

Summer 8-2021

Predicting Eye Movement and Fixation Patterns on Scenic Images Using Machine Learning for Children with Autism Spectrum Disorder

Raymond Anden
Chapman University, anden100@mail.chapman.edu

Follow this and additional works at: https://digitalcommons.chapman.edu/cads_dissertations

Recommended Citation

Firstinitial. Lastname, "Predicting eye movement and fixation patterns on scenic images using Machine Learning for Children with Autism Spectrum Disorder," Ph.D. dissertation, Chapman University, Orange, CA, Year. <https://doi.org/10.36837/chapman.000311>

This Dissertation is brought to you for free and open access by the Dissertations and Theses at Chapman University Digital Commons. It has been accepted for inclusion in Computational and Data Sciences (PhD) Dissertations by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

Predicting eye movement and fixation patterns on scenic images
using Machine Learning for Children with Autism Spectrum
Disorder

A Dissertation by

Raymond Anden

Chapman University

Orange, CA

Schmid College of Science and Technology

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Data Sciences

August 2021

Committee in charge:

Erik Linstead, Ph.D., Chair

Elizabeth Stevens, Ph.D.

Hesham El-Askary, Ph.D.



CHAPMAN UNIVERSITY
SCHMID COLLEGE OF SCIENCE AND TECHNOLOGY

Computational and Data Sciences

The dissertation of Raymond Anden is approved.

**Erik
Linstead**

Digitally signed by
Erik Linstead
Date: 2021.08.11
20:29:17 -07'00'

Erik Linstead, Ph.D., Chair

**Stevens,
Elizabeth**

Digitally signed by
Stevens, Elizabeth
Date: 2021.08.12
15:20:15 -06'00'

Elizabeth Stevens, Ph.D.

A handwritten signature in black ink, appearing to read 'Hesham El-Askary', is written over a horizontal line.

Hesham El-Askary, Ph.D.

August 2021

Predicting eye movement and fixation patterns on scenic images using
Machine Learning for Children with Autism

Copyright © 2021

by Raymond Anden

ACKNOWLEDGEMENTS

Dedicated to my Doggy Angel Baby, Julieta. You are forever loved and missed.

LIST OF PUBLICATIONS

R. Anden, E. Linstead. Predicting eye movement and fixation patterns on scenic images using Machine Learning for Children with Autism Spectrum Disorder. IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM2020), 1950–1956, 2020.

K. Hyde, M. N. Novack, N. Lahaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead, “Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review,” Review Journal of Autism and Developmental Disorders, vol. 6, no. 2, pp. 128–146, 2019.

ABSTRACT

Predicting eye movement and fixation patterns on scenic images using Machine Learning for

Children with Autism

by Raymond Anden

This study uses eye-tracking experiment data to predict the fixation points for children with Autism Spectrum Disorder (ASD) and Typically Developing (TD) for 14 ASD and 14 TD subjects for 300 scenic images [1]. Based on explanatory Logistic Regression models, it is evident that fixation patterns for both ASD and TD subjects focus near the center of each scenic image. Using gradient boosting the researchers successfully identify 31.7% and 39.5% of all fixation points in the top decile of predicted fixation points for ASD and TD subjects respectively. Results conclude that TD subjects have less variability in their eye movement and fixation points leading to increased accuracy in predicting where they will look.

TABLE OF CONTENTS

| | <u>Page</u> |
|--|-------------|
| ACKNOWLEDGEMENTS | IV |
| LIST OF PUBLICATIONS | V |
| ABSTRACT..... | VI |
| LIST OF TABLES | IX |
| LIST OF FIGURES | X |
| LIST OF ABBREVIATIONS | XI |
| 1 INTRODUCTION..... | 1 |
| 1.1 Background..... | 1 |
| 1.2 Machine Learning & Methods | 3 |
| 1.2.1 Logistic Regression..... | 4 |
| 1.2.2 Stepwise Variable Selection | 5 |
| 1.2.3 Variable Importance..... | 5 |
| 1.2.4 Random Forest | 5 |
| 1.2.5 Gradient Boosting | 6 |
| 1.2.6 Principal Component Analysis | 6 |
| 1.2.7 K-Means Clustering | 6 |
| 1.3 Model Measurement Criteria | 7 |
| 1.3.1 Area Under the Curve (AUC) | 8 |
| 1.3.2 Precision..... | 8 |
| 1.3.3 Recall | 8 |
| 1.3.4 F1 Score | 9 |
| 1.3.5 Root Mean Squared Error (RSME)..... | 9 |
| 2 EXPERIMENT SOURCE AND DATA ENGINEERING..... | 10 |
| 2.1 Measures and Experiment Source Data | 10 |
| 2.2 Experiment Design and Participants | 10 |
| 2.3 Scenic Image Samples and Fixation Points | 11 |
| 2.4 Model Training Data..... | 14 |
| 2.5 Data Dictionary | 16 |
| 2.6 Data Engineering Enhancements | 16 |
| 3 UNSUPERVISED LEARNING..... | 18 |
| 3.1 Exploratory Data Analysis..... | 18 |

| | | |
|----------|--|-----------|
| 3.2 | Principal Component Analysis | 22 |
| 3.3 | K-Means Cluster Analysis | 24 |
| 4 | PREDICTING FIXATION BETWEEN ASD AND TD SUBJECTS | 28 |
| 4.1 | Introduction..... | 28 |
| 4.2 | Baseline Logistic Regression Model | 28 |
| 4.3 | Enhanced Stepwise Logistic Regression | 29 |
| 4.4 | Ensemble Methods & Model Performance..... | 31 |
| 4.5 | Discussion | 33 |
| 5 | PREDICTING ASD VS TD FIXATION POINT..... | 35 |
| 5.1 | Introduction..... | 35 |
| 5.2 | Model Data Flow | 35 |
| 5.3 | Baseline Logistic Regression Model | 35 |
| 5.4 | Enhanced Stepwise Logistic Regression | 36 |
| 5.5 | Ensemble Methods..... | 38 |
| 5.6 | Discussion | 39 |
| 6 | PREDICTING DURATION OF SUBJECT FIXATION..... | 40 |
| 6.1 | Introduction..... | 40 |
| 6.2 | Model Data Flow | 40 |
| 6.3 | Baseline Linear Regression Model | 42 |
| 6.4 | Gradient Boosting | 42 |
| 6.5 | Discussion | 43 |
| 7 | APPLYING MACHINE LEARNING FIXATION MODELS ON NEW DATA | 45 |
| 7.1 | Introduction..... | 45 |
| 7.2 | Model Data Flow | 46 |
| 7.3 | Model Performance and Discussion | 46 |
| 8 | CONCLUSION | 48 |
| | REFERENCES..... | 51 |

LIST OF TABLES

| | <u>Page</u> |
|--|-------------|
| Table 2-1: Data Dictionary | 16 |
| Table 2-2: Enhanced Data Dictionary..... | 17 |
| Table 3-1: ASD Subject Independent Variable Summary Statistics | 19 |
| Table 3-2: TD Subject Independent Variable Summary Statistics | 19 |
| Table 3-3: Two Sample T-test for ASD and TD Subjects | 20 |
| Table 3-4: KS Test for ASD and TD Subjects..... | 22 |
| Table 3-5: ASD Subject Principal Components Results..... | 24 |
| Table 3-6: TD Subject Principal Component Results..... | 24 |
| Table 3-7: ASD Subject Cluster Summary | 25 |
| Table 3-8: TD Subject Cluster Summary..... | 25 |
| Table 4-1: ASD and TD Subject Logistic Regression V1 Coefficients and Importance.. | 29 |
| Table 4-2: Subject Enhanced Logistic Regression Coefficients and Importance | 30 |
| Table 4-3: ASD Subject Modeling Test Data Performance..... | 32 |
| Table 4-4: TD Subject Modeling Test Data Performance | 33 |
| Table 5-1: Subject Prediction Logistic Regression V1 Coefficients and Importance | 36 |
| Table 5-2: Subject Prediction Logistic Regression V2 Coefficients and Importance | 37 |
| Table 5-3: ASD vs TD Model Test Data Performance..... | 38 |
| Table 5-4: ASD vs TD Model Performance Improvement..... | 38 |
| Table 6-1: Duration Prediction Linear V1 Coefficients and Variable Importance..... | 42 |
| Table 6-2: Gradient Boosting Variable Importance..... | 43 |
| Table 7-1: ASD & TD Gradient Boosting Performance on New Experiment Data..... | 47 |

LIST OF FIGURES

| | <u>Page</u> |
|------------|---|
| Figure 1-1 | Logistic Regression Equation 4 |
| Figure 1-2 | F1 Score Equation..... 9 |
| Figure 1-3 | Root Mean Squared Error Equation..... 9 |
| Figure 2-1 | Scenic Image of Girl in front of a lake 11 |
| Figure 2-2 | Example Image of Animal 12 |
| Figure 2-3 | Example Scenic Image of Tree and Snow 12 |
| Figure 2-4 | Example Fixation Points based on Figure 2-1 13 |
| Figure 2-5 | Example Heat Map of Figure 2-1 14 |
| Figure 3-1 | Distribution of c.l(red) color transformation for Subject fixation points. 18 |
| Figure 3-2 | Distribution of x pixel fixation point of ASD and TD Subjects 21 |
| Figure 3-3 | Distribution of y pixel fixation point of ASD and TD Subjects 21 |
| Figure 3-4 | PCA for ASD Subjects..... 22 |
| Figure 3-5 | PCA for TD Subjects 23 |
| Figure 3-6 | Cluster Analysis for ASD Subjects..... 26 |
| Figure 3-7 | Cluster Analysis for TD Subjects 27 |
| Figure 6-1 | 99 th Percentile Cap on Scan Fixation Duration..... 41 |
| Figure 6-2 | 95 th Percentile Cap on Scan Fixation Duration..... 41 |
| Figure 7-1 | Multiple Dominant object image 46 |

LIST OF ABBREVIATIONS

| <u>Abbreviation</u> | <u>Meaning</u> |
|----------------------------|---|
| AIC | Akaike information criterion |
| ASD | Autism Spectrum Disorder |
| AUC | Area under ROC Curve |
| KS | Kolmogorov-Smirnov Statistic |
| RGB | Red, Green, Blue Primary colors |
| RMSE | Root Means Squared Error |
| ROC | Receiver operating characteristic curve |
| PCA | Principal Component Analysis |
| TD | Typically Developing |

1 Introduction

1.1 Background

Past work has studied eye tracking, fixation and scan paths to determine age, predict fixation paths, and diagnose patients with conditions such as dyslexia and Autism Spectrum Disorder (ASD). This paper focuses on leveraging machine learning to identify the similarities and differences in predicting fixation points between ASD and Typically Developing (TD) subjects and their inherent behaviors.

Aligned to the intent of this paper, Judd et al studied scan paths to identify where subjects would fixate on a given image. Their study recorded data for tracking paths and fixation location for 15 viewers both male and female between the ages of 18 and 35. They found that 40% of all fixations occurred within the center 11% of an image, and 70% of fixations occur within the center 25% of an image [2].

A study by Dalrymple et al used machine learning to classify the age of toddlers using data from patients 18 and 30 months old studying their fixations and gaze patterns. They were able to identify key differences in age based on bright, dark and length of the fixations between the 18 and 30-month-old toddlers. They had success in classifying age groups with classification accuracy of 0.83 and AUC of 0.90 [3].

Rello and Ballestos used eye movement data to predict with 80.15% accuracy subjects with and without dyslexia using a Support Vector Machine. The dataset used contained observations for Spanish speaking subjects between the ages of 11 to 54 years old, 97 subjects with normal or

corrected-to-normal vision and 48 with diagnosed with dyslexia. They were able to prove that readers with dyslexia have longer fixations which lead to longer reading times [4].

A study by Carette et al used visual representations of eye tracking scan paths to build Artificial Neural Network models that predict ASD with AUC over 0.90 on a group of 59 ASD (29) and TD (30) children with the average age of 8 years old. In this study subjects watched a set of videos to capture the ocular activity which varied in content and length. Here the primary goal early detection and diagnosis of ASD [5].

Pierce et al took a deep dive to study toddlers respective behavior characteristics by studying visual fixation patterns towards geometric images to classify children from 14 to 42 months with ASD. In this study, a one-minute video was used to capture total fixation behavior for 37 toddlers with ASD, 22 with developmental delay and 51 typically developing. They were able to find that toddlers that spent 69% or more time fixating on geometric patterns, they could accurately classify 100% of toddlers as having ASD [6].

A longitudinal study evaluating infants between 2 and 6 months concluded that a decline in eye fixation over time lead to an increase in ASD Diagnosis later in life. This finding is a very early indicator of social disability as very early eye looking appears to begin at normal levels and decline during the observation window which can be leading indicator for early intervention [7].

Prior work studying fixation patterns along with the experiment performed by Duan et. Al that becomes the focus of this research, present an opportunity to leverage one experiment data set for a group of children with Autism Spectrum Disorder (ASD) and Typically Developing (TD) children in multiple ways. Previous studies focus on a single main modeling objective, while

having very interesting findings there is an opportunity to extract more from a given experiment. The initial starting point was to predict fixation points on a given image for the two groups independently, in which the conclusion shows that TD children do have a bit more consistent and predictable behavior. This problem and conclusion in itself aim to aid in potential diagnosis in future studies for similar aged and characteristic children, again similar to prior diagnosis type research in the past.

A novel approach to understanding the differences between the fixation patterns is presented when we model fixation ownership meaning given a fixation point, is it possible to predict whether that prediction comes from an ASD or TD subject. An additional new approach to using models built on child fixations but applied to adult fixations rounds out the research. While there is decay in the model performance for very different subjects, it is still very promising that there is predictive ability of the models when given a new data set. While there were many successful learnings in the research, there was one unsuccessful modeling approach attempting to predict the duration of fixation for our subjects. While unsuccessful in the predictive accuracy sense, a baseline is created for future work.

1.2 Machine Learning & Methods

Machine learning and the algorithms leveraged in the studies of this paper can be classified as either Supervised or Unsupervised learning. Supervised learning is used when we have a known outcome that we are building models on, fixation points, duration of fixation, classification of fixation point subject, etc. Unsupervised learning methods principal component analysis and k-means are used when we are not attempting to predict a dependent attribute rather aiming to understand relationships between independent variables and between the subjects of the study.

The supervised machine learning ensemble random forest and gradient boosting are black boxes algorithms in the sense that we do not have an additive or linear equation given in generalize linear modeling techniques. The benefit of these ensemble methods is that they are able to generate several (dozens, hundreds or thousands etc.) models that outperform a single model. While these methods are built using multiple simple models, in combination they prove to significantly outperform all single model baselines studied in this paper.

1.2.1 Logistic Regression

Generalized Linear Model that can be applied when the response variable is translated into a binary variable (yes or no event). The logistic regression model is used to output a probability of the binary class variable known as the target variable. Logistic regression estimates the parameters of independent input variables that will be summed to create a linear combination summation of log-odds. After the log-odds are generated from the linear combination summation of independent variables, the exponent of the log-odds is taken to then give the predicted odds. Lastly to translate the odds to the final predicted probability, the odds are divided by 1 + odds. The ‘stats’ R package is used to fit the logistic regression models [9].

$$\log odds = \ln \left(\frac{P(yes)}{1 - P(yes)} \right) = B_0 + \dots + B_k$$

$$odds = e^{\log odds}$$

$$Prob(Event) = \frac{odds}{1 + odds}$$

Figure 1-1 Logistic Regression Equation

1.2.2 Stepwise Variable Selection

To reduce the independent variable candidate set in the generalized linear regression models, stepwise variable selection is used to generate the best fitting model based on AIC. Stepwise variable selection begins with no variables and adds the most significant variable in the candidate independent variable set. The selection process then considers adding and subtracting variables in sequence until the minimum AIC is established. The ‘stats’ R package is used for the stepwise variable selection [9].

1.2.3 Variable Importance

Variable importance in logistic regression model ranked by the value of the z statistic for each model parameter used, generated by taking the coefficient divided by its standard error. This ranking gives the relative strength of each variable in the model. The ‘caret’ R package is used to generate the variable importance of our models[10].

1.2.4 Random Forest

Ensemble method using a collection of simple decision trees, each using a random subset of all observations and a random subset of all predictors. Each individual tree casts a prediction and the final prediction is based on a voting scheme, thus creating an ensemble of independent voting trees. For our modeling purposes we set the number of trees to 500 and variables to try to be Number of Variables / 3. Because random forest is a tree-based method, it is able to discover and handle nonlinearities and work with high dimensional data better than that of our generalized linear modeling techniques. Random forests can be used on both classification and regression problems. The ‘ranger’ R package is used [11].

1.2.5 Gradient Boosting

Ensemble method that is additive simple tree models sequentially fit to reduce residual of each step where each iteration uses a random subsample of training data used. Gradient boosting like random forest, uses a combination of simple individual models combined together to create an ensemble. While random forest builds independent models in parallel, gradient boosting business additive simple tree models that are sequentially fit to reduce the residual of each step where each step uses a random subsample of training data. Each successive iteration of the model focuses on accurately predicting the cases the prior model performed poorly on. Again, this ensemble method can be used on both classification and regression problems. The ‘gbm’ R package is used [12].

1.2.6 Principal Component Analysis

Unsupervised learning method that summarizes a dataset based on the most important information for a multivariate dataset that creates new variables called principal components that are a linear combination of the originals. Principal component analysis is traditionally used for variable reduction when the number of independent features is combined into principal components. The principal components and their variable loadings can also be used to identify the key attributes and how much variance of the independent dataset they explain. Later in this paper we leverage PCA for the latter use case, trying to understand the attributes and their combinations that are capturing the variation in the data set. The ‘stats’ R package is used [9].

1.2.7 K-Means Clustering

Unsupervised learning method that groups observations into analyst pre-defined number of similar groups or clusters based on their similarity and closeness. Based on the number of

clusters, that number of centroids are initialized randomly. To find the distance between observations, the Euclidean distance is calculated between each observation and all centroids assigning each observation to its closest centroid cluster. This process is repeated by updating each centroid position based on the updated cluster around that centroid, the process is continued until the centroids for each cluster are no longer changing. To select the number of clusters, the elbow method can be used where a range of clusters can be generated and the within cluster sum of squares is output and the number of clusters is selected based on where the curve flattens out having lower within cluster sum of squares. Reducing within cluster sum of squares is there for reducing the variability within each cluster resulting in similar observations comprising each cluster. The 'stats' R package is used [9].

1.3 Model Measurement Criteria

The standard model measurement statistics leveraged in the analysis to compare performance for each of the studies are in the following section. In supervised learning classification models, it is important to measure not only the overall model performance or fit, but also the performance at a given cut point or threshold of interest. For example, the AUC (c statistic) can be used for an overall measure of model fit and performance in classification outcome models. One misleading metric in classification problems is overall accuracy of prediction depending on the distribution of your target variable. For example, if one were predicting an event with a 1.5% occurrence, if one predicts the event occurs 0% of the time, the model has a 98.5% accuracy.

In most use cases it is more important to apply supervised learning models to a portion of the full population based on a threshold. This threshold can be a probability threshold or relative ranking of probabilities, creating cut points at a given predicted probability or proportion of all

observations. In our fixation prediction models, we establish the probability threshold to be set at the top 10% or first decile of highest predicted probabilities of fixation. Based on this threshold we can use the additional statistics of precision, recall and F1 at the given cumulative proportion of the dataset.

1.3.1 Area Under the Curve (AUC)

Area under the ROC (receiver operating characteristics) curve which plots the true positive rate on the y-axis and the false positive rate on the x-axis for all predicted probabilities [8]. R package ‘pROC’ is used [13].

1.3.2 Precision

The Positive Predictive rate equal to the number of correctly predicted positive observations divided by the total number of predicted positive observations [8]. The ‘caret’ R package is used to create a confusion matrix to derive this metric [10].

1.3.3 Recall

The true positive rate or capture rate is calculated by using the number of observations identified as true positives divided by the total actual number of positive observations [14]. The ‘caret’ R package is used to create a confusion matrix to derive this metric [10].

1.3.4 F1 Score

Model performance metric derived from a balance of Precision and Recall for uneven class distributions (Figure 1-1) [14].

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Figure 1-2 F1 Score Equation

1.3.5 Root Mean Squared Error (RSME)

The Root Mean Squared Error is used to measure how spread out the residuals are between the actual and predicted values. It is calculated by: (i) squaring all residuals between the Predicted and Actual values (ii) Finding the average of all residuals (iii) Taking the square root of the result.

$$RMSE = \left[\sum_{i=1}^N (Predicted_i - Actual_i)^2 / N \right]^{1/2}$$

Figure 1-3 Root Mean Squared Error Equation

2 Experiment Source and Data Engineering

2.1 Measures and Experiment Source Data

This analysis utilizes the eye-tracking experiment performed by Duan et. Al. which contains the fixation and scan paths for children with Autism Spectrum Disorder (ASD) and Typically Developing (TD) children [1]. The images used in the experiment are comprised of natural scenic images such as unique faces, poses, animals, background settings, machinery, food, waves crashing, among others. Each of the 300 source images used in the experiment have resolutions varying between 1,280 x 960 to 768 x 1024 (width x height). The experiment resulted in data for each subject, image shown and the location of each fixation of that image. The fixation points are given by the x and y coordinates indicating a pixel on the image along with the duration of each fixation for the given subject.

2.2 Experiment Design and Participants

The eye tracker used in the experiment was a Tobii T120 with a 17-inch monitor resolution of 1280 x 1024 (width x height). Tobii Studio software was used to obtain the eye fixations of each patient. Each patient was placed approximately 65 cm from the screen. Each image was viewed at its full resolution on the screen. Each image was on screen for 3 seconds with a grey interval of 1 second between each image. The 300 source images used in the experiment were randomly presented to each patient, which 30 selected images were shown in 10 recording sessions.

The experiment comprised of 14 ASD subject and 14 TD subjects seeing each image. The age of ASD subjects ranged from 5 to 12 years old, with a mean age of 8 years old. The mean age of the TD subjects was also 8 years old for a paired control age group. During the experiment, some ASD subjects did not look at the screen while some images were presented, leaving some images with less than 14 ASD subjects data points.

2.3 Scenic Image Samples and Fixation Points

As described in the Experiment overview, the 300 scenic images varied between animals, images of children, landscapes and more. The following examples are provided to demonstrate the types of scenic images that the participants would view.

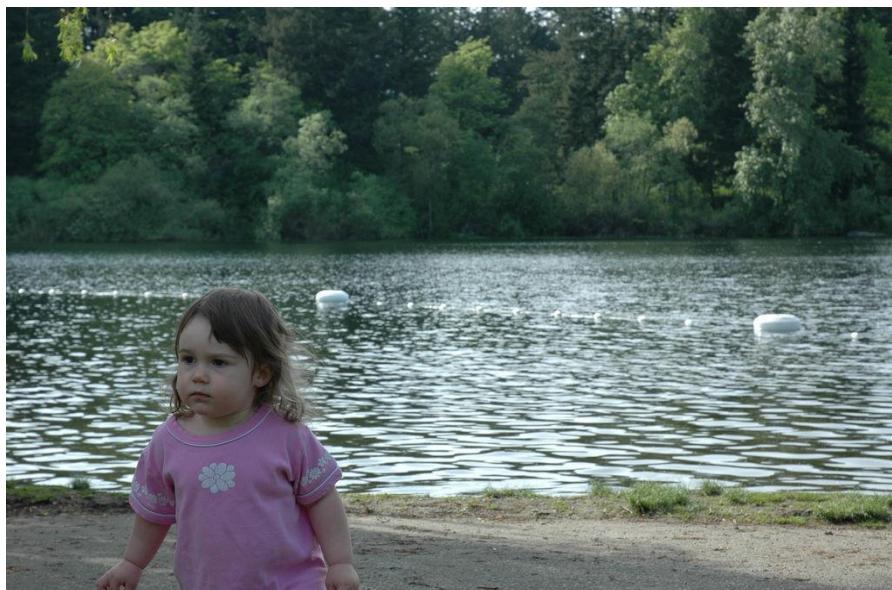


Figure 2-1 Scenic Image of Girl in front of a lake



Figure 2-2 Example Image of Animal



Figure 2-3 Example Scenic Image of Tree and Snow

When Fixation points were recorded using the Tobii T120 software, the x and y coordinates of each fixation along with the length of time in milliseconds was captured. Based on the x and y coordinates we can create a fixation point plot as seen in Figure 2-4.



Figure 2-4 Example Fixation Points based on Figure 2-1

When the element duration of each fixation is added, we can plot not only the fixation point but the length of time for each fixation creating a Heat Map as seen in figure 2-5. Here we can clearly see that the participants were focused on the young girl in the foreground of the photo. One can hypothesize that based on the age of the participants, they were drawn to the young girls, while that also appears to be the focal point of the image. There is also some focus on the buoys in the lake which are a contrast to the lake itself, along with a few fixation points in the background near the forest or trees in the background of the image itself.



Figure 2-5 Example Heat Map of Figure 2-1

2.4 Model Training Data

The output of the experimental data contains a file for each image in the experiment containing each Patient ID in the experiment and their fixation and scan paths. These fixation and scan paths are given by the x and y coordinates of the patients' fixation along with the duration of the fixation. While this provided data will be used to identify the fixation points on each of the 300 images in the experiment, we do not have any data prepared for each of the images giving meta data about each of the fixation points.

The experiment captured 29,089 fixation points for the 14 ASD subjects, compared to 34,538 for the 14 TD subjects. The subject fixation points will be used as the 'duration_flag = 1' for the binomial outcome modeling where any pixel with a duration greater than 0 will be flagged as a

pixel that a subject viewed. While the duration of each fixation was captured by the experiment, it was not used in this analysis as the focus is prediction fixation points rather than the duration of each fixation. The remaining approximately 226 million image pixels that were not viewed will be the 'duration_flag = 0' for the model target.

Due to the amount of non-fixation data points, 1/100th sample was taken on the non-fixation pixels. The random sample was taken across all images to get a fair representation of each image. This 1/100th sample generated approximately 7,500 non fixation pixels for each image, totaling 2,267,205 data points. The modeling data set is therefore under sampled on the fixation pixels compared to the non-fixation pixels.

Training and Test data sets were randomly split 70% and 30% respectively based on the 300 scenic images with 210 images being used for training the models and 90 images held out for testing the model accuracy for each algorithm. The training and test sets were consistent for both the ASD and TD subject groups.

To convert each pixel into a data vector, the R package 'imager' was used to translate each point on the pixel grid into a data frame with the red, green, blue color dimension values in 3 columns [8]. This data transformation gives us a usable data frame for each image that we can merge with the patient fixation paths to understand the characteristics of each fixation point. This data frame is also the use for our non-scanned data points where patients did not fixate, giving us a binomial indicator for each pixel on each of the 300.

2.5 Data Dictionary

The Fixation data provided generated by the experiment contained information on the Patient ID, Image ID, x and y coordinates of each fixation and the length in milliseconds while additional transformations of each pixel created more attributes for the analysis.

Table 2-1: Data Dictionary

| Variable | Description |
|-------------------|---|
| PatientID | Unique Identifier for each Patient |
| ImageID | Unique Identifier for each Scenic Image |
| x | x coordinate of each fixation point |
| y | y coordinate of each fixation point |
| Duration | Milliseconds of fixation |
| c.1(red) | Transformation of image pixel into 0-1 scale for red, green and blue 1.0 |
| c.2(green) | |
| c.3(blue) | |

2.6 Data Engineering Enhancements

With the goal of creating more explanatory attributes to compare relative model importance between ASD and TD subjects, additional Data Engineering was used to create additional attributes for the x and y coordinate of each pixel and each color attribute. To identify where on the photo a subject was fixating, each pixel was grouped into 5 categories for their x position and 5 categories for their y position. The categories for both x and y dimension attributes are 0-200, 201-400, 401-600, 601-800, and 801-Maximum. The color dimensions are split into 10 levels to further explain the color range in the pixel location that is driving the fixation. The color

dimension attributes are already scaled from 0-1, each level increments by 0.10 to the next for the 10 groups.

The data enhancements give us an independent variable list of the original 5 numeric attributes: x and y coordinates, and the RGB (c.1, c.2, c.3) color transformations along with 5 additional attributes with multiple levels for the original numeric attributes. For each of the modeling techniques the newly created attributes will generate binary dummy variables giving us a total of 45 potential attributes for the algorithms to use.

Table 2-2: Enhanced Data Dictionary

| Variable | Description |
|-------------------|--|
| PatientID | Unique Identifier for each Patent |
| ImageID | Unique Identifier for each Scenic Image |
| x | x coordinate of each fixation point |
| y | y coordinate of each fixation point |
| Duration | Milliseconds of fixation |
| c.1(red) | Transformation of image pixel into 0-1 scale for red, green and blue |
| c.2(green) | |
| c.3(blue) | |
| x_pos | Split X and Y axis into 5 ranges based on coordinates |
| y_pos | |
| c.1_dec | r/g/b variables into 10 ranges (0-.1,.11-.2, etc.) to capture more position granularity |
| c.2_dec | |
| c.3_dec | |

3 Unsupervised Learning

3.1 Exploratory Data Analysis

The starting set of independent variables for the analysis are quite simple and straightforward to begin with, focusing on the metadata available for each image and their pixels. The five variables that would be available for any image would be the location of the pixel, their x and y coordinates along with the Red, Green, Blue (RGB) color dimensions indicate by c.1, c.2 and c.3 respectively.

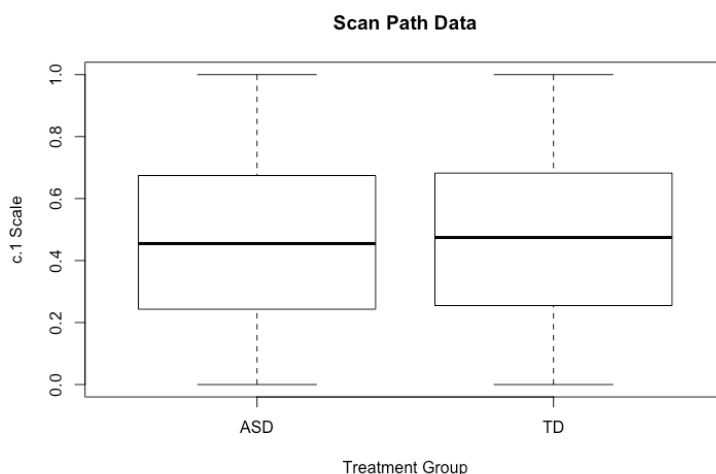


Figure 3-1 Distribution of c.1(red) color transformation for Subject fixation points

When analyzing the independent variables between the subject groups, and their boxplots the similarities stand out more than the differences right away for almost all attributes when completing a univariate analysis in Table 3-1 and 3-2. To identify any differences in the mean values of each variable between the ASD and TD group, Table 3-3 gives results of a two-sample

t-test which proves that the only variables with a statistical difference in mean value are the x and y coordinates along with the c.1 (red) color dimension.

Table 3-1: ASD Subject Independent Variable Summary Statistics

| Measure | Variable | | | | |
|--------------|----------|--------|----------|------------|-----------|
| | x | Y | c.1(red) | c.2(green) | c.3(blue) |
| Min | 1 | 1 | 0 | 0 | 0 |
| 1st Quartile | 365 | 278 | 0.243 | 0.212 | 0.169 |
| Median | 487 | 380 | 0.455 | 0.396 | 0.341 |
| Mean | 490.6 | 393.7 | 0.467 | 0.418 | 0.382 |
| 3rd Quartile | 611 | 493 | 0.675 | 0.596 | 0.5657 |
| Max | 1020 | 1024 | 1 | 1 | 1 |
| St. Dev. | 186.84 | 166.29 | 0.2693 | 0.252 | 0.260 |

Table 3-2: TD Subject Independent Variable Summary Statistics

| Measure | Variable | | | | |
|--------------|----------|--------|----------|------------|-----------|
| | x | Y | c.1(red) | c.2(green) | c.3(blue) |
| Min | 2 | 1 | 0 | 0 | 0 |
| 1st Quartile | 392 | 292 | 0.255 | 0.220 | 0.172 |
| Median | 501 | 378 | 0.475 | 0.404 | 0.341 |
| Mean | 505.6 | 390.1 | 0.478 | 0.421 | 0.382 |
| 3rd Quartile | 613 | 472 | 0.682 | 0.596 | 0.557 |
| Max | 1023 | 1019 | 1 | 1 | 1 |
| St. Dev. | 172.122 | 145.34 | 0.266 | 0.248 | 0.257 |

Focusing first on the color dimension, there is almost no discernable difference between the subject groups when comparing the RGB attributes. For example, the c.1 boxplot is shown in Fig. 3-1 where the distributions look almost identical. When looking at the actual summary statistics we also see very similar statistics for each group across color dimensions. However, the two-sample t test does indicate a statistical difference between the mean values of the c.1 (red) color dimension.

Table 3-3: Two Sample T-test for ASD and TD Subjects

| Variable | ASD Mean | TD Mean | t-Statistic | p-value |
|-------------------|-----------------|----------------|--------------------|------------------|
| x | 490.6 | 505.6 | -10.458 | < 0.01 |
| y | 393.7 | 390.1 | 2.884 | < 0.01 |
| c.1(red) | 0.4669 | 0.4779 | -5.159 | < 0.01 |
| c.2(green) | 0.4183 | 0.4213 | -1.505 | 0.1324 |
| c.3(blue) | 0.382 | 0.3819 | 0.029 | 0.9771 |
| duration | 260.6 | 318.2 | -19.824 | < 0.01 |

The x and y coordinates of scan paths have very similar minimum and maximums between each subject group however the x and y coordinate of the scan path mean, median and standard deviation values begin to show differences between the consistency of the location of the scan path. This is visually evident in the boxplots in Fig. 3-2 and Fig. 3-3 respectively where the TD group has much tighter bounds around the median and quartiles for the x and y coordinates of scan paths. The boxplot also gives better indication of the outliers presented in each of the groups, while the minimum and maximum statistics do not seem inherently different, the boxplot calculates the minimum and maximum values as 1.5 multiplied by the interquartile range (1st and 3rd Quartile). These visual representation indicate that the variability of the ASD scan paths is going to be larger than TD subjects and may impact the ability to predict in the later modeling. The two-sample t-test proves that there is indeed a statistical difference between the mean x and y coordinate between the two subject groups.

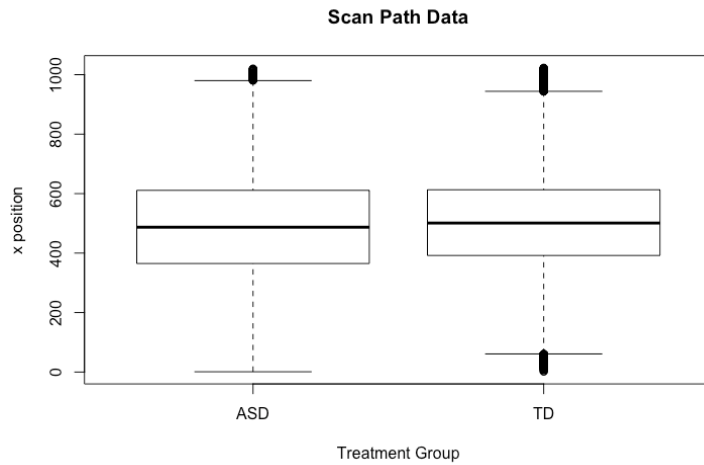


Figure 3-2 Distribution of x pixel fixation point of ASD and TD Subjects

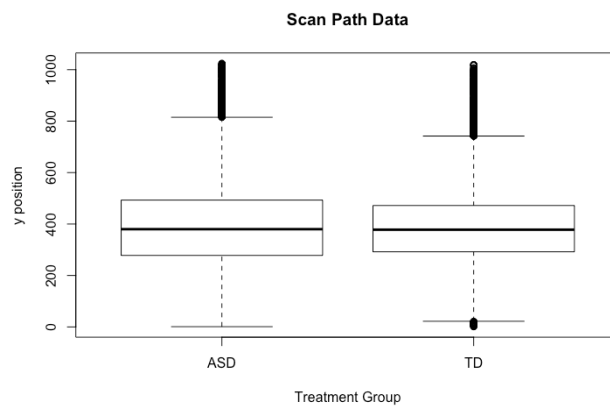


Figure 3-3 Distribution of y pixel fixation point of ASD and TD Subjects

Being that the distribution of each independent variable used in the modeling are part of a finite distribution on the x and y axis of an image or on a scale from 0 to 1, we also explored a two-sample KS test to compare the distributions of the values these variables take between TD and ASD subject fixations in Table 3-4. When running these tests, we found that all variables are significantly different in the distributions that they take for fixation points. While the averages and t-tests show slight differences between the populations, are in fact significantly different.

Table 3-4: KS Test for ASD and TD Subjects

| Variable | D Statistic | p-value |
|-------------------|---------------|-------------------|
| x | 0.0555 | < 0.001 |
| y | 0.0407 | < 0.001 |
| c.1(red) | 0.0243 | < 0.001 |
| c.2(blue) | 0.0133 | 0.007 |
| c.3(green) | 0.0113 | 0.0351 |
| duration | 0.0773 | < 0.001 |

3.2 Principal Component Analysis

Principal Component Analysis was used to better understand the similarities and differences between the two subject groups and their fixation data. Before getting into the loadings that comprise each principal component, it is staggering at the similarity between the variance explained for the 2 groups being almost identical presented in Figure 3-4 and 3-5.

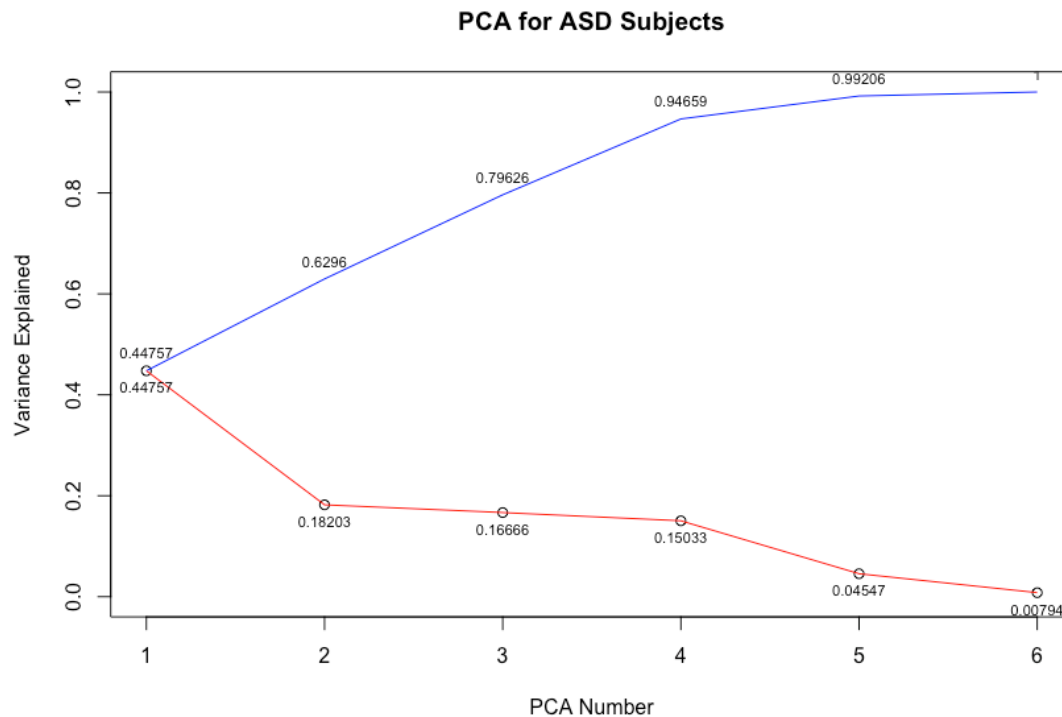


Figure 3-4 PCA for ASD Subjects

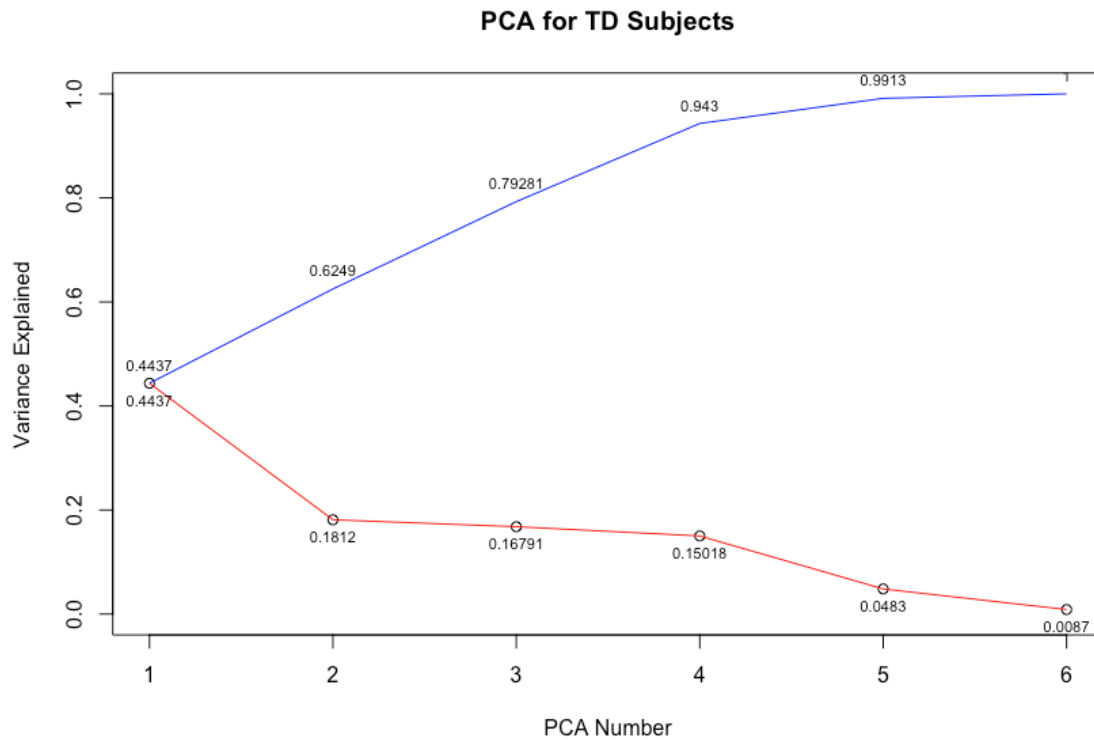


Figure 3-5 PCA for TD Subjects

When we review the loadings of each Principal component, we do see some similarities and differences between the two groups in Table 3-5 and 3-6. Each of the Principal Components contain very similar proportion of variation along with main variable loadings. For both groups Principal Component 1 is heavily driven by Color Combination, Principal Component 2 is comprised of the x and y position of each fixation and Principal Component 3 is all about the Duration of the fixation. These 3 make up almost 95% of the Variance in the dataset which now can be reduced to Color, Location and Duration of subjects' fixation. Prior to running this analysis, no hypothesis was formed or expected conclusion to be reached but it is very interesting that the behaviors and results are so similar between the groups.

Table 3-5: ASD Subject Principal Components Results

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|------------|---------------|---------------|--------------|--------|--------|--------|
| x | 0.012 | -0.689 | 0.279 | -0.669 | -0.015 | 0.000 |
| y | 0.062 | 0.695 | -0.001 | -0.715 | -0.032 | 0.014 |
| c.1(red) | -0.561 | 0.007 | -0.002 | -0.001 | -0.740 | 0.372 |
| c.2(green) | -0.600 | 0.030 | 0.003 | -0.041 | 0.054 | -0.797 |
| c.3(blue) | -0.567 | 0.023 | 0.003 | -0.048 | 0.670 | 0.476 |
| duration | -0.001 | 0.201 | 0.960 | 0.194 | 0.000 | 0.002 |

Table 3-6: TD Subject Principal Component Results

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 |
|------------|--------------|---------------|--------------|--------|--------|--------|
| x | 0.002 | -0.710 | 0.220 | -0.669 | -0.010 | -0.007 |
| y | -0.056 | 0.701 | 0.158 | -0.692 | -0.027 | -0.020 |
| c.1(red) | 0.560 | 0.010 | -0.003 | 0.004 | -0.738 | -0.377 |
| c.2(green) | 0.602 | 0.033 | 0.009 | -0.039 | 0.051 | 0.796 |
| c.3(blue) | 0.567 | 0.027 | 0.016 | -0.028 | 0.672 | -0.474 |
| duration | -0.004 | 0.047 | 0.962 | 0.267 | -0.007 | 0.004 |

3.3 K-Means Cluster Analysis

K-Means Cluster Analysis was leveraged to identify similarities and differences between the groups of subjects within our ASD and TD subject groups. The original fixation attributes were used to create the clusters: x and y pixel locations, red/green/blue color scale and fixation duration. Each subject group was placed in to 6 clusters for comparison, summarized in Table 3-

7 and Table 3-8 respectively. For each group the 6th cluster is the largest comprised of over 28% of each subject group. The ASD cluster #5 and TD Cluster #2 are both the smallest cluster for each subject group, however they have the longest duration for each subject cluster. Being that each color attribute is a 0-1 scale, we can create a color for each combination in which we see that they are all unique colors, although the color scales look similar across groups.

Table 3-7: ASD Subject Cluster Summary

| ASD Cluster | n | % of n | x | y | c.1(red) | c.2(green) | c.3(blue) | color | duration (ms) | dur(rank) |
|-------------|-------|--------|-------|-------|----------|------------|-----------|---------|---------------|-----------|
| 1 | 4,993 | 17.13% | 259.5 | 333.5 | 0.48 | 0.42 | 0.38 | #796C61 | 208.37 | 3 |
| 2 | 5,084 | 17.44% | 415.0 | 626.7 | 0.44 | 0.41 | 0.37 | #70685F | 161.54 | 5 |
| 3 | 4,996 | 17.14% | 758.7 | 374.7 | 0.45 | 0.40 | 0.37 | #72675E | 188.30 | 4 |
| 4 | 4,994 | 17.13% | 497.5 | 385.7 | 0.48 | 0.42 | 0.39 | #796C62 | 522.01 | 2 |
| 5 | 894 | 3.07% | 507.8 | 413.2 | 0.45 | 0.41 | 0.38 | #736961 | 1167.38 | 1 |
| 6 | 8,188 | 28.09% | 507.2 | 299.6 | 0.49 | 0.43 | 0.39 | #7C6D65 | 114.51 | 6 |

Table 3-8: TD Subject Cluster Summary

| TD Cluster | n | % of n | x | y | c.1(red) | c.2(green) | c.3(blue) | color | duration (ms) | dur(rank) |
|------------|-------|--------|-------|-------|----------|------------|-----------|---------|---------------|-----------|
| 1 | 4,229 | 12.24% | 546.4 | 407.6 | 0.48 | 0.42 | 0.39 | #7B6C62 | 658.23 | 2 |
| 2 | 1,336 | 3.87% | 508.0 | 394.4 | 0.46 | 0.41 | 0.39 | #756962 | 1280.55 | 1 |
| 3 | 7,399 | 21.42% | 721.5 | 373.9 | 0.47 | 0.41 | 0.37 | #77695F | 221.97 | 4 |
| 4 | 5,310 | 15.37% | 385.9 | 595.0 | 0.45 | 0.41 | 0.37 | #74695D | 192.07 | 5 |
| 5 | 6,405 | 18.54% | 384.2 | 324.5 | 0.49 | 0.43 | 0.39 | #7D6D63 | 354.99 | 3 |
| 6 | 9,859 | 28.55% | 469.2 | 326.4 | 0.49 | 0.43 | 0.39 | #7D6D64 | 105.51 | 6 |

Visually we can see one key difference between each group's clusters. Our Clusters are made up of x and y pixel coordinates along with color and duration attributes. To visualize this, we have plotted each of the color scale combinations with the average x and y coordinate while the size of each point is the average duration of fixation for the cluster group.

In our ASD clusters, we begin to see a bit more variability as only 3 of the clusters compared to 5 of the TD clusters fall between the x (300,600) and y (200,600) coordinates (Callout A in Figure 3-6 and 3-7). We also see that the largest cluster for each of the subject groups is the lowest point on the Y axis (Callout B). The smallest cluster for both groups has the longest duration and is the most center (Callout C).

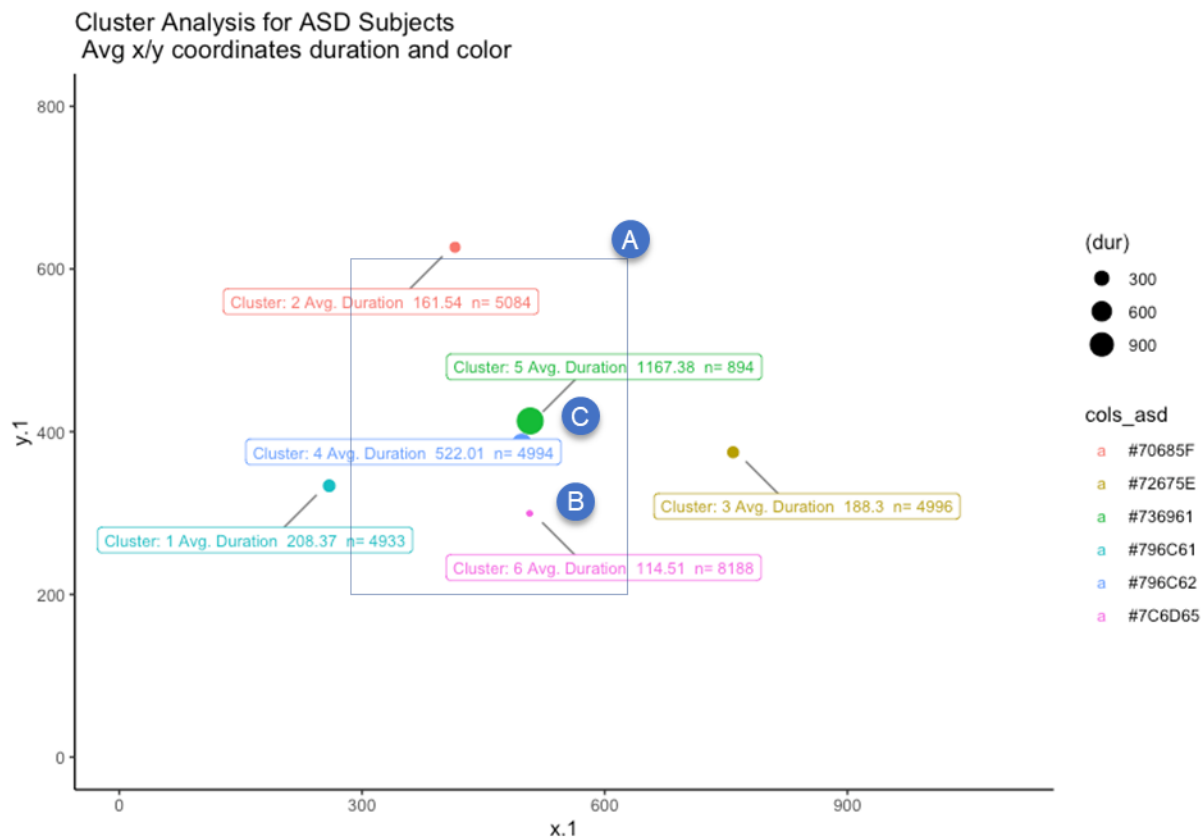


Figure 3-6 Cluster Analysis for ASD Subjects

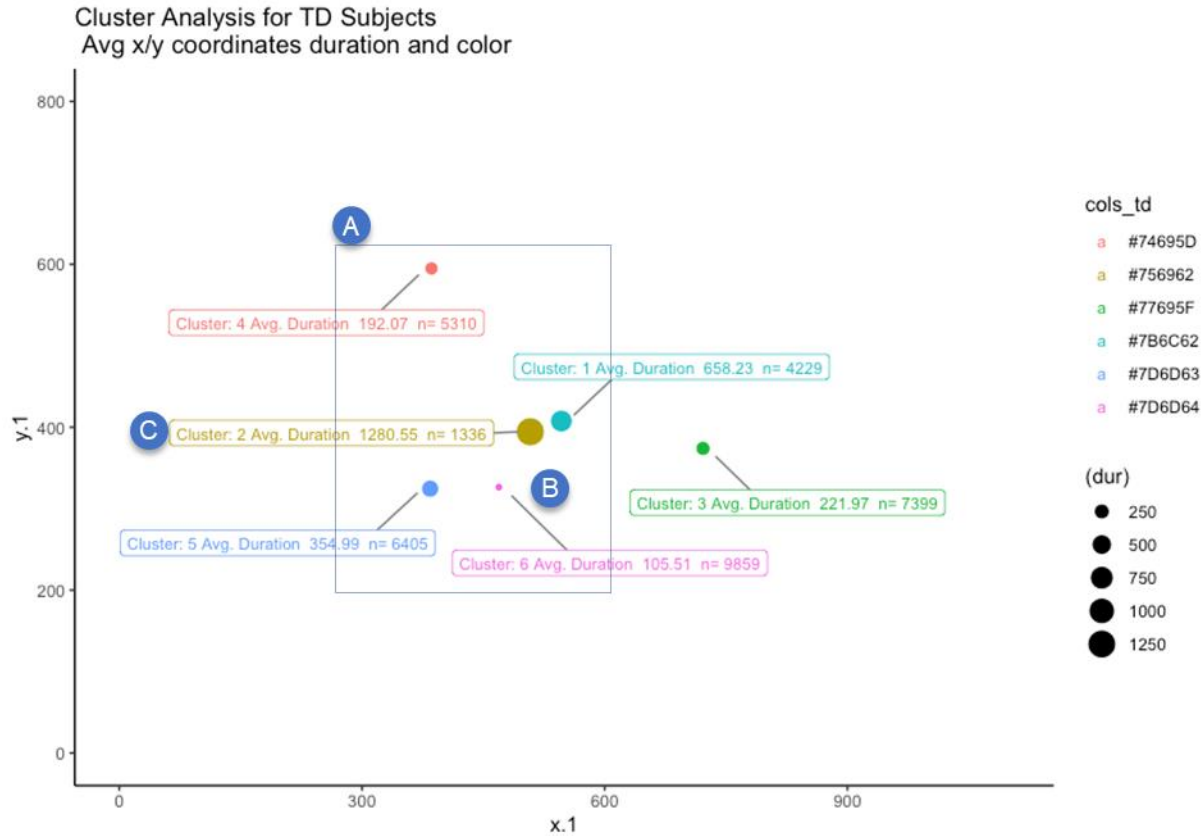


Figure 3-7 Cluster Analysis for TD Subjects

While there are similarities observed between the subject groups, it leads to the hypothesis that the TD subjects may be more consistent in their fixation patterns having less variability in the x and y coordinates of each fixation.

4 Predicting Fixation between ASD and TD Subjects

4.1 Introduction

In this section we provide the results for multiple machine learning algorithms built to predict the fixation points for both ASD and TD subject. Models are built for each subject group independently to compare the key drivers in each logistic regression model and determine if ASD and TD subject's fixation points can be predicted to the same level of accuracy. First, we begin with a baseline logistic regression model from each subject group using the initial 5 variables in the dataset: x and y coordinates of each pixel along with their RGB color transformation. Data engineering is used to create interpretable attributes for deeper insight in another round of an enhanced logistic regression. Lastly, additional machine learning techniques are leveraged to compare performance and incremental lift.

4.2 Baseline Logistic Regression Model

Initially logistic regression is used to build separate models for the ASD and TD Subject fixation points respectively with the goal of identifying the similarities and differences in the model drivers for each population's fixation points. These models will be used as the baseline to compare the predictive accuracy of additional modeling techniques in addition to the value of additional data transformations.

In the initial logistic regression equations, we see that the coefficients of the two models all have the same positive or negative direction implying the same effect on predicting fixation points.

We also see that the most important variable, based on the z value of each coefficient, is the c.1(red) color transformation, followed by the c.2(green) color transformation and the y axis position of the pixel. The only differences in variable importance between the models is the x axis position is the lowest variable of importance in the ASD subject model and it is the fourth most important variable in the TD subject model. The similarities in the variable importance between the models again leads us to conclude that fixation points appear to be consistent between the ASD and TD subjects.

Table 4-1: ASD and TD Subject Logistic Regression V1 Coefficients and Importance

| Variable | ASD Subjects | | | TD Subjects | | |
|-------------------|---------------------|-----------------|-------------|--------------------|----------------|-------------|
| | Coeff. | Var Imp. | Rank | Coeff. | Var Imp | Rank |
| x | 2.37E-05 | 0.955 | 5 | 0.00020 | 8.85 | 4 |
| y | -0.00022 | 7.256 | 3 | -0.00026 | 9.374 | 3 |
| c.1(red) | 1.842 | 35.173 | 1 | 2.303 | 50.04 | 1 |
| c.2(green) | -2.082 | 22.191 | 2 | -2.333 | 28.064 | 2 |
| c.3(blue) | 0.1944 | 2.912 | 4 | 0.094 | 1.547 | 5 |

4.3 Enhanced Stepwise Logistic Regression

To select the best model using the additional attributes, stepwise variable selection in the enhanced logistic regression model. Looking first at the ASD Subject model in Table VI, we see that the c.1(red) color dimension is no longer the most important variable rather the largest drivers are associated to pixel position in the image. In the original model, the x position had the lowest variable importance but with the additional transformations we see that the pixel position being on the x axis between 401 and 600 is the most important factor in the model. The second highest driver is the pixel position being between the 201 and 400 x axis position, and the third highest driver is the pixel being between the 601 and 800 x axis position. This indicates that

horizontal location plays a major role in the fixation patters for ASD subjects. These top 3 variables all indicate the scan paths being primarily in the center of the photo per the x axis association. The next 2 variable drivers are all associated to the y position of the pixel. In order of importance the y position between 201-400 and 401-600 which again focus on the center of the image. The coefficients for all these top 5 variables have a positive coefficient meaning that there is a higher likelihood that an ASD patient will look at the pixel towards the center of a given image. These main drivers support the earlier work by Judd et al who found that 70% of fixation occur within the center of an image [1].

The TD subject model has the same top 5 variables based on variable importance of the model; however, the order is different with the y position variables being the top 2 variables, followed by the x position of the pixel. This again supports those subjects will fixate on the center of images based on the x and y coordinates. We also see that the signs of each coefficient are negative except for x position of 401-600 which is opposite from the ASD subject model.

As observed in the baseline model, the most important color dimension in the enhanced model is the c.1(red) dimension is the top color variable in each of the subject models, while the c.2(blue) and c.3(green) color dimensions fall lower on the importance list.

Table 4-2: Subject Enhanced Logistic Regression Coefficients and Importance

| Variable | ASD Subjects | | | TD Subjects | | |
|----------------------|---------------------|-----------------|-------------|--------------------|----------------|-------------|
| | Coeff. | Var Imp. | Rank | Coeff. | Var Imp | Rank |
| x_pos_401_600 | 1.896 | 61.940 | 1 | 2.044 | 40.482 | 4 |
| y_pos_201_400 | 1.179 | 34.965 | 4 | -0.517 | 72.038 | 1 |
| x_pos_201_400 | 1.441 | 45.435 | 2 | -0.017 | 41.251 | 3 |
| y_pos_401_600 | 0.760 | 13.897 | 5 | -0.208 | 52.873 | 2 |
| x_pos_601_800 | 1.247 | 38.299 | 3 | -0.221 | 19.657 | 5 |

| | | | | | | |
|-----------------------|--------|-------|----|--------|-------|----|
| c.1 | 1.452 | 5.779 | 7 | -0.219 | 8.825 | 6 |
| y_pos_801_1024 | -0.759 | 6.605 | 6 | -0.245 | 7.610 | 7 |
| c.3_d6 | 0.245 | 5.368 | 8 | -0.348 | 4.789 | 8 |
| c.2 | -1.334 | 5.197 | 9 | -0.423 | 2.592 | 22 |
| X | | | NA | -0.579 | 4.045 | 9 |
| c.3_d5 | 0.168 | 3.971 | 10 | -0.611 | 3.736 | 11 |
| y_pos_601_800 | -0.267 | 3.413 | 13 | -0.540 | 3.913 | 10 |

4.4 Ensemble Methods & Model Performance

The 30% test set of 90 images is used and their respective data to compare the performance of the ASD and TD Subject models. To compare the model performance, we set a threshold of the Top Decile for predicted probability of a pixel being a fixation point, where a fixation point was defined in the training data set as a duration greater than 0 and a target variable ‘duration_flag=1’. Due to the imbalance of the dataset, Accuracy is of little value in comparing model performance as we would be over 98% accurate if we predicted every pixel to be 0.

The main model performance metrics we will be reviewing are Precision, Recall and F1 Score. Precision is a valuable metric because it indicates the true positive rate of the top decile, meaning what is the actual fixation rate of all data points in the decile of predicted probabilities. Recall is valuable in evaluating model performance because it indicates how many of the actual fixation points, we are able to capture in the top decile threshold cutoff. The F1 score uses both precision and recall values to evaluate the overall performance of a model into a single metric. There is low cost to false positive categorization in the model prediction, therefore recall is going to be the primary metric we use to evaluate model performance, again what is the fixation capture rate in the top decile for each of the models.

First evaluating the Performance of the ASD Subjects baseline logistic regression model compared to the enhanced logistic regression models, there is over a 71.5% increase in Precision, Recall and F1 score. For ASD subjects, Recall increases from 18.5% to 31.7%. The performance increases even more between the Baseline and enhanced logistic regression model for the TD subjects, with 84.2% increases across precision, recall and F1 score. Recall increases from 19.8% to 39.6% of all fixation points being identified in the top decile.

With the value of the enhanced data set clear, additional machine learning models were run to improve the predictive ability of fixation points. Random forest and gradient boosting models were run using the enhanced variable data set to improve each model set. For both ASD and TD subjects, the random forest models performed slightly lower than the enhanced logistic regressions across all performance metrics. However, the gradient boosting models outperformed the enhanced logistic regression models. For the ASD subjects, the best model performance is seen with gradient boosting, where recall is 33.1%. The same is seen for the TD subjects, where gradient boosting increases Recall to 39.6%. These top decile capture rates prove there is large predictive lift in predicting fixation paths for both ASD and TD subjects.

Table 4-3: ASD Subject Modeling Test Data Performance

| Algorithm | Performance Metric | | | | |
|-------------------------------------|--------------------|--------------|--------------|--------------|--------------|
| | AUC | Accuracy | Precision | Recall | F1 Score |
| Baseline Logistic Regression | 0.585 | 0.892 | 0.023 | 0.185 | 0.041 |
| Enhanced Logistic Regression | 0.743 | 0.895 | 0.039 | 0.317 | 0.070 |
| Random Forest | 0.720 | 0.895 | 0.035 | 0.282 | 0.062 |
| Gradient Boosting | 0.751 | 0.898 | 0.042 | 0.331 | 0.074 |

Table 4-4: TD Subject Modeling Test Data Performance

| Algorithm | Performance Metric | | | | |
|-------------------------------------|--------------------|--------------|--------------|--------------|--------------|
| | AUC | Accuracy | Precision | Recall | F1 Score |
| Baseline Logistic Regression | 0.594 | 0.891 | 0.029 | 0.198 | 0.051 |
| Enhanced Logistic Regression | 0.798 | 0.896 | 0.054 | 0.365 | 0.094 |
| Random Forest | 0.769 | 0.053 | 0.359 | 0.359 | 0.092 |
| Gradient Boosting | 0.788 | 0.897 | 0.059 | 0.396 | 0.102 |

4.5 Discussion

By comparison it is possible to predict the fixation points of TD subjects more accurately than ASD subjects when comparing the performance metrics it's observe higher performance metrics of the TD models with a 37% higher F1 Score, 19.7% higher, recall and 40% higher. In both subject populations, adding the new data transformations along with the machine learning ensemble methods improved model performance compared to logistic regression models only using the x and y coordinates and the RGB color transformation of a give images pixels. This type of analysis can be used to study future patients eye movement and fixation patterns and compare to their expected scan patterns for an ASD or TD subject.

This research can be paired with prior research in predicting the ASD diagnosis of patients, by focusing on how similar a subject's scan path is to ASD or TD behaviors. This can be done with both fixation point predictions along with coupling prior research that sims to understand patient perception applications identifying where patients look and what they are interested in. Prior

work mainly focused on the diagnosis of patients, however leveraging fixation patterns, durations and visual shapes of focus in developing children can aid in early detection of ASD.

A future research opportunity is to predict the duration of each fixation for each data point given the fixation has occurred on a given pixel. Based on the two-sample t-test and KS test there is a significant difference between the duration of ASD and TD scan durations. Understanding the drivers of scan duration would deepen the insight gained from this experiment.

5 Predicting ASD vs TD Fixation Point

5.1 Introduction

In this Chapter, the analysis conducted leverages all fixation point data for both the ASD and TD Subjects. Rather than predicting the fixation points for each subject, models are built to predict whether the fixation came from an ASD or TD Subject. The modeling approach is the same as Chapter 4 where a baseline logistic regression is built, then an enhanced logistic regression with additional data enhancements followed lastly by our machine learning methods.

5.2 Model Data Flow

The key data in building this model is the Patient Fixation points, where we have 29,089 and 34,538 for the ASD and TD subjects respectfully. We also have all the Image Metadata transformed into Pixel information for each of the fixation points. The full fixation data set is then randomly split into a 70% training and 30% test data set maintaining the ratio of ASD vs TD subject fixation points. The modeling objective is now to predict the likelihood the fixation is an ASD subject or not.

5.3 Baseline Logistic Regression Model

When predicting ASD vs TD subjects we are now able to leverage duration in our model as we are no longer only trying to predict a fixation but leverage the information about the given fixation to predict the subject group. Table 5-1 displays the attributes along with their importance to the baseline logistic regression model. We see that duration is indeed the top attribute in the model with a slight negative coefficient. This indicates that the longer a fixation duration, the

less likely the duration is from an ASD Subject. This supports one of the exploratory findings where we saw TD subjects have longer fixation on average. The performance of this model using the test data set and a top 30% probability threshold cutoff gives a performance recall of 0.3360, cumulative lift of 1.12.

Table 5-1: Subject Prediction Logistic Regression V1 Coefficients and Importance

| Variable | Coeff. | Odds Ratio | Var. Imp | Rank |
|--------------------------------|------------------------|-------------------|-----------------|-------------|
| duration (milliseconds) | -0.001 | 0.9993 | 19.7 | 1 |
| x | -0.0004 | 0.9996 | 8.3 | 2 |
| c.1 (red) | -0.528 | 0.5895 | 6.7 | 3 |
| c.2 (green) | 0.482 | 1.6199 | 3.5 | 4 |
| y | 0.0001 | 1.0001 | 1.2 | 5 |
| c.3 (blue) | -0.046 | 0.9549 | 0.5 | 6 |
| Test Set Performance | Recall = 0.3360 | | | |

5.4 Enhanced Stepwise Logistic Regression

Once again to further improve the performance of our model, stepwise variable selection was used to build a model based on our enhanced data set attributes. In this model we again see that duration of a given fixation is the key attribute in the model. The negative coefficient indicates that longer durations represent a lower likelihood of the fixation subject being an ASD Subject. For example, a 250 millisecond longer fixation results in 23% lower likelihood of the fixation being an ASD Subject. Next, we see that y axis position of each fixation pixel is more predictive than the x axis position. Recall in our cluster analysis we did see that all ASD clusters were

higher on the y axis which holds true in the model. Model performance increases 5.9% with the Recall Test data set performance increasing from 0.33 to 0.36

Table 5-2: Subject Prediction Logistic Regression V2 Coefficients and Importance

| Variable | Coeff. | Odds Ratio | Var. Imp | Rank |
|-----------------------------|------------------------|-------------------|-----------------|-------------|
| duration(ms) | -0.001 | 0.9993 | 19.3 | 1 |
| y_posy_201_400 | -0.447 | 0.6398 | 12.9 | 2 |
| y_posy_401_600 | -0.390 | 0.6772 | 11.0 | 3 |
| c.1(red) | -0.425 | 0.6535 | 5.7 | 4 |
| c.2(green) | 0.353 | 1.4232 | 4.4 | 5 |
| x_posx_401_600 | -0.331 | 0.7184 | 4.2 | 6 |
| y_posy_601_800 | -0.188 | 0.8286 | 4.1 | 7 |
| x_posx_201_400 | -0.214 | 0.8073 | 3.7 | 8 |
| y_posy_801_1024 | 0.243 | 1.275 | 2.7 | 9 |
| x | 0.000 | 0.9996 | 2.6 | 10 |
| x_posx_601_800 | -0.205 | 0.8149 | 1.9 | 11 |
| x_posx_801_1024 | 0.016 | 1.0158 | 0.1 | 12 |
| Test Set Performance | Recall = 0.3559 | | | |

5.5 Ensemble Methods

Further leveraging machine learning ensemble methods again prove to result in the best performance across our performance metrics given in Table 5-3. Random Forest results in a 9.4% increase in recall while gradient boosting results in a 10.6% increase in recall. When evaluating the improvement of the models in Table 5-4, we can conclude that 5.9% of the improvement is contributed to adding the enhanced data to our model evident in the enhanced logistic regression model. The gradient boosting algorithm then attributes for the addition 4.7% of improvement to the recall statistic in our model.

Table 5-3: ASD vs TD Model Test Data Performance

| Algorithm | Performance Metric | | |
|------------------------------|--------------------|--------|----------|
| | Precision | Recall | F1 Score |
| Baseline Logistic Regression | 0.5195 | 0.3360 | 0.4081 |
| Enhanced Logistic Regression | 0.5502 | 0.3559 | 0.4322 |
| Random Forest | 0.5684 | 0.3676 | 0.4468 |
| Gradient Boosting | 0.5743 | 0.3715 | 0.4511 |

Table 5-4: ASD vs TD Model Performance Improvement

| Algorithm | Performance Metric | | |
|------------------------------|--------------------|--------|----------|
| | Precision | Recall | F1 Score |
| Baseline Logistic Regression | | | |
| Enhanced Logistic Regression | 5.9% | 5.9% | 5.9% |
| Random Forest | 9.4% | 9.4% | 9.5% |
| Gradient Boosting | 10.5% | 10.6% | 10.5% |

5.6 Discussion

Similar to predicting fixation points for ASD and TD subject, we see that our models are more predictive than random where even our baseline logistic regression model has a slight lift over random. Again, we have seen that gradient boosting produces the highest recall of 0.3715 and lift of 1.24 at the 3rd decile. Data transformations along with machine learning methods improve our ability to predict the fixation behaviors. Gradient Boosting increased recall by 10.6%, 5.9% attributed to new variables and 4.7% to the method. Being that data is limited in this study, one hypothesizes that additional features about each subject will likely lead to increased model predictiveness.

6 Predicting Duration of Subject Fixation

6.1 Introduction

Continuing to leverage the dataset and experiment performed by Duan et. Al., in this chapter the analysis conducted attempts to predict the duration of each fixation for the ASD and TD Subjects independently. This allows for a second comparison into the ability for machine learning to be leveraged in predicting subject behaviors.

6.2 Model Data Flow

The target variable in the analysis is duration in milliseconds for each subject fixation. Due to the variability and outliers of fixation duration, especially in the ASD subject group a decision was made to create a cap of each fixation length at the 95th percentile for each respective group. The boxplot in Figure 6-1 presents the distribution with a 99th percentile cap in comparison to the 95th percentile cap boxplot in Figure 6-2. Based on the exploratory analysis and insight from the boxplots, the target variable duration of fixation was capped at 800 milliseconds.

The modeling population utilizes the ASD and TD Fixation points randomly split into a 70% training and 30% testing data set (210 images in training, 70 images in testing). In total we have 29,089 and 34,538 fixation points for the ASD and TD respective models. The independent variables used in our models are the image and pixel metadata for each of the fixation points.

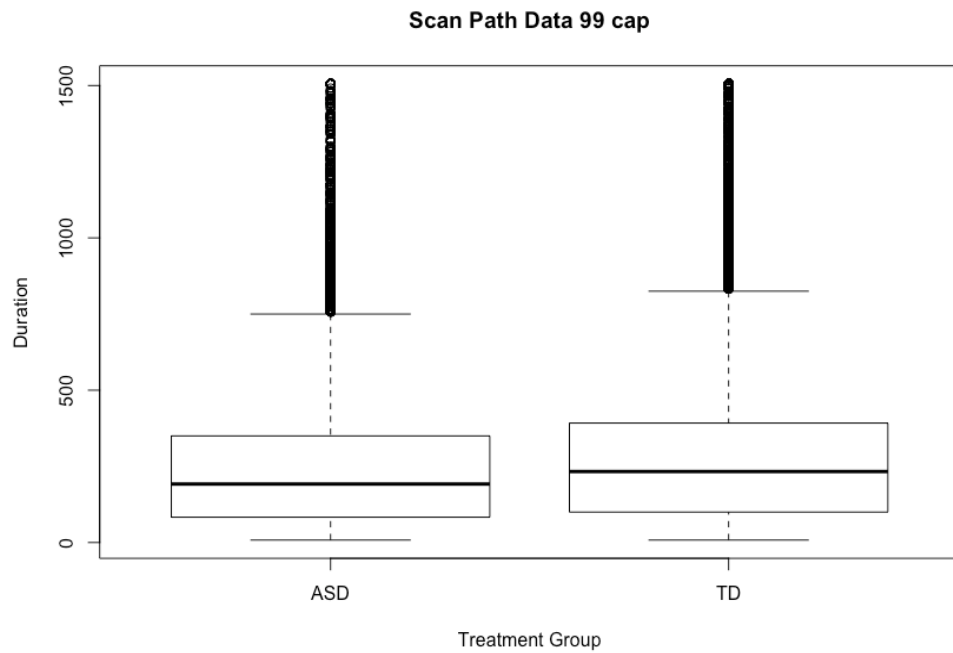


Figure 6-1 99th Percentile Cap on Scan Fixation Duration

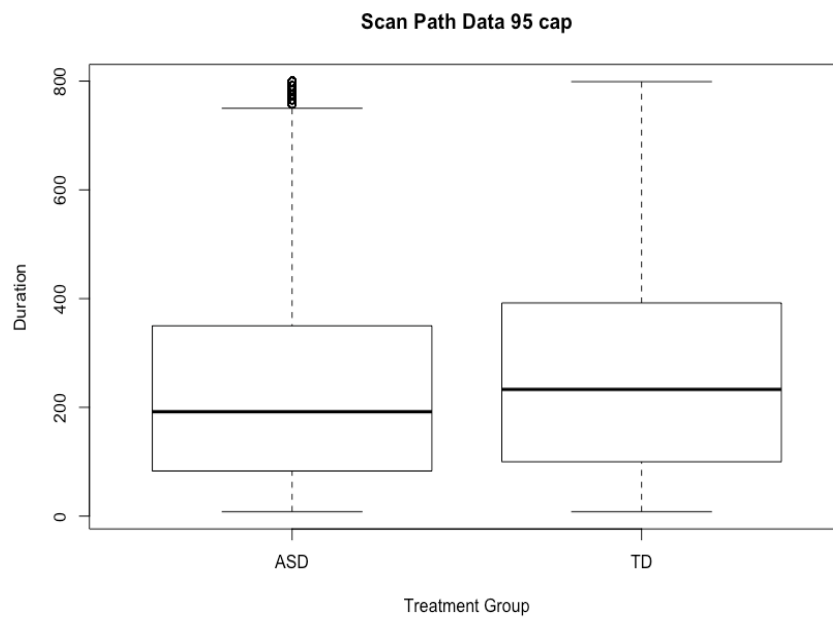


Figure 6-2 95th Percentile Cap on Scan Fixation Duration

6.3 Baseline Linear Regression Model

Due to modeling a continuous dependent outcome in this model iteration, a linear regression is first used to establish our baseline model summarized in Table 6-1. In terms of variable importance, the y axis position of the fixation point proves once again to be the most important variable in predicting subject behaviors. In comparison, the model predicting ASD subject fixation duration slightly outperforms the TD subject model with a lower RMSE of 206.97 compared to 219.40. The correlation between the model predicted duration and actual duration is also given, both are very low at 4.55% and 1.17% for the ASD and TD subjects respectfully.

Table 6-1: Duration Prediction Linear V1 Coefficients and Variable Importance

| | ASD Subjects | | | TD Subjects | | |
|----------------------|---------------------------|----------|------|---------------------------|---------|------|
| Variable | Coeff. | Var Imp. | Rank | Coeff. | Var Imp | Rank |
| x | -0.003 | 0.335 | 4 | 0.029 | 3.492 | 2 |
| y | 0.020 | 2.324 | 1 | 0.047 | 3.567 | 1 |
| c.1(red) | -0.687 | 0.088 | 5 | 14.532 | 3.267 | 3 |
| c.2(green) | 16.867 | 0.811 | 3 | -60.620 | 3.043 | 4 |
| c.3(blue) | -13.638 | 0.973 | 2 | 44.148 | 1.273 | 5 |
| Test Set Performance | RMSE = 206.97, $r=4.55\%$ | | | RMSE = 219.40, $r=1.17\%$ | | |

6.4 Gradient Boosting

While the baseline model performance did not appear to be a great beginning to the predictive nature of this analysis, gradient boosting is again leveraged to explore model improvement and variable importance for each respective subject group. The variable importance for this model iteration along with the test set performance are given in Table 6-2.

Table 6-2: Gradient Boosting Variable Importance

| | ASD Subjects | | TD Subjects | |
|-----------------------------|-----------------------------------|------|----------------------------------|------|
| Variable | Var Imp. | Rank | Var Imp | Rank |
| y | 34.800 | 1 | 26.032 | 1 |
| x | 14.355 | 2 | 23.240 | 2 |
| c.3_dec | 7.256 | 6 | 13.284 | 3 |
| c.1_dec | 8.047 | 5 | 12.371 | 4 |
| c.2_dec | 10.912 | 4 | 3.151 | 8 |
| y_pos (ranges) | 12.649 | 3 | 0.000 | 10 |
| c.1 | 3.116 | 8 | 7.915 | 5 |
| c.2 | 4.536 | 7 | 7.253 | 6 |
| c.3 | 1.874 | 10 | 4.217 | 7 |
| x_pos (ranges) | 2.453 | 9 | 2.537 | 9 |
| Test Set Performance | RMSE = 206.67, <i>r</i> =6.53% | | RMSE =219.40, <i>r</i> =2.16% | |

For each of the Subject groups, location is the key driver as the y and x axis location of each fixation point is the most important attribute. This is consistent with prior model iterations and attempts. The ability for gradient boosting to capture nonlinear relationships give us a new and different insight from prior model iterations. In this model, the raw location x and y axis position is more important than the grouped location ranges that were predictive in the logistic regression models presented in earlier chapters. When comparing the subject groups, we see a significant mix in the difference of variable importance beyond the raw x and y axis fixation location. While the RMSE and correlation on the test data sets have increased in this iteration, it's hard to claim that either of the models are doing well in predicting duration with such high RMSE and very low correlation.

6.5 Discussion

By comparison it is possible to predict the fixation points of ASD subjects slightly more accurately than TD subjects when comparing RMSE and correlations. While this iteration of

modeling lacks accuracy, it creates a baseline for future models with additional data points about the subjects. Creating segments prior to modeling duration may improve accuracy. Cluster analysis shows that duration differs based on the location of fixation point.

7 Applying Machine Learning Fixation Models on New Data

7.1 Introduction

In this chapter we apply our fixation prediction gradient boosting models for ASD and TD subjects onto a new population. The experiment and research study are like that of Duan and team, conducted by Shuo et. Al [15], where the subjects were twenty high functioning people with ASD along with nineteen neurologically and psychiatrically healthy subject with no family history of ASD. In this study, the major difference is that the ASD group had a mean age of 30.8 years compared to mean age of 8 years old in our prior research data. The dataset used in the experiment was comprised of 700 images that contained multiple dominant objects in each scene. The models were not built on specified multiple dominant objects per image nor on fully functioning adult subjects with ASD which will factor into the application of the models themselves.

A sample image in Figure 7-1 demonstrates the multiple dominant objects that appear in the images. In this example we see multiple puppies in a line creating multiple dominant images along the bottom of the image. In sharp contrast there are also four bright blue plates above the dogs creating even further dominant objects within a single image.



Figure 7-1 Multiple Dominant object image

7.2 Model Data Flow

Each of the 700 images had an approximate aspect ratio of 800 x 600 creating 480,000 pixels of metadata for each image. Due to the large scale of this data set, 3 out of every 7 images were selected from the pool of images and fixation points for the analysis resulting in 210 images and over 100-million-pixel observation points to predict fixation points. The twenty ASD subject had a total number of fixation points of (). This again leads to a very under sampled target variable when trying to predict whether a fixation occurs.

7.3 Model Performance and Discussion

The best performing models when attempting to determine a subject fixation in Chapter 4 were built using gradient boosting. This machine learning technique proved to have the best statistical performance for both ASD and TD subjects. When applying each respective gradient boosting

model to this new experiment data source and looking at performance of the top 10% highest probability of fixation cutoff the AUC and recall statistics are presented in Table 7-1. The model built on ASD children has a recall of 0.223 and AUC of 0.694, while the model built on TD children has a recall of 0.244 and AUC of 0.691. Thus, the model built on TD children does a slightly better at identifying the fixation points of the highly functional subjects in the Shuo experiment.

There is significant decay in the performance of each model where earlier applied on the new data set where we have seen the Recall degrade substantially, however using the top 10% cutoff, there is still significant lift over random when applying these models. While the subjects and the types of images used in the experiment are substantially different it is a terrific outcome that the models were predictive in this application.

Table 7-1: ASD & TD Gradient Boosting Performance on New Experiment Data

| Gradient Boosting Algorithm | Performance Metric | |
|---|---------------------------|---------------|
| | AUC | Recall |
| ASD Development Data | 0.751 | 0.331 |
| TD Development Data | 0.788 | 0.396 |
| ASD Model Applied to New Dataset | 0.694 | 0.223 |
| TD Model Applied to New Dataset | 0.691 | 0.244 |

8 Conclusion

While there is significant research done applying supervised learning on ASD subjects as documented in the literature review completed by Hyde et. Al [16], this paper was able to generate a multitude of findings. The research presented in this paper extracts several uses and modeling opportunities using the same dataset spanning from unsupervised learning techniques to supervised machine learning techniques.

When applying unsupervised learning (chapter 3), it became evident that ASD and TD subjects did in fact have many similar behaviors and variation in terms of fixation location and patterns given by the likeness between the principal components for each subject group. However, when digging deeper into the clusters created using leveraging k-means begin to show us some inherent differences in the fixation locations along with the consistency of the fixation points where ASD subjects had more variability in the locations they would fixate.

The baseline logistic regression models used to predict fixation for the respective ASD, and TD groups provided a starting point at further understanding if there are differences in key attributes that lead to fixation in chapter 4. Initially we see that color is the dominant factor, but when additional attribute transformations are created and fed into our enhanced logistic regression it becomes evident that the location of each pixel is the dominant factor in whether a subject will fixate on a given pixel. Furthermore, we see the additional value in leveraging machine learning as gradient boosting proves to be the most accurate model in predicting the likelihood of fixation on a given pixel. Across all models and building upon the insights derived from our unsupervised

learning chapter it is proven that the TD subjects and their models perform slightly better than the ASD subjects and respective models based on the reduce variability in their behavior.

Given a fixation point for either ASD or TD subject group, there was a slight ability to predict the subject group when using our baseline logistic regression approach. While the results were not nearly as predictive in chapter 5 compared to chapter 4, we were able to quantify the performance improvement that can be attributed to additional data transformations along with more advanced machine learning techniques. The results conclude that a 10.6% recall improvement could be made by leveraging gradient boosting and additional data transformation, of which 5.9% are attributed to the additional data and 4.7% is attributed to gradient boosting. This finding is key as data or method alone will not extract the full predictive capability in the modeling.

Next in chapter 6 we had a very interesting yet unsuccessful application of machine learning when attempting to predict the duration of fixation for ASD and TD subjects respectively. Much of the poor performance is attributed to the variability observed in the fixation durations themselves, however we did see model improvement when leveraging random forest and gradient boosting compared to linear regression. This model performance can be used as a baseline for future work of the same nature.

The main purpose of the research had been to extract several use cases and finding out of the given experiment performed by Duan et. Al. until the research and experiment conducted by Shuo et. Al. presented the opportunity to apply our fixation models onto a dataset comprised of significantly different images and subjects. In chapter 7, we applied our gradient boosting fixation models build on ASD and TD subjects onto a fixation dataset of subjects almost 4 times

older than they were built. While there was significant decay in model performance with recall statistics dropping, significant lift over random was observed. One hypothesizes that for all the modeling and research presented, increased predictability will come with additional information about subjects.

References

- [1] H. Duan, G. Zhai, X. Min, Z. Che, Y. Fang, X. Yang, J. Gutiérrez, P. Le Callet, "A Dataset of Eye Movements for the Children with Autism Spectrum Disorder", ACM Multimedia Systems Conference (MMSys'19), Jun. 2019
- [2] T. Judd, K. Ehinger, F. Durand and A. Torralba, "Learning to predict where humans look," 2009 IEEE 12th International Conference on Computer Vision, Kyoto, 2009, pp. 2106-2113, doi: 10.1109/ICCV.2009.5459462
- [3] Dalrymple, K.A., Jiang, M., Zhao, Q. et al. Machine learning accurately classifies age of toddlers based on eye tracking. *Sci Rep* 9, 6255 (2019). <https://doi.org/10.1038/s41598-019-42764-z>
- [4] Luz Rello and Miguel Ballesteros. 2015. Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proceedings of the 12th Web for All Conference (W4A '15)*. Association for Computing Machinery, New York, NY, USA, Article 16, 1–8. DOI:<https://doi.org/10.1145/2745555.2746644>
- [5] Carette, Romuald & Elbattah, Mahmoud & Cilia, Federica & Dequen, Gilles & Guerin, Jean-Luc & Bosche, Jérôme. (2019). Learning to Predict Autism Spectrum Disorder based on the Visual Patterns of Eye-tracking Scanpaths. 103-112. 10.5220/0007402601030112.
- [6] Pierce, K., Conant, D., Hazin, R., Stoner, R. and Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, 68(1), 101–109
- [7] Jones, W., Klin, A. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* 504, 427–431 (2013).
- [8] Simon Barthelme (2019). *Imager: Image Processing Library Based on ‘Cimg’*. R package version 0.41.2. <http://CRAN.R-project.org/package=imager>
- [9] R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [10] Max Kuhn (2020). *caret: Classification and Regression Training*. R package version 6.0-86.<https://CRAN.R-project.org/package=caret>
- [11] Marvin N. Wright, Andreas Ziegler (2017). *ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R*. *Journal of Statistical Software*, 77(1), 1-17. doi:10.18637/jss.v077.i01

- [12] Brandon Greenwell, Bradley Boehmke, Jay Cunningham and GBM Developers (2020). gbm: Generalized Boosted Regression Models. R package version 2.1.8. <https://CRAN.R-project.org/package=gbm>
- [13] Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, p. 77. DOI: 10.1186/1471-2105-12-77
- [14] Myatt, G. J., & Johnson, W. P. (2009). Making sense of data II: A practical approach to data visualization, advanced data mining methods, and applications. Oxford: Wiley-Blackwell, pp.112-114,122-124.
- [15] Shuo Wang, Ming Jiang, Xavier Morin Duchesne, Elizabeth A. Laugeson, Daniel P. Kennedy, Ralph Adolphs, Qi Zhao. Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking, *Neuron*, Volume 88, Issue 3, 2015, Pages 604-616, ISSN 0896-6273, <https://doi.org/10.1016/j.neuron.2015.09.042>.
- [16] K. Hyde, M. N. Novack, N. Lahaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead, “Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review,” *Review Journal of Autism and Developmental Disorders*, vol. 6, no. 2, pp. 128–146, 2019.