it is included here as a supplementary comparison metric across each model.

> RQ1: Can supervised and unsupervised machine learning models help in differentiating eye movements of TD individuals from individuals with ASD?

The results show that none of three unsupervised approaches were able to achieve a meaningful level of accuracy. Table II summarizes the results of the unsupervised methods. Depending on the relative importance of false positives/negatives, a specific method may be slightly more suitable; however, no unsupervised approach in this paper was able to provide a reliable grouping of ASD and TD.

Table III summarizes the results of each supervised approach. The model trained on predefined features that some of the literature had shown to differentiate ASD and TD performed the worst, with 68.9% accuracy compared to 62.4% which would occur by a model that predicted ASD for every participant. However, the deep learning approach showed far more promising results with an 84.2% accuracy and 85.0% precision. This is much closer to the extremely high level of accuracy found in the literature for supervised methods.

RQ1 Takeaway: Based on the results, current unsupervised machine learning methods may be inadequate in differentiating ASD and TD saccadic eye movements. However, labeled data provided to supervised machine learning approaches can provide meaningful levels of accuracy and precision based on the results in this study. This is generally consistent with the literature; however, by performing both methods on the same dataset, this study shows that dataset differences between supervised and unsupervised methods in other studies may not be a factor in the under-performance of unsupervised methods.

TABLE II
UNSUPERVISED RESULTS

Method	Actual ASD		Actual TD	
	Group 1	Group 2	Group 1	Group 2
Perfect (Ex.)	100%	0%	0%	100%
K-Means	44%	56%	22%	78%
DBSCAN	39%	61%	18%	82%
Agglomerative	80%	20%	96%	4%

TABLE III
SUPERVISED RESULTS

Method	Accuracy	Precision	Recall	F1
Perfect	100%	100%	100%	1.0
All ASD	62.4%	100%	100%	1.0
Feature	68.9%	75.5%	100%	0.86
Deep	84.2%	85.0%	94.7%	0.90
Hybrid	85.1%	86.0%	95.2%	0.90

> RQ2: Can a hybrid model approach combining feature engineering and end-to-end deep learning improve the detection of ASD-related differences in eye-tracking data compared to ML models that use either feature engineering or deep learning alone?

The overall best performing model in this study was the hybrid model pre-trained on specific identified features and then trained further on the raw data. This model had an accuracy level of 85.1% and a precision of 86.0%. The model trained only on extracted features achieved just 68.9% accuracy, and the model trained on all the raw data with no feature engineering achieved 84.2% accuracy. Thus, the hybrid approach performed far better than the feature-based approach (85.1% vs 68.9%), but only slightly better than the deep learning approach (85.1% vs 84.2%).

That being said, the feature model was the only model that was able to detect every single ASD participant (100% recall). Although, the practical use of that model may be overstated when looking at recall alone, considering the relatively low accuracy and precision. Thus, depending on what the priorities are for an end user, there is a balance to be struck between the weight of feature engineering and deep learning. If the goal is to maximize the detection of ASD, the results show that feature engineering is superior with 100% recall; however, if the goal is to detect the vast majority of ASD while balancing other costs associated with an excessively high number of falsely labeled TD participants, then the hybrid approach is clearly superior with still high (95.2%) recall, but also much higher precision and accuracy.

RQ2 Takeaway: Since the hybrid approach performed much better than the feature-based approach and only slightly better than the deep learning approach, it can be inferred that the features extracted from the feature-based approach may be somewhat irrelevant compared to the unknown features being learned by the deep learning approach. Furthermore, given the minimal difference between the hybrid and deep learning approaches, it may be inferred that the features extracted for the feature-based approach are mostly learned by the deep learning approach in addition to some other unknown features. Thus, there may exist some overlap in the learned features between feature-based and deep learning approaches in the literature, but this study shows that deep learning approaches may be much more comprehensive in the features that are learned. That being said, it is clear that a hybrid approach, combined known features with deep learning, can improve the detection of ASD-related differences in eye-tracking data compared to either approach alone.

V. Threats to Validity

This study faces several threats to validity that may impact the generalizability and reliability of its findings:

Dataset Bias: The dataset used in this research consists of eye-tracking data from completing a very specific task designed to record the process of saccadic adaptation, which may not represent eye movements in normal settings.

Preprocessing Limitations: The preprocessing steps, such as filtering trials and scaling features, may inadvertently remove important information or introduce artifacts that affect model performance. The exclusion of control trials, for instance, focuses the analysis on adaptation trials but might overlook other relevant patterns. Although done intentionally, training a model with the same architecture without excluding any trials should be done to determine the importance of the control trials.

Computational Resource Limitations: Limited access to the University of Hawaii HPC (AKA Koa) forced the research team to downsize data and reduce model complexity. This data exclusion and simplified model may have reduced the level of success that the supervised deep-learning learning approaches attained.

Model Generalizability: The machine learning models were trained and tested on a single dataset with a specific experimental setup. Differences in eye-tracking hardware, recording conditions, or trial designs in other studies could limit the generalizability of the models to new datasets.

Unsupervised Learning Challenges: The unsupervised learning methods applied in this study showed limited success in differentiating ASD and TD groups. This could be due to the inherent difficulty of the clustering task, the chosen metrics, or the high dimensionality of the data, which may not lend itself well to unsupervised analysis without additional dimensionality reduction techniques.

Evaluation Metrics: The accuracy, precision, and recall metrics used to evaluate model performance provide valuable insights but may not fully capture the nuances of the classification task. Future studies could benefit from exploring additional evaluation metrics, such as the area under the ROC curve (AUC), to provide a more comprehensive assessment.

Unsupervised Clustering Parameters: Given the many possible binary splits in the data (male/female, old/young, ASD/TD, etc.), it could be that there are more prominent differences in eye movement features between groups in a different binary split other than ASD/TD (whether that be age, gender, or a different split in the data). Since the unsupervised methods didn't perform well grouping ASD and TD, it may be that the data was being grouped by a binary characteristic other than ASD/TD. In future research, the data should be processed to remove as many known binary splits as possible, to leave only ASD vs TD in the data, before attempting to cluster the participants into two groups.

Addressing these threats requires careful consideration in future work, including validating the findings on diverse datasets, refining preprocessing methods, securing longer-term and more consistent access to high-performance computing resources, and exploring alternative model architectures and evaluation strategies.

VI. Future Work

The findings in this study open several avenues for further research and development:

Future work should explore the use of more sophisticated machine learning and deep learning models to improve classification accuracy. Approaches like transformer-based models or ensemble techniques might better capture the nuances of eye-tracking data. Incorporating diverse datasets with varied demographics, eye-tracking tasks, and experimental settings would ensure broader generalizability and mitigate dataset-specific biases.

Expanding the set of features extracted from eye-tracking data, such as incorporating temporal dynamics or more advanced saccadic metrics, may improve model performance.

Developing methods to interpret deep learning models, such as using SHAP (Shapley additive explanations) values or attention maps, would provide insights into which features or data segments drive model predictions. Given the superior performance of deep learning compared to manual feature extraction, a focus on the explainability of deep learning is especially important.

Further refinement of the hybrid model approach could help balance the strengths of feature-based and raw data-based learning. Experimentation with different pre-training strategies and fine-tuning techniques could yield better results.

Perhaps most importantly, securing long-term access to high-performance computing could enable the timely training of a more complex supervised model and could avoid the data exclusion procedure necessary in this paper. This alone may result in a more highly accurate machine learning model.

Investigating how different eye-tracking tasks or experimental designs impact the performance of ML models could provide insights into task-specific strengths of eye-tracking as a diagnostic tool. Developing models that can adapt to multiple tasks or dynamically select task-relevant features may enhance versatility.

Combining eye-tracking data with other biometric or behavioral data, such as EEG, motion tracking, or physiological measures, could provide a more comprehensive understanding of ASD-related differences. Multi-modal learning frameworks could be applied to integrate these diverse data types effectively.

Addressing the challenges faced by unsupervised learning in this study by experimenting with advanced clustering methods or autoencoders could yield better grouping results. There are also semi-supervised approaches that leverage a mix of labeled and unlabeled data that may bridge the gap between unsupervised and supervised learning that has not been applied to this type of data in the literature.

Investigating the generalizability of the models in this study and the literature is also necessary. Transfer learning should be used to train a subset of successful pre-trained models from the literature and evaluated on a new dataset to determine how well each model generalizes to new sets and types of eye-tracking data.

Translating research findings into practical diagnostic tools requires exploring lightweight models that can run on portable or low-resource devices. Developing user-friendly interfaces for clinicians to visualize and interact with model outputs could improve adoption in healthcare settings. Research should be done to produce a product that can deliver the promising findings of this and related research to the public.

By addressing these areas, future studies can build on the insights of this research to enhance the utility of machine learning and eye-tracking data in understanding and diagnosing Autism Spectrum Disorder.

VII. Acknowledgments

The SR support forum and SR research support and information specialists are gratefully acknowledged.

The technical support and advanced computing resources from the University of Hawaii Information Technology Services – Cyberinfrastructure, funded in part by the National Science Foundation CC* awards # 2201428 and # 2232862 are gratefully acknowledged.

References

[1] B. Johnson, N. Rinehart, O. White, L. Millist, and J. Fielding, "Saccade adaptation in autism and asperger's disorder," *Neuroscience*, vol. 243, pp. 76–87, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306452213003060

[2] M. W. Mosconi, B. Luna, M. Kay-Stacey, C. V. Nowinski, L. H. Rubin, C. Scudder, N. Minshew, and J. A. Sweeney, "Saccade adaptation abnormalities implicate dysfunction of cerebellar-dependent learning mechanisms in autism spectrum disorders (asd)," *PloS one*, vol. 8, no. 5, p. e63709, 2013.

[3] L. M. Schmitt, E. H. Cook, J. A. Sweeney *et al.*, "Saccadic eye movement abnormalities in autism spectrum disorder indicate dysfunctions in cerebellum and brainstem," *Molecular Autism*, vol. 5, p. 47, 2014. [Online]. Available: https://doi.org/10.1186/2040-2392-5-47

[4] K. Tarrit, E. G. Freedman, A. A. Francisco, D. J. Horsthuis, S. Molholm, and J. J. Foxe, "No evidence for differential saccadic adaptation in children and adults with an autism spectrum diagnosis," *Frontiers in Integrative Neuroscience*, vol. 17, 2023. [Online]. Available: https://www.frontiersin.org/journals/integrative-neuroscience/articles/10.3389/fnint.2023.1232474

[5] R. Asmetha Jeyarani and R. Senthilkumar, "Eye tracking biomarkers for autism spectrum disorder detection using machine learning and deep learning techniques: Review," *Research in Autism Spectrum Disorders*, vol. 108, p. 102228, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1750946723001289

[6] F. Cilia, R. Carette, M. Elbattah, G. Dequen, J.-L. Guérin, J. Bosche, L. Vandromme, and B. Le Driant, "Computer-aided screening of autism spectrum disorder: Eye-tracking study using data visualization and deep learning," *JMIR Hum Factors*, vol. 8, no. 4, p. e27706, Oct 2021. [Online]. Available: https://humanfactors.jmir.org/2021/4/e27706

[7] I. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. H. Ali, H. S. A. Shatnawi, S. M. Alwazer, and M. Alshahrani, "Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques," *Electronics*, vol. 11, no. 4, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/4/530

[8] Z. Ahmed and M. E. Jadhav, "Convolutional neural network for prediction of autism based on eye-tracking scanpaths," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 05, 2020.

[9] A. Gaspar, D. Oliva, S. Hinojosa, I. Aranguren, and D. Zaldivar, "An optimized kernel extreme learning machine for the classification of the autism spectrum disorder by using gaze tracking images," *Applied Soft Computing*, vol. 120, p. 108654, 2022.

[10] M. Elbattah, J.-L. Guérin, R. Carette, F. Cilia, and G. Dequen, "Nlp-based approach to detect autism spectrum

disorder in saccadic eye movement," pp. 1581–1587, 2020.

[11] J. H. Yoo, C. Kang, J. S. Lim, B. Wang, C.-H. Choi, H. Hwang, D. H. Han, H. Kim, H. Cheon, and J.-W. Kim, "Development of an innovative approach using portable eye tracking to assist adhd screening: a machine learning study," *Frontiers in Psychiatry*, vol. 15, p. 1337595, 2024.

[12] M. Elbattah, R. Carette, F. Cilia, J.-L. Guérin, and G. Dequen, "Chapter 5 - applications of machine learning methods to assist the diagnosis of autism spectrum disorder," in *Neural Engineering Techniques for Autism Spectrum Disorder*, A. S. El-Baz and J. S. Suri, Eds. Academic Press, 2023, pp. 99–119. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128244210000138

[13] "Sr support forum," https://www.sr-research.com/support/thread-7675.html, accessed: 2024-10-06.

[14] P. L. Howard, L. Zhang, and V. Benson, "What can eye movements tell us about subtle cognitive processing differences in autism?" *Vision*, vol. 3, no. 2, 2019. [Online]. Available: https://www.mdpi.com/2411-5150/3/2/22

[15] M. Sadria, S. Karimi, and A. T. Layton, "Network centrality analysis of eye-gaze data in autism spectrum disorder," *Computers in Biology and Medicine*, vol. 111, p. 103332, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001048251930201X

[16] C. Wu, S. Liaqat, S.-c. Cheung, C.-N. Chuah, and S. Ozonoff, "Predicting autism diagnosis using image with fixations and synthetic saccade patterns," pp. 647–650, 2019.

[17] D. Roth, M. Jording, T. Schmee, P. Kullmann, N. Navab, and K. Vogeley, "Towards computer aided diagnosis of autism spectrum disorder using virtual environments," in *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, 2020, pp. 115–122.

[18] M. Lai, J. Lee, S. Chiu, J. Charm, W. Y. So, F. P. Yuen, C. Kwok, J. Tsoi, Y. Lin, and B. Zee, "A machine learning approach for retinal images analysis as an objective screening method for children with autism spectrum disorder," *EClinicalMedicine*, vol. 28, p. 100588, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2589537020303321

[19] S. Kumar R and D. M, "Differential evolution tuned support vector machine for autistic spectrum disorder diagnosis," p. 3861–3870, Jul. 2019. [Online]. Available: http://dx.doi.org/10.35940/ijrte.B3063.078219

[20] G. Wan, X. Kong, B. Sun, S. Yu, Y. Tu, J. Park, C. Lang, M. Koh, Z. Wei, Z. Feng, Y. Lin, and J. Kong, "Applying eye tracking to identify autism spectrum disorder in children," *Journal of Autism and Developmental Disorders*, vol. 49, no. 1, pp. 209–215, Jan 2019. [Online]. Available: https://doi.org/10.1007/s10803-018-3690-y

[21] J. Kang, X. Han, J. Song, Z. Niu, and X. Li, "The identification of children with autism spectrum disorder by svm approach on eeg and eye-tracking data," *Computers in Biology and Medicine*, vol. 120, p. 103722, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482520301074

[22] J. S. Oliveira, F. O. Franco, M. C. Revers, A. F. Silva, J. Portolese, H. Brentani, A. Machado-Lima, and F. L. S. Nunes, "Computer-aided autism diagnosis based on visual attention models using eye tracking," *Scientific Reports*, vol. 11, no. 1, p. 10131, May 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-89023-8

[23] Z. Zhao, H. Tang, X. Zhang, X. Qu, X. Hu, and J. Lu, "Classification of children with autism and typical development using eye-tracking data from face-to-face conversations: Machine learning model development and performance evaluation," *J Med Internet Res*, vol. 23, no. 8, p. e29328, Aug 2021. [Online]. Available: https://www.jmir.org/2021/8/e29328

[24] M. Alcañiz, I. A. Chicchi-Giglioli, L. A. Carrasco-Ribelles, J. Marín-Morales, M. E. Minissi, G. Teruel-García, M. Sirera, and L. Abad, "Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: A proof of concept for diagnosis," *Autism Research*, vol. 15, no. 1, pp. 131–145, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aur.2636

[25] T. Vu, H. Tran, K. W. Cho, C. Song, F. Lin, C. W. Chen, M. Hartley-McAndrew, K. R. Doody, and W. Xu, "Effective and efficient visual stimuli design for quantitative autism screening: An exploratory study," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 2017, pp. 297–300.

[26] Y. Fang, H. Duan, F. Shi, X. Min, and G. Zhai, "Identifying children with autism spectrum disorder based on gaze-following," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 423–427.

[27] X.-J. Kong, Z. Wei, B. Sun, Y. Tu, Y. Huang, M. Cheng, S. Yu, G. Wilson, J. Park, Z. Feng, M. Vangel, J. Kong, and G. Wan, "Different eye tracking patterns in autism spectrum disorder in toddler and preschool children," *Frontiers in Psychiatry*, vol. 13, 2022. [Online]. Available: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2022.899521

Appendix A
Related Work Papers

Ref.	Summary / Findings	Research Questions / Goals
[1]	1. Adaptation of saccade gain is slower in high-functioning autism (HFA). 2. Longer corrective saccade latencies suggest poor utilization of visual feedback in both HFAs. 3. Findings support protracted motor learning in autism, but not Asperger's disorder	RQ 1. Do children with high-functioning autism (HFA) show differences in saccade adaptation and motor learning compared to typically developing (TD) children?
[2]	1. Slower Saccade Adaptation: Individuals with ASD adapted their saccade amplitudes more slowly than healthy controls, with 30% of ASD individuals failing to adapt significantly, compared to only 6% of controls. 2. Increased Amplitude Variability: There was greater trial-to-trial variability in saccade amplitude across baseline, adaptation, and recovery phases in ASD individuals, indicating reduced consistency in saccade accuracy. 3. Association with Manual Motor Control: Impairments in saccade adaptation and increased amplitude variability were linked to poorer performance on a manual motor test, suggesting a broader motor control deficit in ASD. 4. Impaired Neural Plasticity in Cerebellar Circuits: The pattern of impaired adaptation and saccade variability in ASD indicates reduced neural plasticity within the learning circuits of the oculomotor vermis, supporting functional abnormalities in the cerebellar vermis that are consistent with postmortem and neuroimaging studies of ASD.	RQ 1. Are there differences in saccade adaptation rates and variability between individuals with Autism Spectrum Disorders (ASD) and healthy controls? RQ 2. Is impaired saccade adaptation in individuals with ASD related to abnormalities in the cerebellar vermis, and how does this relate to motor control impairments in both oculomotor and manual motor systems?
[3]	1. Reduced Saccade Accuracy and Increased Variability: Individuals with ASD demonstrated reduced accuracy in saccades and greater trial-to-trial variability compared to healthy controls, particularly when making larger saccades. 2. Altered Saccade Dynamics: Saccades in ASD were characterized by lower peak velocity and prolonged duration. Specifically, individuals with ASD took longer to accelerate to peak velocity but showed no difference in deceleration duration. 3. Similar Latency Responses: While saccade latencies were similar across ASD and control groups, individuals with ASD exhibited greater variability in these latencies across trials. 4. No Deficit in Visual Orienting and Attention: Gap and overlap paradigms revealed no significant differences in latency effects between groups, suggesting that basic visual orienting and attention systems are relatively intact in ASD. 5. Age-Related Improvements: Both ASD and control groups showed similar age-related improvements in saccade performance (accuracy and latency variability), indicating that developmental trajectories are comparable across groups.	RQ 1. How do saccadic eye movements (latency, accuracy, and dynamics) differ between individuals with Autism Spectrum Disorder (ASD) and healthy controls? RQ 2. Are there abnormalities in the functional integrity of cerebellar and brainstem circuitry related to the sensorimotor control of saccades in ASD?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[4]	1. No Differences in Saccadic Adaptation: Children and adults with ASD showed similar abilities to adapt saccades in response to experimentally introduced visual errors compared to TD peers. The rate of adaptation was comparable across ASD and TD groups during both early and late adaptation phases. 2. Similar Saccade Variability: The study found no significant differences in within-participant variability of saccade amplitude in ASD compared to TD individuals, suggesting that saccade variability is not increased in ASD as some previous studies reported. 3. Hypermetric Saccades in ASD: A post hoc analysis indicated that individuals with ASD tended to make slightly longer (hypermetric) saccades to non-adapted targets than TD participants. However, the effect size was small and requires further replication. 4. No Developmental Differences: The study found no evidence that saccadic adaptation differed between children and adults within the ASD group, indicating that the adaptation process is not significantly influenced by age in this population.	RQ 1. Does saccadic adaptation (a measure of eye movement adjustments in response to visual errors) differ between individuals with Autism Spectrum Disorder (ASD) and typically developing (TD) individuals? RQ 2. Are there developmental differences in saccadic adaptation across children and adults with ASD?
[5]	1. Eye Tracking as a Biomarker: Eye tracking is a promising tool for assessing social attention and visual processing anomalies in ASD, providing measurable indicators of social communication deficits. 2. ML and DL Model Performance: ML models such as SVM, Random Forest, and KNN have shown considerable accuracy (up to 100% in some studies) in classifying ASD, with SVM being the most commonly used. DL models like CNN and LSTM effectively handle complex data and temporal sequences for ASD detection. 3. Variability in Accuracy: There is considerable variation in the accuracy of ASD classification across different studies, largely depending on the dataset size, eye-tracking metrics used, and chosen model. 4. Gaze Prediction for Interventions: Eye-tracking data combined with ML/DL techniques hold the potential for designing personalized intervention strategies, predicting gaze behavior, and assessing the effectiveness of ASD therapies over time.	RQ 1. Can eye-tracking data serve as reliable biomarkers for Autism Spectrum Disorder (ASD)? RQ 2. How effective are machine learning (ML) and deep learning (DL) techniques in differentiating individuals with ASD from typically developing (TD) peers using eye-tracking data? RQ 3. What are the strengths and limitations of various ML and DL models when applied to ASD classification based on eye movement patterns? RQ 4. How can eye-tracking data be leveraged for early diagnosis and intervention design in ASD?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[6]	1. Visual Representation of Scanpaths: Eye-tracking scan paths can be transformed into visual representations, effectively encoding gaze dynamics (e.g., velocity) using color gradients. These visualizations allow the classification task to be framed as an image classification problem. 2. High Classification Accuracy: A convolutional neural network (CNN) was able to achieve a high classification accuracy (90%) for distinguishing between ASD and non-ASD participants based on these visual scan path representations, suggesting the approach's effectiveness in ASD screening. 3. Correlation with Autism Severity: The study found a strong correlation between the CARS scores (measuring autism severity) and eye movement velocity, indicating that eye movement dynamics are indicative of the level of autism symptoms. 4. Generalizability to Other Disorders: The approach demonstrates the potential to be transferable to the screening of other neurodevelopmental disorders by using eye-tracking data, visualization, and deep learning techniques. 5. Practical Application and Parental Acceptance: Eye-tracking measures are seen as a practical tool that can be quickly managed during diagnostic interviews, with high acceptance from parents due to the clarity of the visual results. However, costs related to hardware and software could be a limitation for widespread clinical adoption.	RQ 1. How can eye-tracking technology be effectively combined with visualization and deep learning to assist in the early diagnosis of Autism Spectrum Disorder (ASD)? RQ 2. Can visual representations of eye-tracking scan paths be used as a reliable feature for the classification of ASD? RQ 3. Is there a correlation between the severity of autism, as measured by the Childhood Autism Rating Scale (CARS), and the dynamics of eye movements? RQ 4. Can the approach of integrating eye-tracking data with machine learning be generalized to screen for other neurodevelopmental disorders?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[7]	1. High Accuracy in Different Models: 1.1 Neural Networks (ANN and FFNN): Achieved the highest accuracy for ASD diagnosis, reaching 99.8%. This was based on the classification of eye-tracking data using features extracted through a hybrid of LBP and GLCM. 1.2 Pre-trained CNN Models (GoogleNet and ResNet-18): Showed strong classification performance with accuracies of 93.6 1.3 Hybrid Approach (CNN + SVM): Combining deep learning models (GoogleNet and ResNet-18) with a Support Vector Machine (SVM) classifier yielded accuracies of 95.5 2. Feature Extraction and Data Processing: 2.1 A hybrid method combining LBP and GLCM algorithms effectively extracted critical features from eye-tracking data, contributing to the high performance of the neural networks. 2.2 Image enhancement techniques (e.g., average and Laplacian filters) were applied to optimize images before feature extraction and classification. 3. Comparison of Methods: 3.1 The study concluded that neural networks (FFNN and ANN) outperformed both the pre-trained CNN models and the hybrid deep learning-machine learning approaches in terms of accuracy and overall performance. 3.2 The ResNet-18 model demonstrated better accuracy compared to GoogleNet, and the GoogleNet + SVM hybrid technique achieved slightly better accuracy than the ResNet-18 + SVM.	RQ 1. How can eye-tracking data be used effectively in conjunction with machine learning and deep learning techniques for the early diagnosis of Autism Spectrum Disorder (ASD)? RQ 2. What is the impact of using different artificial intelligence techniques (neural networks, convolutional neural networks, and hybrid approaches) on the accuracy of ASD classification? RQ 3. How do the combinations of feature extraction methods like Local Binary Pattern (LBP) and Grey Level Co-occurrence Matrix (GLCM) improve the performance of neural networks in classifying ASD from typically developing (TD) cases? RQ 4. Can a hybrid approach combining deep learning and machine learning enhance the efficiency and accuracy of ASD diagnosis based on eye-tracking data?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[8]	1. Deep Learning Model Development: 1.1 A deep learning model utilizing a Convolutional Neural Network (CNN) was developed to classify ASD and TD children based on eye-tracking scan paths. 1.2 The dataset consisted of 59 participants: 29 ASD children (25 males, 4 females) and 30 TD children (13 males, 17 females), with an average age of around 8 years. This dataset contained 547 eye-tracking scanpath images. 2. Data Augmentation and Pre-Processing: 2.1 Image augmentation techniques, such as rotation and shearing, were applied to reduce overfitting due to the small sample size, creating additional synthetic data. 2.2 Images were resized and converted to grayscale to reduce the computational complexity of the model. 3. CNN Architecture and Training: 3.1 The CNN architecture consisted of four convolutional layers followed by max pooling and one fully connected layer. The network utilized ReLU as the activation function and used a sigmoid function for binary classification. 3.2 The model was trained for 50 epochs using a batch size of 32, with the Adam optimizer set to a learning rate of 0.001 and a dropout rate of 0.20 to prevent overfitting. 4. Experimental Results: 4.1 The model achieved an accuracy of 98% when tested on 30% of the data, outperforming previously reported results on the same dataset. 4.2 A confusion matrix analysis revealed that the model performed well in distinguishing between ASD and TD scan paths, with a high number of true positive and true negative classifications.	RQ 1. Can eye-tracking scan paths be used to distinguish between children with Autism Spectrum Disorder (ASD) and typically developing (TD) children using deep learning models? RQ 2. How effective is a convolutional neural network (CNN) model in classifying children as ASD or TD based on their eye-tracking data?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[9]	1. Proposed GPC-KELM Methodology: 1.1 A novel machine learning approach, GPC-KELM, was introduced for ASD classification using gaze-tracking data. 1.2 The Kernel Extreme Learning Machine (KELM) was optimized using the Giza Pyramids Construction (GPC) algorithm to enhance the accuracy of ASD classification. 2. High Classification Accuracy: 2.1 The methodology achieved an average accuracy of 98.8% in classifying ASD subjects using gaze-tracking images. 2.2 The GPC algorithm outperformed other optimization techniques such as Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), Ant Lion Optimizer (ALO), Bat Algorithm (BA), and Harris Hawks Optimizer (HHO) when applied to KELM. 3. Data Processing and Validation: 3.1 The GPC-KELM approach included data augmentation, dimensionality reduction, and normalization steps to ensure accurate classification. 3.2 Statistical tests and analyses were performed to validate the methodology, demonstrating its robustness in ASD classification based on eye movement patterns. 4. Comparison with Other Machine Learning Approaches: The GPC-KELM model was compared to traditional machine learning techniques such as Naive Bayes, Logistic Regression, and Artificial Neural Networks (ANN), demonstrating superior performance in ASD classification.	RQ 1. Can a Kernel Extreme Learning Machine (KELM) model be effectively optimized for classifying Autism Spectrum Disorder (ASD) using gaze-tracking images? RQ 2. How does the Giza Pyramids Construction (GPC) algorithm improve the performance of KELM in ASD classification compared to other optimization methods?
[10]	1. NLP-Based Transformation of Eye-Tracking Data: 1.1 Eye-tracking records, which consist of saccades and fixations, were transformed into textual strings describing the sequences of eye movements using NLP techniques. 1.2 This transformation allowed for the use of sequence-based classification models to predict ASD. 2. Classification Models and Performance: 2.1 Standard Convolutional Neural Network (ConvNet) and Long Short-Term Memory (LSTM) models were trained on the transformed sequences. 2.2 The ConvNet models consistently outperformed the LSTM models, achieving an ROC-AUC of up to 0.84, suggesting that the sequence-based representation of eye movements is a viable feature for ASD classification.	RQ 1. Can a sequence-learning approach using saccadic eye movements be effective in classifying Autism Spectrum Disorder (ASD)? RQ 2. How can Natural Language Processing (NLP) techniques transform eye-tracking data into a sequence-based representation suitable for machine learning (ML) classification models?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[11]	1. Eye-Tracking Features and ADHD Classification: 1.1 33 eye-tracking features were identified across five tasks (pro-saccade, anti-saccade, memory-guided saccade, change detection, and Stroop tasks) that could distinguish between children with ADHD and typically developing children (TDC). 1.2 Participants with ADHD showed increased saccade latency and degree, and shorter fixation time compared to TDC. 2. Machine Learning Model Performance: 2.1 A soft voting model integrating extra tree and random forest classifiers achieved a high accuracy of 76.3% in identifying ADHD using eye-tracking features alone. 2.2 When comparing the model based only on eye-tracking features with models using conventional screening methods like the Advanced Test of Attention (ATA) or Stroop test, there was no significant difference in the area under the curve (AUC). 3. Ensemble Models and Improvement: The integration of demographic, behavioral, and clinical data with eye-tracking features improved the accuracy of classification but did not significantly alter the AUC, indicating that eye-tracking data alone is a robust feature set for ADHD classification.	RQ 1. Can eye-tracking data and machine learning be used to develop a reliable screening model to classify ADHD in children? RQ 2. How do different eye-tracking tasks related to selective attention, working memory, and response inhibition contribute to the identification of ADHD using machine learning models?
[12]	1. The study developed a method to convert eye-tracking scan paths into compact image formats that visually encode gaze movements and their dynamics (such as velocity) using color gradients. 2. By applying unsupervised ML techniques like clustering on these visual representations, the researchers discovered inherent patterns that could differentiate between ASD-diagnosed individuals and typically developing (TD) participants. The clusters revealed correlations related to gaze behavior dynamics, such as higher gaze velocity and acceleration in ASD participants. 3. Supervised ML models, particularly convolutional neural networks (CNNs), were trained on the scan path images to classify ASD. The CNN model achieved a high prediction accuracy with a Receiver Operating Characteristic Area Under the Curve (ROC-AUC) of approximately 0.9, demonstrating the effectiveness of this approach. 4. The experimental results indicate that integrating ML with eye-tracking data holds significant promise for developing data-driven techniques to assist in the early and accurate diagnosis of ASD. The approach offers a non-invasive, efficient means of capturing and analyzing behavioral gaze patterns associated with autism.	RQ 1. Can machine learning (ML) techniques be effectively applied to eye-tracking data to assist in the diagnosis of Autism Spectrum Disorder (ASD)? RQ 2. Is it possible to represent gaze patterns visually and use these representations for both supervised and unsupervised ML models to detect ASD-related behaviors? RQ 3. What are the potentials and limitations of integrating ML and eye-tracking technologies in supporting the diagnostic processes of ASD?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[14]	1. Eye-tracking studies reveal that while basic oculomotor control in individuals with ASD is intact, there are subtle processing differences, especially in attention allocation during social and complex tasks. 2. Atypical gaze patterns in ASD often result in delays in detecting key social cues and differences in processing language and social interactions. 3. Findings highlight the role of increased task complexity and the presence of competing stimuli in influencing attentional differences in ASD.	RQ 1. How do eye movement patterns provide insight into cognitive processing differences in individuals with ASD, particularly in language and social domains?
[15]	1. Betweenness centrality was the most effective network analysis method, identifying significant differences in four AOIs between ASD and TD children. 2. ASD children exhibited significantly longer fixation times on the mouth and shorter times on the eyes compared to TD children, consistent with prior studies. 3. Degree centrality and fixation time analysis alone were less effective in revealing broader differences in eye-gaze patterns.	RQ 1. What network analysis approaches best distinguish the eye-gaze patterns of ASD and TD children?
[16]	1. Two Machine Learning Approaches for ASD Classification: 1.1 The synthetic saccade approach uses a generative model (STAR-FC) to simulate typical non-ASscan paths. These synthetic paths are compared with real scan paths from children using various distance measures, which are then used as features for a deep learning classifier. 1.2 The image-based approach uses a state-of-the-art convolutional neural network (CNN) to classify ASD based on both the input image and the fixation maps generated from the scan path data. 2. Model Performance and Accuracy: 2.1 The synthetic saccade approach achieved an accuracy of 65.41% on the validation dataset. 2.2 The image-based model utilizes a dual-branch CNN architecture to jointly learn features from both the image and the scan path data, leveraging the visual context and fixation sequences to make ASD/TD classifications. 3. Dataset and Experimentation: 3.1 The models were trained on a dataset provided by the "Saliency4ASD" challenge, which includes 300 images viewed by 14 ASD and 14 TD children each, resulting in 5542 scan paths for training and 1411 for testing. 3.2 Due to the small size of the dataset and the need to avoid overfitting, data augmentation was applied, including jittering the color of images and adding random noise to fixation locations and durations.	RQ 1. Can eye-tracking scan paths be used as an effective feature to automatically classify children as having Autism Spectrum Disorder (ASD) or being typically developing (TD)? RQ 2. How do synthetic saccade patterns and image-based deep learning approaches compare in terms of their ability to classify ASD based on eye-tracking data?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[17]	1. A neural network classifier achieved 92.9% accuracy in classifying ASD based on nonverbal behavior recorded in avatar-mediated communication. 2. Gaze behavior, specifically horizontal gaze shifts and focus on the eye region, was identified as a significant discriminator between ASD and typically developing individuals. 3. The proposed system offers potential as a tool to complement clinical ASD diagnostics by providing objective, quantitative data on nonverbal behavior during naturalistic interactions.	RQ 1. Can ASD be classified based on nonverbal behavior (gaze, voice, head motion) recorded through avatar-mediated communication?
[18]	1. Retinal images revealed significant differences in optic disc and cup diameters between ASD and control groups. 2. A machine learning classifier achieved sensitivity of 95.7% and specificity of 91.3 3. The study supports the use of non-invasive retinal image analysis as an objective screening tool for ASD.	RQ 1. Can retinal image analysis be employed as an objective screening method for ASD in children?
[19]	1. A Differential Evolution (DE) tuned Support Vector Machine (SVM) was proposed for classifying autism spectrum disorder (ASD) data, achieving a classification accuracy of 100 2. Feature selection using Sequential Forward Selection (SFS) reduced the dimensionality of the data by 82%, improving computational efficiency. 3, The DE-tuned SVM outperformed Artificial Neural Networks (ANN) and other SVM configurations in accuracy, precision, recall, and F-measures.	RQ 1. Can Differential Evolution optimization improve the performance of Support Vector Machines for accurate ASD classification?
[20]	1. Eye tracking data revealed significantly reduced fixation times for ASD children at the eyes, mouth, and body compared to TD children. 2. Fixation times at the moving mouth and body provided significant discrimination between ASD and TD children, achieving a classification accuracy of 85.1%, sensitivity of 86.5%, and specificity of 83.8%. 3. The study demonstrated that a brief 10-second video could effectively differentiate ASD from TD children, supporting its potential for early detection of ASD.	RQ 1. Can eye-tracking data from short video clips effectively distinguish ASD from TD children and support early ASD detection?
[21]	1. Combining EEG and eye-tracking data for classification achieved 85.44% accuracy with AUC 0.93 using 32 selected features. 2. ASD children exhibited higher theta band power and lower beta and gamma band power in EEG compared to TD children. 3. ASD children fixated less on core facial areas (nose and mouth) and more on the background in eye-tracking tests, consistent with the eye avoidance hypothesis.	RQ 1. Can the combination of EEG and eye-tracking data improve the classification accuracy of ASD and TD children using machine learning methods?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[22]	1. A novel approach using Visual Attention Models (VAMs) based on eye-tracking data and videos was developed to classify ASD and TD groups with an average precision of 90%, specificity of 93%, and sensitivity of 69%. 2. The method uses a Genetic Algorithm for feature selection, identifying biological and geometric movement features as significant for ASD visual attention patterns. 3. The proposed method eliminates the need for manually defined Regions of Interest (ROIs), reducing bias and data loss in ASD classification.	RQ 1. Can Visual Attention Models trained with eye-tracking data and video stimuli classify ASD and TD individuals with high accuracy?
[23]	1. Eye-tracking data collected during face-to-face conversations classified children with ASD and TD with a maximum accuracy of 92.31%, using a support vector machine (SVM). 2. Combining visual fixation features (e.g., mouth and body AOIs) with session length achieved higher classification accuracy compared to using either modality alone. 3. The study suggests eye-tracking during naturalistic interactions as a feasible tool for ASD screening, emphasizing the need for validation in diverse populations.	RQ 1. Can eye-tracking data from face-to-face conversations effectively classify children with ASD and TD, and does combining visual fixation with session length improve classification performance?
[24]	1. A supervised machine learning model combining immersive virtual reality and eye-tracking data achieved 86% accuracy and 91% sensitivity in classifying ASD and TD children. 2. Autistic children showed higher visual attention to adults over children and demonstrated distinct gaze patterns in dynamic, social-rich VR scenarios compared to TD children. 3. This study is the first to integrate immersive VR and eye-tracking data for ASD recognition, offering a proof of concept for more objective and ecologically valid assessments.	RQ 1. Can immersive virtual reality and eye-tracking paradigms combined with machine learning effectively distinguish ASD from TD children based on social attention behaviors?
[25]	1. ”Social scene” stimuli with a 5-second exposure time achieved the highest ASD screening accuracy of 98.24 2. Human face stimuli also performed well, with an accuracy of 97.22%, while object stimuli yielded lower accuracy at 90.26%. 3. The study highlights the importance of both stimulus content and exposure time in optimizing quantitative ASD screening methods.	RQ 1. What is the impact of different visual stimuli and exposure times on the accuracy of quantitative ASD screening?

Continued on next page

TABLE IV – continued from previous page

Ref.	Summary / Findings	Research Questions / Goals
[26]	1. A novel gaze-following dataset (GazeFollow4ASD) was created, including 300 images and corresponding eye movement data from 8 children with ASD and 10 typically developing (TD) children. 2. Proposed an LSTM-based model to extract discriminative features from fixation maps, achieving a classification accuracy of 79.94%. 3. Gaze-following stimuli were shown to effectively differentiate between ASD and TD groups, with key findings on gaze-following biases and saliency.	RQ 1. Can gaze-following features in eye movement data be used to classify children with ASD and TD children?
[27]	1. ASD children showed significant differences in fixation time percentages compared to TD children across most areas of interest (AOIs), except for certain stimuli like the moving helicopter. 2. Toddler and preschool-aged children with ASD exhibited distinct fixation patterns, with notable age-related interactions, such as reduced eye fixation in preschool-aged ASD compared to TD. 3. Machine learning (SVM) achieved 80% accuracy in discriminating ASD from TD toddlers and 71% accuracy for preschoolers, highlighting the potential for early ASD screening.	RQ 1. Do toddlers and preschoolers with ASD show distinct fixation patterns compared to TD children? RQ 2. Can eye-tracking data combined with machine learning accurately distinguish ASD from TD children at different developmental stages?