PREDICTING THE PRESENCE OF AUTISM SPECTRUM DISORDER BASED ON EYE-TRACKING SCAN PATH IMAGES

Sara Chandorkar
Panther Creek High School
Cary, North Carolina, U.S.A

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects communication, social interaction, and behavior. With over 5.4 million individuals in the United States diagnosed with ASD [1], early detection is critical for improving outcomes through timely intervention. Eye-tracking technology, which monitors visual attention patterns, has emerged as a potential tool for early diagnosis [2]. Our research aimed to explore whether eye-tracking scan path images could be used to predict the presence of ASD by applying machine learning techniques.

We collected eye-tracking data from individuals with ASD and non-ASD controls, processed the images to enhance clarity, and labeled them accordingly (ASD = 1, non-ASD = 0). Several machine learning models were tested, with logistic regression being the most effective, achieving an accuracy of 74.6% using a regularization parameter (C) of 0.1. This indicates that while eye-tracking can aid in ASD classification, there is still room for improvement in model performance.

Our findings suggest that eye-tracking combined with machine learning has the potential to assist in early ASD detection. This non-invasive method may also be useful in diagnosing other neurological conditions, such as Alzheimer's, offering healthcare professionals a new approach for early intervention.

INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition marked by difficulties in social interaction, communication, and repetitive behaviors [3]. The disorder is highly heterogeneous, with varying degrees of severity among individuals, and it affects people across all races, ethnicities, and backgrounds. Early and accurate diagnosis is crucial for timely interventions that can significantly enhance the quality of life for individuals with ASD. However, traditional diagnostic methods, which rely on behavioral assessments by clinicians, can be subjective and time-consuming. With the increasing prevalence of ASD, there is a growing need for more objective, efficient, and scalable diagnostic tools.

Recent advancements in technology have introduced new possibilities for diagnosing ASD, including the use of eye-tracking technology. Eye-tracking captures where and how long a person looks at different elements within a visual scene, providing insights into their visual attention patterns. These patterns can be visualized as scan path images, which map the sequence and duration of eye fixations and movements. Many studies have found that individuals with ASD often exhibit distinct visual attention patterns compared to neurotypical individuals, such as reduced focus on socially relevant stimuli, like human faces, and increased attention to objects [4][5]. This divergence in gaze patterns suggests that eye-tracking data could serve as a valuable biomarker for ASD. Therefore, the objective of this research is to investigate the feasibility of using machine learning techniques to classify eye-tracking scan path images as either ASD or non-ASD.

Machine learning, a subset of artificial intelligence, focuses on developing algorithms that can learn from data to make decisions or predictions. These algorithms are particularly adept at handling large datasets and identifying complex patterns that may be difficult for humans to discern. Our research addresses a supervised classification problem, where machine learning algorithms are trained on labeled datasets to categorize input data. In this case, the input data consists of grayscale scan path images derived from eye-tracking technology, and the output is a set of labels (ASD or non-ASD) that indicate the predicted category for each image.

We employed logistic regression as the primary classification method, evaluating the model's performance using metrics such as accuracy, precision, recall, F1 score, and the area under the ROC curve (AUC). This approach allowed us to assess the effectiveness of eye-tracking scan path imaging in distinguishing between ASD and non-ASD cases.

BACKGROUND

In exploring early autism spectrum disorder (ASD) detection through eye-tracking technology, several studies offered valuable perspectives. In a study by Daniel Fernandez-Lanvin et. al. researchers aimed to develop a screening tool for toddlers under 24 months [6]. They used eye-tracking devices to monitor how toddlers responded to videos designed to test their social engagement and gaze-following behaviors. By applying various machine learning algorithms, like Random Forests and SVM, the study found that these methods could effectively classify ASD cases. While this approach offers a promising advancement in early screening, a limitation of this study was that only toddlers were utilized. There is a need for a more diverse sample of subjects to accurately diagnose gaze behaviors across all age groups.

Another important study by Qiuhong Wei et. al, reviewed several research papers on using eye-tracking data for ASD diagnosis [7]. This meta-analysis looked at different machine learning techniques and their success rates in identifying ASD based on gaze patterns. It showed that while methods like Random Forests and SVM were effective, the results varied depending on factors such as sample size and participant age.

In a study by Warren Jones et. al, eye-tracking was evaluated as a tool for early autism diagnosis in children aged 16 to 30 months [8]. By measuring social visual engagement, the study compared eye-tracking results to clinical diagnoses made by specialists. It found that eye-tracking demonstrated a sensitivity of 71.0% and a specificity of 80.7%, with stronger results in confirmed autism cases. The study showed that eye-tracking data correlated well with clinical assessments of social disability and cognitive abilities, suggesting that eye-tracking could complement existing methods for diagnosing autism.

Overall, these studies demonstrate the potential of eye-tracking technology and machine learning in early ASD detection and highlight the ongoing need to refine these approaches for better accuracy and accessibility. In our study, we focused on utilizing a broader set of eye-tracking data to explore whether machine learning can generalize across various age groups and find the optimal model that brings out the most accurate results.

DATASET

The dataset used in this project is a collection of Eye-Tracking Scan Path (ETSP) images contributed by Mahmoud E. and accessed through figshare, designed to explore the visual attention patterns of individuals with Autism Spectrum Disorder (ASD) [9]. This dataset comprises 219 ASD and 328 neurotypical images for a total of 547 ETSP images generated from eye-tracking data collected from 59 participants. The participants were exposed to a series of autism-specific visual stimuli during the experiment, with their eye movements recorded using an SMI RED eye tracker. These stimuli were carefully chosen to highlight differences in gaze patterns between individuals with ASD and neurotypical individuals.

The ETSP images serve as the vision-based data for this study, providing a visual representation of the sequence and duration of eye fixations and saccades (rapid eye movements) as participants viewed the stimuli. The dataset includes both ASD and non-ASD participants, with each image labeled accordingly to facilitate supervised machine learning tasks.

The preprocessing of the dataset involved several key steps to prepare the eye-tracking scan path images for analysis. Initially, images were converted to grayscale to reduce complexity while retaining crucial gaze pattern details. Contrast enhancement was applied to improve feature visibility. The images were then converted into NumPy arrays, which are suitable for machine learning. Labels were assigned to distinguish between ASD and non-ASD images. Following this, the dataset was combined and split into training and testing sets using an 80-20 split, and the images were flattened into 1D arrays to match the input requirements of the machine learning models.

The features extracted from these images—such as the length and pattern of fixations—are directly relevant to our research question, as they provide measurable differences between ASD and non-ASD participants. By analyzing these features, we aim to develop a reliable method for classifying individuals based on their eye-tracking data, contributing to more objective and scalable diagnostic tools for ASD.

METHODS

To address the problem of classifying eye-tracking scan path images into Autism Spectrum Disorder (ASD) and non-ASD categories, we utilized several machine learning algorithms and a structured workflow. The process involved data preprocessing, model training, and evaluation using various classification techniques.

We began by preprocessing the eye-tracking scan path images. The images were initially converted to grayscale to simplify the data while preserving critical visual information necessary for gaze pattern analysis. Contrast enhancement was applied to improve the clarity of features within the images. Each image was then converted into a NumPy array, which is a format compatible with machine learning frameworks. Labels were assigned to indicate ASD (1) or non-ASD (0) status. The dataset, consisting of 547 images (from 59 participants), was split into training and testing sets using an 80-20 split ratio to evaluate model performance effectively. The images were flattened from their original 2D format into 1D arrays to suit the input requirements of the machine learning models.

We tested out several machine learning models from sklearn to find the model with the optimal results.

Logistic regression was employed as our primary classification technique. This model estimates the probability of binary outcomes based on input features.

A Ridge Classifier was also tested. It is similar to logistic regression but includes L2 regularization to prevent overfitting. This classifier performs well in scenarios with multicollinearity and helps to stabilize the model by penalizing large coefficients.

To test more complex decision-making, we used the Random Forest Classifier, an ensemble method that constructs multiple decision trees and aggregates their outputs. This approach improves robustness and accuracy by leveraging the collective wisdom of various trees.

A Decision Tree Classifier was also tested to model decisions based on feature values. This algorithm splits the data into branches to make predictions, providing a clear, interpretable decision-making process.

Finally, we applied a Support Vector Classifier (SVC) with a linear kernel to separate the data into classes. SVC aims to find the optimal hyperplane that maximizes the margin between different classes, making it effective for classification problems.

After preprocessing, the models were trained on the training set and evaluated on the testing set. The performance of each model was assessed using several metrics:

Accuracy: The proportion of correctly classified instances.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

Figure 1

Precision: The ratio of true positive predictions to the total predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Figure 1

Recall: The ratio of true positive predictions to the total actual positives.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Figure 3

F1 Score: The harmonic mean of precision and recall, providing a balance between the two metrics.

$$\text{F1} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}}$$

Figure 4

Confusion Matrix: Used to visualize the distribution of true positives, false positives, true negatives, and false negatives.

ROC Curve and AUC: The Receiver Operating Characteristic curve and Area Under the Curve (AUC) were used to evaluate the model's ability to distinguish between ASD and non-ASD cases across different thresholds.

Each model's performance was compared to determine the most effective approach for classifying eye-tracking scan path images, with a focus on accuracy and generalization to new data.

To enhance the performance of the logistic regression model, we performed hyperparameter tuning using both grid search and random search methods. An initial random search was conducted to explore a wide range of hyperparameters for logistic regression, including the regularization type (penalty), regularization strength (C), and the number of iterations (max_iter). This search provided insights into potentially effective hyperparameter ranges. Following the random search, a grid search was performed to systematically evaluate a predefined hyperparameter. Grid search involved testing various values for the regularization strength 'C', the optimal hyperparameter.

RESULTS & DISCUSSION

RIDGE CLASSIFIER

The Ridge Classifier model achieved an accuracy of 63%, with a precision of 0.7143, recall of 0.3774, and an F1 score of 0.4938. These metrics indicate that while the Ridge Classifier was able to reduce some false positives, it was less effective at correctly identifying individuals with ASD (lower recall), which in turn decreased its overall performance. The lower F1 score compared to Logistic Regression further confirms that Ridge Classifier struggled to balance precision and recall, limiting its practical applicability.

RANDOM FOREST CLASSIFIER

The Random Forest Classifier achieved a slightly better accuracy of 65%, with a precision of 0.7778 and a recall of 0.3962. The F1 score was 0.525, highlighting that this model performed reasonably well in minimizing false positives while maintaining a modest recall. Random Forest's ensemble nature helped it generalize better than some single-decision models like Ridge Classifier, but it still fell short of Logistic Regression in overall performance.

DECISION TREE CLASSIFIER

The Decision Tree Classifier performed the worst of all models, with an accuracy of 60%. Precision was 0.6098, recall was 0.4717, and the F1 score was 0.5319. While the model was relatively balanced between precision and recall, its lower overall accuracy and F1 score suggest that Decision Tree struggled with overfitting or was too simple to capture the complexity of the data. Despite its interpretability advantages, this model's performance was not competitive with others.

SUPPORT VECTOR CLASSIFIER (SVC)

The Support Vector Classifier achieved a performance similar to the Random Forest, with an accuracy of 65%, precision of 0.7586, recall of 0.4151, and an F1 score of 0.5366. Although SVC's precision was high, indicating effective identification of non-ASD cases, its recall was suboptimal, meaning it missed a significant proportion of ASD cases. The F1 score and accuracy were lower than Logistic Regression, further supporting the decision to focus optimization efforts on the latter.

LOGISTIC REGRESSION

The Logistic Regression model, prior to further tuning, achieved an accuracy of 70%. Precision was 0.8125, indicating that the model was fairly good at minimizing false positives, while recall was lower at 0.4906, showing that it struggled somewhat with false negatives. The F1 score of 0.6118 balanced these two metrics, suggesting that the model was moderately effective in classifying ASD cases, though there was room for improvement.

The confusion matrix (Figure 5) further illustrates the model's classification performance. The ROC curve (Figure 6) for this model yielded an AUC (Area Under the Curve) of 0.66, which suggests that while the model is better than random classification, its discriminatory power between ASD and non-ASD cases was not optimal. Improving the recall metric, in particular, was a key focus in subsequent iterations.

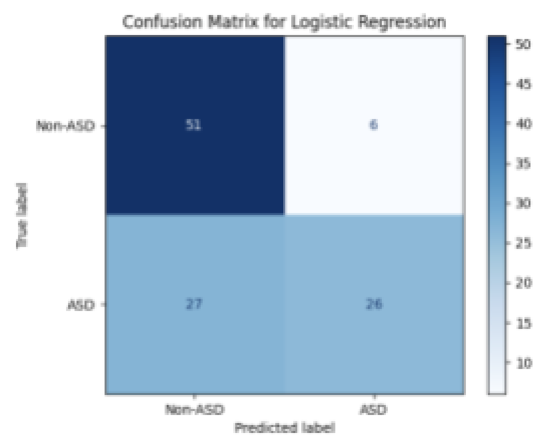

Figure 5: Confusion Matrix for initial Logistic Regression model with a True Positive (TP) of 51, False Negative (FN) of 6, False Positive (FP) of 27, and True Negative (TN) of 26.

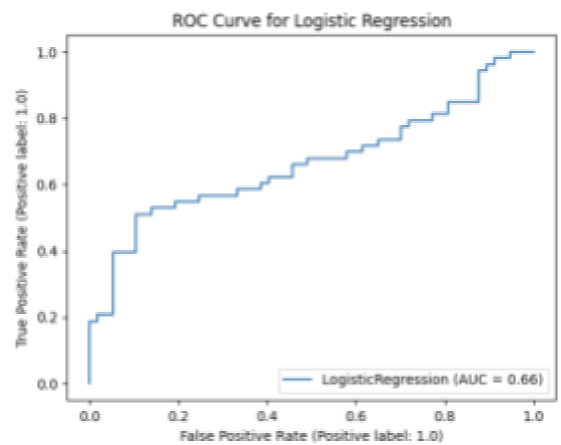

Figure 6: ROC Curve for initial Logistic Regression model. Shows an AUC score of 0.66.

Given its strong initial performance, we improved the Logistic Regression model even further by fine-tuning its hyperparameters. We used a grid search (see Figure 7) to find the best value for the regularization parameter "C," which turned out to be 0.1. After this adjustment, the model's accuracy increased to 74.6%, which is a nearly 5% improvement from before.

This boost in accuracy shows how important regularization is for reducing overfitting and making the model more reliable. After fine-tuning, Logistic Regression consistently outperformed the other models in various metrics. The models like Ridge Classifier and SVC showed high precision but lower recall, meaning they were good at identifying non-ASD cases but missed many ASD cases. On the other hand, Logistic Regression managed to find a better balance between precision and recall, resulting in a higher F1 score.

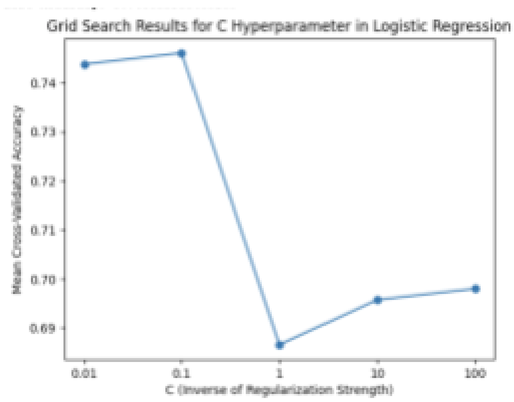

Figure 7: Grid search results for hyperparameter "C".

Despite the promising results, the Logistic Regression model still exhibited some limitations. The relatively low recall in the initial iteration suggests that the model struggled to identify all ASD cases. The presence of false positives, particularly in the pre-tuned model, could be attributed to noise in the data or insufficient feature representation. The hyperparameter tuning process significantly improved accuracy, but further improvements may be possible with more sophisticated techniques such as feature selection or alternative regularization methods.

In addition, the AUC of 0.66 (before tuning) and subsequent improvement indicate that the model may still misclassify individuals near the decision boundary. Enhancing the model's ability to handle borderline cases, perhaps by incorporating additional data or applying advanced algorithms, remains a future area for exploration.

CONCLUSION

A potential problem with this research is the limited diversity in the dataset, particularly concerning age and gender representation among individuals with Autism Spectrum Disorder (ASD). The current dataset may not fully represent all age groups and genders, which could introduce bias and affect the generalizability of the AI model. This lack of diversity may lead to inaccurate predictions or reduced performance when the model encounters underrepresented groups.

To address these issues, we could expand the dataset to include more diverse samples, with a particular focus on improving representation across various age groups and genders. By increasing the diversity of the dataset, we could enhance the model's accuracy and robustness. Additionally, incorporating data from different sources could help mitigate bias and improve the model's performance across different demographics.

Currently, our logistic regression model achieved an accuracy of 75%, which is lower than expected. This could be due to the model's linear nature, which may not adequately capture the complex, non-linear patterns in scan path images. Given that scan path data involves intricate patterns that logistic regression might not model effectively, exploring advanced techniques like Convolutional Neural Networks (CNNs) is warranted. CNNs are designed to handle image data and can learn and extract complex patterns through multiple layers, potentially improving classification accuracy compared to logistic regression.

Furthermore, additional preprocessing techniques for the scan path images to enhance classification accuracy may lead to more accurate classification. This may include data augmentation, multiscale processing, or applying advanced feature extraction methods to better represent the scan paths and improve the model's ability to classify different ASD patterns.

Overall, our logistic regression model has shown some promising results, but there is significant potential for improvement with a more diverse dataset and the incorporation of advanced machine learning techniques. By addressing these areas, we hope to develop a more accurate and generalizable model, ultimately contributing valuable insights to the field of ASD research.

Acknowledgments (ACK. CLAUSE TITLE)

We would like to thank the Inspirit AI mentorship program for providing the opportunity to conduct this research. A special thanks to Emily Broadhurst whose invaluable mentorship and feedback were instrumental to properly carrying out this research in the form of machine learning.

References

[1] Key Findings: Estimated Number of Adults Living with Autism Spectrum Disorder in the United States, 2017. (2024, May 16). Autism Spectrum Disorder (ASD). https://www.cdc.gov/autism/publications/adults-living-with-autism-spectrum-disorder.html#:~:text=An%20estimated%205%2C437%2C988%20(2.21%25),according%20to%20findings%20from%202017.

[2] Sqalli, M. T., Aslonov, B., Gafurov, M., Mukhammadiev, N., & Houssaini, Y. S. (2023). Eye tracking technology in medical practice: a perspective on its diverse applications. Frontiers in Medical Technology, 5. https://doi.org/10.3389/fmedt.2023.1253001

[3] Autism spectrum disorder (ASD) | Autism Speaks. (n.d.). Autism Speaks. https://www.autismspeaks.org/what-autism

[4] Machine learning study sheds light on gaze patterns in adults with autism. (2022, February 11). National Institute of Mental Health (NIMH). https://www.nimh.nih.gov/news/science-news/2022/machine-learning-study-sheds-light-on-gaze-patterns-in-adults-with-autism#:~:text=Many%20studies%20have%20indicated%20that,meaningful%20differences%20in%20social%20processing.

[5] Reimann, G. E., Walsh, C., Csumitta, K. D., McClure, P., Pereira, F., Martin, A., & Ramot, M. (2021). Gauging facial feature viewing preference as a stable individual trait in autism spectrum disorder. Autism Research, 14(8), 1670–1683. https://doi.org/10.1002/aur.2540

[6] Fernandez-Lanvin, D., Gonzalez-Rodriguez, M., De-Andres, J., & Camero, R. (2023). Towards an automatic early screening system for autism spectrum disorder in toddlers based on eye-tracking. Multimedia Tools and Applications, 83(18), 55319–55350. https://doi.org/10.1007/s11042-023-17694-8

[7] Wei, Q., Cao, H., Shi, Y., Xu, X., & Li, T. (2023). Machine learning based on eye-tracking data to identify Autism Spectrum Disorder: A systematic review and meta-analysis. Journal of Biomedical Informatics, 137, 104254. https://doi.org/10.1016/j.jbi.2022.104254

[8] Jones, W., Klaiman, C., Richardson, S., Aoki, C., Smith, C., Minjarez, M., Bernier, R., Pedapati, E., Bishop, S., Ence, W., Wainer, A., Moriuchi, J., Tay, S., & Klin, A. (2023). Eye-Tracking–Based measurement of social visual engagement compared with expert clinical diagnosis of autism. *JAMA*, *330*(9), 854. https://doi.org/10.1001/jama.2023.13295

[9] figshare. (2023, May 30). Visualization of Eye-Tracking scanpaths in Autism Spectrum Disorder: Image Dataset. Figshare. https://figshare.com/articles/dataset/Visualization_of_Eye-Tracking_Scanpaths_in_Autism_Spectrum_Disorder_Image_Dataset/7073087