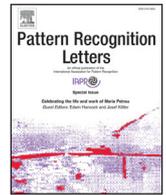




ELSEVIER

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Gaze-based classification of autism spectrum disorder

Diego Fabiano, Shaun Canavan*, Heather Agazzi, Saurabh Hinduja, Dmitry Goldgof

University of South Florida, 4202 E. Fowler Ave, Tampa, FL 33620, USA



ARTICLE INFO

Article history:

Received 22 October 2019

Revised 23 March 2020

Accepted 20 April 2020

Available online 27 April 2020

Keywords:

Machine learning

Autism spectrum disorder

Classification

Gaze

ABSTRACT

People with autism spectrum disorder (ASD) display impairments in social interaction and communication skills, as well as restricted interests and repetitive behaviors, which greatly affect daily life functioning. Current identification of ASD involves a lengthy process that requires an experienced clinician to assess multiple domains of functioning. Considering this, we propose a method for classifying multiple levels of risk of ASD using eye gaze and demographic feature descriptors such as a subject's age and gender. We construct feature descriptors that incorporate the subject's age and gender, as well as features based on eye gaze patterns. We also present an analysis of eye gaze patterns validating the use of the selected hand-crafted features. We assess the efficacy of our descriptors to classify ASD on a National Database for Autism Research dataset, using multiple classifiers including a random forest, C4.5 decision tree, PART, and a deep feedforward neural network.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social interactions (e.g., making appropriate eye contact, gestures, facial expressions), communication (e.g., reduced reciprocal conversation exchanges, poor use of gestures), and the presence of restricted interests and repetitive behaviors (e.g., rigid routines, stereotyped language, unusual behaviors) [41]. ASD is often considered a childhood disorder, yet it persists through the life span and results in severe psychosocial impairments. Among these impairments for example, is difficulty forming close relationships, which can lead to social isolation and bullying [25]. Early detection of ASD is critical to helping children overcome such disorder-related impairments. An important social interaction that develops as a part of infancy, is gaze-to-face. From a neural and cognitive systems perspective, gaze-to-face plays a critical role in creating social meaning as it acts rapidly to change attention, and arouse and modulate responses (e.g., mimicry and imitation) [19]. Children diagnosed with ASD tend to exhibit delays in this critical social behavior [15]. The salient association between eye movements, cognitive processes and general cognitive abilities makes the study of the relationship between eye gaze and ASD significant [20].

Most diagnoses of ASD occur after or around 4 years of age [3], which is long after the social behavior gaze-to-face develops. The gold standard for ASD diagnosis involves a lengthy evaluation pro-

cess that requires a certified professional to administer behavioral observation instruments in a clinical setting [43]. The lengthy evaluation coupled with the need for a specialist operating in a clinical environment contributes toward delays in early identification, as these resources are not universally available and tend to cluster in urban areas [24,26,38].

In recent years, a burgeoning cache of complex computational and engineering methodologies have been applied to neurocognitive research. One such technology, machine learning, has demonstrated promise as a reliable tool to classify young children with ASD by analyzing eye gaze patterns. Given the heterogeneity of ASD symptoms, machine learning studies have demonstrated the importance of incorporating children across the autism spectrum (e.g., severe and milder presentations), as well as typically developing children, when finding deviations in eye gaze patterns [5,36]. An accurate algorithm that incorporates various factors like age and gender, as well as patterns in eye gaze behavior can make measurements that may not be possible with the human eye, thus providing a more reliable screening tool to enhance diagnostic and intervention research in ASD. In doing so, we should account for details that regard the design and principles of the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2 [25]), a gold-standard ASD behavioral assessment, and the correct application of machine learning, as well as the limitations of datasets [5]. It is imperative to screen at-risk children (often identified by familial history and presence of developmental delays) and potentially diagnose ASD as early as possible, given that the largest treatment effects (e.g. intellectual, language, and adaptive functioning) have been linked to early intervention [10]. In the impending sections,

* Corresponding author.

E-mail address: scanavan@usf.edu (S. Canavan).

we introduce the study of gaze and ASD as they apply to medical and psychology literature, as well classifying ASD, using different modalities (e.g. gaze, MRI) with computer vision and machine learning. It is also worth noting some of the works fall under both categories. The following literature review is not meant to be exhaustive but an introduction to the most related works.

1.1. Related medical and psychology works

In medical and psychology literature, there is a wealth of material related to the study of gaze and ASD. Frazier et al. [14] performed a meta-analysis on the differences in gaze for social and nonsocial information. They used 122 independent studies and compared 1155 characteristics such as eye-tracking methods, stimulus features, and regions of interest. They found a reliable pattern of gaze abnormalities for individuals with ASD which suggested a problem with selecting socially relevant versus irrelevant information. Helminen et al. [22] used heart rate deceleration to investigate orienting response to direct versus averted gaze among 2–5-year-olds with low-functioning ASD. Their results suggest that children with ASD lack the perceptual detection advantage of direct gaze, as well as fail to respond to the gaze with enhanced physiological orienting. Pierce et al. [33] studied preferences for looking at dynamic geometric images versus social images in children ages 14–42 months. The study found that children with ASD spent significantly more time looking at dynamic geometric images suggesting that what children look at is as important as where they look. Guillon et al. [17] looked at 20 preschoolers and their typically developing peers, while scanning over faces. When typically developing children scan, they found that the left eye might be an anchor; however, children who have ASD did not have this anchor. This study highlights multiple differences in gaze patterns of children with ASD compared to typically developing (TD) children. Previous studies have used eye tracking devices to measure and categorize eye gaze patterns in children with and without ASD. Sasson et al. [40] discovered, when presented with arrays of social and non-social objects, the visual attention of children with ASD was more circumscribed (fewer images were explored), more perseverative (more time was spent on each image) and more detail-oriented (greater amount of fixation on details in images) than TD children. Bekele et al. [4] found children with ASD focus more on the forehead (an information irrelevant area) than on the mouth (information relevant) when compared to TD children in a study of their response to facial expressions while looking at a computer-generated avatar in a virtual environment.

1.2. Related computer vision and machine learning works

Recently, computer vision and machine learning techniques have been used to analyze eye contact and gaze in relation to potential risk for ASD. Ye et al. [44] proposed detecting eye contact with glasses. An examiner wears point-of-view glasses to capture video of a child's gaze direction. In this solution, the examiner's point of gaze is simultaneously captured from the glasses. In the captured videos the child's face is detected and a feature set is extracted which includes the location of the examiner's gaze point with respect to the child's eye center. A random forest was trained for eye contact detection, with a precision of 80% and recall of 72%. Seminal work by Rehg et al. [37] was done in decoding the social behavior of children, which focuses on activity recognition of children including gaze detection. They detected gaze by using an overhead Microsoft Kinect camera and a front facing camera to understand the social interactions of 121 children in over 160 sessions. This study used video, audio, physiological recordings, scoring datasheets, and parent questionnaires to analyze social behavior. This study presented the first baseline analysis of decoding so-

cial behavior of children, which showed they can reliably predict child social data with multiple modalities. The research also resulted in a new multimodal dyadic behavior dataset consisting of adult-child social interactions.

While the above listed computer vision and machine learning works have focused on social interactions, there are studies that look to analyze gaze for classification of ASD. One of the first works to do this, from Alie et al. [1], trained Markov Models to classify children at 3 years of age with and without a diagnosis of ASD. The study included 26 TD children and 6 with an ASD diagnosis. A total of 93.75% of the subjects were successfully classified. Puisol et al. [34] looked at classifying developmental disorders based on eye-movements of subjects ages 12 to 28-years-old. The study focused on subjects with a Fragile X Syndrome diagnosis, which is a common genetic cause of ASD [18]. Using temporal gaze features and a recurrent neural network, they achieved a top precision of 0.91. Liu et al. [29] identified children with ASD by evaluating their face scanning patterns. This is accomplished through eye movement patterns. They measured the frequency distribution of the gaze points without temporal information. Using K-means, they quantized these coordinates from 29 children with ASD and 29 control children. Using a leave-one-out cross-validation strategy, they achieved an average accuracy of approximately 88%.

Along with gaze, recent works have also successfully used other modalities to classify ASD in children and adults. Kong et al. [28] used subject dependent deep neural networks (DNN) trained on MRI images to learn features based on brain connectivity. Using a subset of images from the Autism Brain Imaging Data Exchange I [2], they evaluated 182 subjects (78 with ASD, 104 without), and created a separate DNN for each one. They showed encouraging results on this data, achieving an accuracy of 90.39%. Using a larger set of data from the same data exchange, Heinsfeld et al. [21] investigated detecting ASD from 505 subjects with ASD and 530 control subjects. Similar to Kong et al., they also used deep neural networks to classify the subjects with ASD. On this larger set of data, they achieved an accuracy of 70%, also showing that the DNN can outperform a support vector machine (SVM) and random forest, with those classifiers achieving an accuracy of 65% and 63%, respectively.

Using full-body, non-verbal data, Georgescu et al. [16] detected ASD in adult subjects. They manually selected regions of interest (ROI) that included the head in one and rest of body in the other. Changes in greyscale in these ROIs were used to train a support vector machine on 58 subjects (29 with ASD, 29 control). This approach achieved an accuracy of 75.9%, with a sensitivity of 96.3% (one subject with ASD was incorrectly classified). Nakai et al. [30] investigated how machine learning compares to human experts at classifying ASD, in children, using voice analysis. For the machine learning, they calculated 24 features from the fundamental frequency representing pitch. Using these features, they trained an SVM to learn the voice patterns of children with and without ASD. Their proposed approach showed a higher accuracy (F-measure) compared to human experts at classifying children with ASD using voice analysis.

Another interesting study involves using machine learning along with items (features) from the Social Responsiveness Scale (SRS). Duda et al. [9] classified subjects with ASD compared to those with attention deficit hyperactivity disorder (ADHD) using this approach. They used a crowdsourcing technique, to get extra data, to help their model generalize to unseen subjects. Training machine learning classifiers on SRS features they achieved an Area Under the Curve (AUC) of > 0.90 (classifying ASD vs. ADHD). They investigated multiple machine learning classifiers that include support vector classification, logistic regression, and linear discriminant analysis.

Table 1
Important elements in ETS-E dataset.

Element Name	Data Type	Description
Subject ID	String	ID of subject
Age	Integer	Age in months
Phenotype	String	Diagnosis
Gaze point	(x, y)	Gaze coordinate
Fixation duration	Integer	Length of fixation
Gender	String	Gender of subject
Stimuli name	String	Name of stimuli (image/video)

1.3. Motivation

Although these studies provide a great deal of information regarding classification of ASD, relatively few focus on automatically classifying ASD based on observed gaze patterns of subjects with multiple levels of risk for ASD (low, medium, high, and ASD). Most of the studies concentrate on differences between those diagnosed with ASD and their typically developing peers. In this study, we extend our proposed method for ASD classification using eye gaze and subject demographic information [7]. Raw gaze, and feature descriptors, created using subject's age, gender, and eye gaze data were tested on four different classifiers; random forest [6], C4.5 decision tree [35], PART [13], and a deep feedforward neural network. Our proposed method resulted in an ASD classification rate of 93.45% on the National Database for Autism Research (NDAR) [31]. We also present an analysis of eye gaze patterns validating the use of the selected hand-crafted features, as well as a comparison of hand-crafted vs. deep features for classification of ASD risk. Our contributions be summarized as follows:

- (1) Two sets of gaze-based feature descriptors along with demographic information such as age and gender are proposed for classification of ASD risk. One set is derived from raw eye gaze points (x, y), and the second consists of hand-crafted features (e.g. gaze velocity).
- (2) Analysis of eye gaze patterns is provided across low, medium, and high risk, as well as ASD. The analysis is conducted across age ranges that include 2–12, 13–33, and 64–132 months.
- (3) Feature descriptors (e.g. age) are analyzed to determine which are strongest for classification of ASD risk.

2. Dataset

The National Database for Autism Research (NDAR) [31] is a website that serves as the primary entry point for the National Institute of Mental Health Data Archive (NDA), which shares de-identified human subjects with the scientific community. The NDA also gives researchers access to tools, methods and forms of analysis for scientific discovery. For our experiments, we use the "Eye Tracking Subject-Experiment" (ETS-E) dataset, which is available for download through the NDA. The ETS-E dataset contains gaze information from subjects that have been classified with various levels of risk for ASD which include low, medium, and high risk, as well as an ASD diagnosis. These subjects watched images and videos on a computer screen. The dataset contains a total of 41 subjects diagnosed with a low risk, 122 subjects diagnosed with a medium risk, 49 subjects diagnosed with a high risk, and 17 subjects with a diagnosis of ASD, for a total of 229 subjects. The dataset also includes demographic information such as age (2 months to 11 years) and gender of the subjects (Table 1).

As can be seen in Table 2, there is some overlap in age among low and high-risk subjects, however, medium and ASD classes share no overlap in age with any other risk class. Considering this, we investigate the impact of age on the classification of ASD risk, by removing it as a feature (Section 4.2).

Table 2
Age distribution (in months), of ETS-E dataset.

ASD Risk	Age Range
Low	3–12
Medium	13–33
High	2–12
ASD	64–132

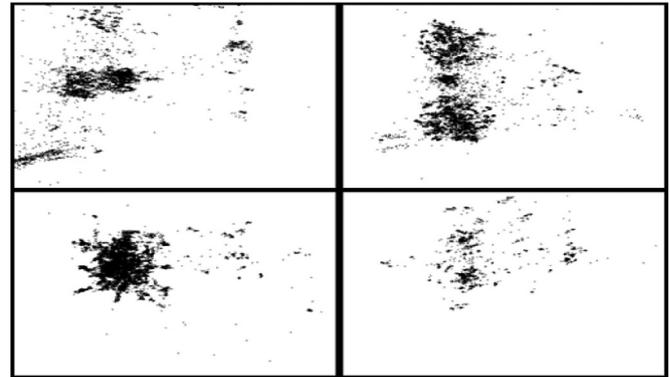


Fig. 1. Gaze maps of subjects. Top left: ASD; Top right: high risk; Bottom left: low risk; Bottom right: medium risk.

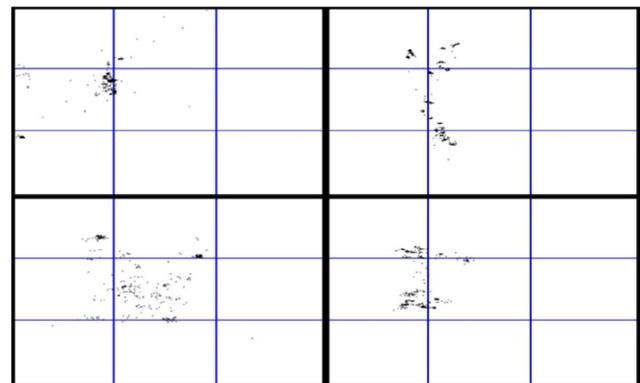


Fig. 2. 3x3 grid, of ETS-E subject data, for analyzing gaze patterns. Top row: ASD (left side) and high risk (right side); Bottom row: low risk (left side) and medium (right side).

Through observations of the subject gaze data (x and y coordinates), from each of the different risk levels in the ETS-E dataset, it can be seen that the focus of the subject decreases as the risk moves from low to high, and to an ASD diagnosis (Fig. 1). These observations are a motivation for the selected, hand-crafted features used in this work. Details on the hand-crafted features are given in the next section along with further analysis validating their use.

3. Hand-crafted feature descriptors

Given eye gaze data, to classify risk for ASD, we hypothesized that important hand-crafted features could be created based on segmentation of the gaze points into a 3x3 grid (Fig. 2). We chose to use a 3x3 grid as this corresponds to the nine possible directions of eye gaze. Tran et al. [42] used a 3x3 grid, based on the nine possible directions of eye gaze, around the gaze points for analyzing affective communication domains that include interrogation-based deception and communication skills. We are motivated by this to use a 3x3 grid for analyzing gaze points for each class of ASD risk (low, medium, high, and ASD).

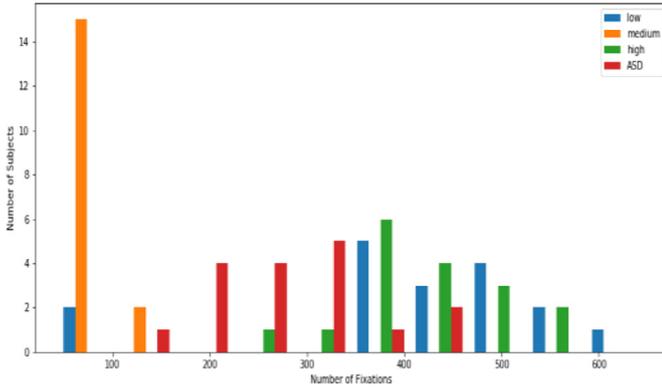


Fig. 3. Number of subject fixations per class. *Best viewed in color* (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

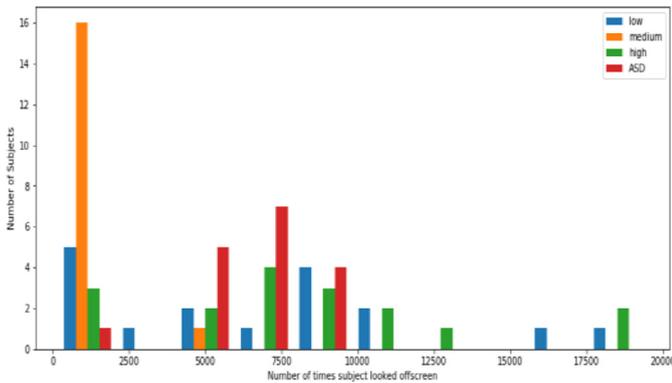


Fig. 4. Number of times subjects looked offscreen per class. *Best viewed in color* (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

3.1. Hand-crafted features

To the best of our knowledge, the NDAR ETS-E dataset is the only one available that contains gaze and demographic information about subjects that have been diagnosed with levels of ASD risk. Considering this, the following hand-crafted features and subsequent analysis (Section 3.2) is currently specific only to this dataset. Due to this lack of other datasets, it is difficult to test the generalizability of the proposed hand-crafted features to other data/subjects. However, there is supporting evidence in medical works [8,39] that justify the use of these hand-crafted features and their ability to generalize to other datasets. Motivated by this, we investigate the use of two sets of features. The first includes hand-crafted features plus the raw gaze data, the second includes hand-crafted features from patterns in the gaze. As can be seen in Table 2, each risk is associated with a specific age range that is relatively small (e.g. low risk is 3–12 months). *Due to this, the conducted evaluations have a fair distribution of age ranges, suggesting that the results are based on the gaze data and not age (i.e. the differences in age for each risk level will not impact the distributions from Figs. 3,4 and Tables 4,5).*

Features based on raw gaze. These proposed hand-crafted features use the ETS-E set and combine raw eye gaze data, as well as demographic information to classify ASD. We propose the use of four different feature descriptors that include (1) raw eye gaze points (x, y); (2) average fixation duration; (3) age; and (4) gender. Eye gaze information can detail important features that can be used to classify ASD [11]. The raw eye gaze points contain the (x, y) coordinates indicating where the gaze of each test subject was

focused. The number of gaze points per subject is not consistent. Considering this, we used the first 2580 gaze points for each subject, which corresponds to 43 s of data (at 60 fps) for each subject. This approach ensures the feature vectors are the same length and we are analyzing the same time segment for each subject (i.e. the first 43 s). It is important to note, that while Fig. 1 plots the gaze maps without visually considering temporal information, when using the raw gaze features for classification, temporal information is considered as they are used sequentially in order (i.e. time) based on the subject watching a video.

Along with the raw eye gaze data, the NDAR provides eye gaze fixation information. Average fixation was calculated by dividing the total length of fixations over the total number of fixations per subject. Average fixation length was used as subjects with ASD tend to fixate on images for a longer amount of time [40]. For example, one test subject classified with a medium risk had an average fixation length of approximately 0.6 s, while another subject at high risk had an average fixation length of 0.8 s. The last two feature descriptors are age and gender.

Features based on gaze patterns. Motivated by the visual differences between each class (low, medium, high, ASD as seen in Fig. 1), we construct a 21-dimension feature vector, based on the grid-based approach (Fig. 2), that includes features from gaze patterns and demographic information (i.e. age). The grid-based approach allows us to extract more fine-grained information about where the subjects are looking, how often they look there, and their gaze velocity, where

$$velocity = \left(\frac{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}{t} \right). \quad (1)$$

Here, $t = time$, (x_1, y_1) is the gaze point at time t_1 and (x_2, y_2) is the gaze point at time t_2 . We hypothesized that this information would have key features that can be used to classify ASD risk. Considering this, using each cell of the grid, we took the average velocity, and the frequency of gaze (total number of gaze points in each cell) of all cells (9 + 9 features) and inserted them into our feature vector. We also used the total fixations (over all cells – 1 feature), the number of times the subject was looking off screen (out of monitor – 1 feature), and age of subject (1 feature). We take each of these as features to classify ASD risk. These features give us our 21-dimension feature vector (9 + 9 + 1 + 1 + 1 = 21), which allow for accurate classification of ASD risk. To validate the use of these hand-crafted features, we next detail an analysis of the gaze data using the 3x3 grid-based approach.

3.2. Validation of hand-crafted features

Through the empirical study of the gaze maps for each of the different classifications in the ETS-E dataset, we have observed noticeable gaze patterns that subjects follow (Fig. 1). The observed patterns contain important features that can be used to classify subject's risk for ASD. From these observations we conducted in-depth analyses of the patterns. This analysis is presented as validation for the hand-crafted features detailed in Section 3.1. When analyzing risk for ASD between low, medium, high, and ASD we can study where the subjects are focusing when watching videos. To study these focus areas, we created a 3x3 grid around each of the gaze maps. Each cell of this grid corresponds to a potential area of interest on the screen where the subjects focus their attention. From the grids we can perform a more in-depth analysis of the gaze maps by focusing on smaller regions. As can be seen in Fig. 2, the focus of the low risk sample (bottom left) is mainly in the middle cell (middle of screen), while medium, high, and ASD are focused in other regions, and in some cases more than one region.

Table 3

Average age in months (with standard deviation), average number of gaze points that were not focused on the screen, and average number of fixations for each risk class.

Class	Age (SD)	Out of Monitor	Fixations
Low	6.3 (3.3)	7140.4	415.7
Med	22.2 (4.3)	1319.8	75.8
High	6.1 (3.1)	8048.3	425.6
ASD	97.6 (21.3)	6407.6	283.9

When analyzing gaze patterns, there are important features that we want to consider. A core feature of ASD is diminished gaze fixation [8], and it has been found that saccadic velocity was diminished in children with ASD or autistic-like conditions [39]. Based on these two important aspects of ASD, we analyzed the following gaze patterns: (1) fixation of subjects; (2) frequency of gaze for each cell of our grid (how often each subject looked at a cell); (3) average velocity of the subject's gaze per cell; and (4) the number of times the subject was not looking at the screen, which was determined by verifying if the gaze was in the bounds of the screen size used. From these patterns we can see how often the different classes look at certain regions of the screen (or if they are looking off screen), if they change their fixation on the screen, and how their velocity changes per class of diagnosis. Fig. 3 shows the number of fixations per class, and Fig. 4 shows how often each class was looking off screen. Tables 4 and 5 illustrate the average gaze velocities and frequencies for each cell respectively. Table 3 shows the average subject age, frequency of looking off-screen, and the number of fixations.

As can be seen in Table 4, the velocities change with each class. Low has the smallest average velocity, across all cells, at 0.7 and a standard deviation of 0.3. This shows that subjects with a low risk for ASD are slowly scanning the screen with little variation (i.e. they are following the video closely). Those subjects with an ASD diagnosis have the highest average velocity at 1.1 and a higher standard deviation of 0.5. This suggests that subjects with an ASD diagnosis more rapidly scanned the screen, and this rapid scanning was not consistent (e.g. they make quick movements with their eyes to different parts of the screen). An interesting class to look at is medium risk as there are some outliers that effect that analysis. Most of the velocities for all classes are less than 4; however, in the medium class there are large outliers in the bottom 3 cells of the grid. These velocities are 7.7 in the bottom left cell, 5.9 in the bottom middle cell, and 14.5 in the bottom right cell. These outliers have a large effect on the overall average velocity of 0.9 and standard deviation of 0.6 (see Table 4). Due to these outliers the class has the largest standard deviation and its average velocity is higher than both the low and high classes. A possible explanation for this is that a small subset of subjects in medium class are exhibiting patterns like ASD (e.g. eyes quickly darting to cell location on screen).

There are also interesting gaze frequencies (Table 5), for each of the classes in the specific cell of the grid. Most of the time was spent looking at the center of the screen (middle cell), the differences lie in the frequencies of the other cells. The subjects in the

low class exhibited the smallest standard deviation of frequency across all cells, at 37.1. The average gaze frequency for the center of the screen was 120.6 with the bottom middle and top left cells having the second and third highest frequencies, respectively. The low standard deviation is showing consistency across all cells. This suggests that the low-class scans across the screen in a consistent manner, also supported by the consistent velocity across all cells for this class. Those subjects with an ASD diagnosis have the highest standard deviation of frequency gaze with 90.8, which when combined with the high average velocity shows subjects that make fast “jumps” in gaze from the center of the screen to another random cell, and back to the center again.

We also analyzed how many times the subject was not focused on the screen and the total number of average fixations (Table 3). Low, high, and ASD classes exhibits similar characteristics with the number of times they were not focused on the screen, however, ASD subjects had noticeably different fixations with an average of 283.9 fixations versus the 415.7 and 425.6 fixations for low and high respectively. Again, subjects with a diagnosis of medium risk showed a noticeable difference in patterns compared to the other classes. Medium had an average of 1319.8 gaze points not focused on the screen, and only 75.8 average fixations. While these outliers exist, in the medium class, we will show in the next section that they do not negatively impact classification of ASD risk.

By analyzing gaze points for ASD risk we have observed patterns that can be useful for classifying risk of ASD. A summary of these observed patterns is detailed below.

- (1) Average gaze velocity, generally, increases as subjects move from low to high risk, and to an ASD diagnosis.
- (2) Average gaze velocity across the viewing screen is more consistent in classes with less risk of ASD compared to those with a diagnosis of ASD.
- (3) Subjects with a diagnosis of ASD display patterns of rapid eye movement across the screen, while lower risk subjects consistently scan the screen.
- (4) The frequency of gaze points changes from class to class, as well as where the subjects focus. This, along with velocity, is consistent with previous literature which has shown that where children are looking and at what, matters [33].

4. ASD classification

4.1. Classification techniques

Based on the hand-crafted features from the raw gaze data, and the gaze patterns detailed in Section 3, we trained a random forest, C4.5 and PART classifiers, and a deep feedforward neural network. For our proposed classification method, we used 10-fold cross validation for training and testing. Justification and details on each of the classifiers is given below.

Random forests. A random forest [6] is a collection of random classification trees where the mean classification of all trees is taken as the output. They have been used in a wide variety of classification tasks such as real-time hand gesture recognition [45],

Table 4

Average gaze velocities for each of the grid cell locations for low, medium, high, and ASD risk classes. The last 2 columns show the average velocity across all cells, and the standard deviation respectively.

Class	Bottom Left Velocity	Middle Left Velocity	Top Left Velocity	Bottom Mid Velocity	Middle Mid Velocity	Top Mid Velocity	Bottom Right Velocity	Middle Right Velocity	Top Right Velocity	Across All Cells	STDEV
Low	0.9	0.8	0.1	0.7	0.7	0.7	1.2	0.9	0.5	0.7	0.3
Med	1.1	0.7	-0.4	1.5	1.1	-0.03	2.1	0.8	-0.7	0.9	0.6
High	0.9	0.9	0.4	0.7	0.9	0.5	1	0.8	0.7	0.8	0.2
ASD	1.8	1.6	0.1	1.4	1	1.2	1.5	0.6	0.9	1.1	0.5

Table 5

Average gaze frequencies for each of the grid cell locations for low, medium, high, and ASD risk classes. The last 2 columns show the average frequency across all cells, and the standard deviation respectively.

Class	Bottom Left Freq	Middle Left Freq	Top Left Freq	Bottom Mid Freq	Middle Mid Freq	Top Mid Freq	Bottom Right Freq	Middle Right Freq	Top Right Freq	Across All Cells	STDEV
Low	8.8	22.5	48.1	55.8	120.6	15	16.5	5	10.6	33.7	37.1
Med	5.7	37.4	24.1	32.2	230.6	4.4	0.7	39.1	2.8	41.9	72.4
High	10.7	16.8	37.6	59.8	155.2	37.9	6.2	20.5	2.9	38.6	47.4
ASD	2.5	9.6	14.6	32.5	283.8	19.4	3.7	6.5	12.6	42.8	90.8

Table 6

Raw-gaze ASD classification accuracies.

	Random Forest	PART	C4.5	Neural Network
Classification Accuracy	86.43%	90.83%	90.83%	85.18%

and head pose estimation [12]. They can classify diverse types of data making them a good fit for classifying ASD risk.

C4.5. C4.5 [35] is a statistical classifier that builds decision trees based on information entropy. At each node of the tree, the algorithm looks to split the subsets based on the most information gain. From all the available attributes, the one that has the highest information gain makes the decision. The algorithm then recursively does this for each of the available subsets that are left. The C4.5 algorithm can handle both discrete and continuous data making them a good fit for classifying ASD risk.

PART. PART [13] makes use of partial decision trees and a separate and conquer strategy for rule induction. For PART to make a rule it builds a pruned decision tree with the current set of instances, which have not been removed, and the leaf that has the largest amount of coverage is turned into a rule. The rest of the tree is then discarded. In doing this, it can avoid over-pruning making them a good fit for classifying ASD risk.

Deep feedforward neural network. There are many variants of deep neural networks such as convolutional neural networks and recurrent neural networks. For this study, we use a deep feedforward neural network to train and classify risk of ASD. They use hidden layers between the inputs and outputs. The weights on incoming connections of these hidden layers are adapted to learn features to predict the correct output of input data. These networks are a natural fit for our experiments as they have successfully been used on a variety of classification tasks [23].

Our network has an initial input layer with the same number of neurons as the input vector, one hidden layer where the number of neurons = $\lfloor (\text{number of neurons in input layer} + \text{number of neurons in output layer})/2 \rfloor$, and the final output layer output layer where the number of neurons = 4 (the number of classes to predict – Low, Medium, High, ASD). The softmax activation function was used, and the adamax optimizer [27] with a learning rate of 0.001. All features used with network were normalized by removing the mean and scaling to unit variance.

4.2. Results

Hand-crafted features from raw gaze. Using raw gaze-based features for training data, a max classification rate of 90.83% from both the PART and C4.5 classifiers was achieved. Table 6 shows the accuracies for each of the tested classifiers, and Tables 7–9 show the confusion matrices for each.

For each of the classifiers tested, medium (out of low, medium, and high risk) has the highest classification rate, with only one misclassification among all the tests. This can partially be attributed to the relatively large amount of data available for medium (122 compared to 41 low, 49 high, and 17 ASD). For PART

Table 7

Confusion matrix using raw gaze-based features with the PART and C4.5 classifiers.

	low	medium	high	ASD
Low	32	0	9	0
Medium	0	121	0	1
High	11	0	38	0
ASD	0	0	0	17

Table 8

Confusion matrix using raw gaze-based features with the random forest classifier.

	low	medium	high	ASD
Low	29	5	7	0
Medium	0	122	0	0
High	9	7	33	0
ASD	0	3	0	14

Table 9

Confusion matrix using raw gaze-based features with a deep feedforward neural network.

	low	medium	high	ASD
Low	31	3	7	0
medium	0	122	0	0
high	3	7	39	0
ASD	0	2	1	14

Table 10

Gaze pattern ASD classification accuracies.

	Random Forest	PART	C4.5	Neural Network
Classification Accuracy	93.45%	91.70%	88.20%	92.59%

and C4.5, the ASD diagnosis was successfully classified 100% of the time, with random forest and the deep network successfully classifying 82% of the ASD data. These results are encouraging as they show that gaze can be used as a marker for diagnosis of ASD [3]. It is also important to note that while a deep feedforward neural network had a lower accuracy these results are still encouraging. Deep neural networks can require a large amount of data to train, where we only used 229 instances of the gaze-based features to train.

Hand-crafted features from gaze patterns. Using features derived from gaze patterns for training data, a max classification rate of 93.45% using a random forest. Table 10 shows the accuracies for each tested classifier. As we also tested hand-crafted features in a deep network, we make note that this is not common compared to using raw gaze data. This was done to directly compare raw gaze vs. gaze patterns. See Tables 11–14 for the confusion matrices of the four tested classifiers.

Cases where a misclassification occurs in an instance such as high being misclassified as medium can, at least partially, be attributed to the classes displaying similar features. The interesting cases of the misclassified risks come from the classes that should

Table 11

Confusion matrix for ASD risk classification, based on 21-dimension feature vector and random forest classifier.

	Low	medium	High	ASD
Low	35	1	5	0
Medium	0	122	0	0
High	5	1	41	2
ASD	0	1	0	16

Table 12

Confusion matrix for ASD risk classification, based on 21-dimension feature vector and deep feedforward neural network.

	Low	medium	High	ASD
Low	35	0	6	0
Medium	1	120	0	1
High	7	0	42	0
ASD	0	1	0	16

Table 13

Confusion matrix for ASD risk classification, based on 21-dimension feature vector and PART.

	Low	medium	High	ASD
Low	33	0	8	0
Medium	0	121	0	1
High	10	0	39	0
ASD	0	0	0	17

Table 14

Confusion matrix for ASD risk classification, based on 21-dimension feature vector and C4.5.

	Low	medium	High	ASD
Low	30	0	11	0
Medium	0	121	0	1
High	15	0	34	0
ASD	0	0	0	17

be far apart (e.g. high vs. low). One possible explanation for this is some of the features used for classification are not well separated by the classes. We study this in more detail, by not only looking at the classification rate, but also by considering which features are most important. To study this, we calculate the correlation between each attribute and the classification output using the Pearson correlation coefficient [32]. By looking at the correlation between these samples (each attribute/feature and the output), we can rank the features based on highest correlation (where 1 is the highest correlation and closer to 0 signifies a low correlation). This ranking can give us insight into how to reduce the dimensionality of our feature vector further. As can be seen in Table 15, fixation has the highest correlation with 0.6816, which is consistent with findings that problems with gaze fixations are a core feature of ASD [8]. See Table 16 for error statistics.

Along with overall classification rates, we also wanted to know which feature gave us the lowest rate. In other words, can we determine features not suitable for classification of ASD risk. For the ETS-E dataset this feature is the frequency of gaze in the bottom left cell. When this feature alone is used for classifying ASD risk, it results in 62.0087% of the instances being correctly classified (142 of 229). It is also interesting to look at the worst case, as this can give us insight to what is (or in this case is not) happening on screen. The mean frequency of gaze in this cell is 57.127 which implies that nothing of importance was happening on the screen in this cell and the gaze from all classes was not drawn to this region of interest. This cell also has a higher max velocity of gaze than many of the other cells. One explanation for this is when subjects

Table 15

Correlation between features and classification output. Close to 1 signifies a high correlation.

Correlation	Feature
0.6816	Fixation
0.5988	Middle Mid Frequency
0.5681	Bottom Mid Frequency
0.5613	Out of Monitor
0.5107	Top Left Frequency
0.4431	Top Mid Frequency
0.3018	Top Right Velocity
0.2784	Top Mid Velocity
0.2712	Top Right Frequency
0.2618	Bottom Left Frequency
0.2595	Bottom Right Frequency
0.2319	Left Middle Frequency
0.209	Middle Mid Velocity
0.1948	Age
0.1884	Middle Right Frequency
0.1749	Bottom Mid Velocity
0.0873	Top Left Velocity
0.0837	Middle Right Velocity
0.0758	Middle Left Velocity
0.05	Bottom Left Velocity
0.019	Bottom Right Velocity

Table 16

Error statistics for ASD classification based on proposed method and 9-dimension feature vector.

Statistic	Statistic Rate	
	9 dimensions	21 dimensions
Kappa statistic	0.9586	0.8962
Mean absolute error	0.0665	0.072
Root mean squared error	0.1536	0.1578
Relative absolute error	20.9483%	22.6602%
Root relative squared error	38.6085%	39.6602%

Table 17

Classification results on both raw gaze and gaze pattern features when age is removed as a feature.

	Random Forest	PART	C4.5	Neural Network
Raw Gaze	86.03%	80.35%	79.04%	86.92%
Gaze Patterns	92.58%	88.21%	87.77%	91.70%

did look at this cell, they quickly looked to and focused on another region of interest.

Impact of age on classification results. As can be seen in Table 2, the ETS-E dataset is imbalanced in terms of age. While low and high risk have similar age ranges, the medium and ASD classes do not overlap with any other age range. Due to this, we can't directly perform classification on individual age ranges as the classes would be the same for specific age ranges (e.g. ≥ 64 months would all have class label of ASD). Considering this, it is important to verify if age or the gaze, is causing the classification results that are shown with both the features from raw gaze, and the gaze patterns. To facilitate these experiments, we ran the same machine learning classifiers (e.g. random forest) on the features from raw gaze and patterns, however, we removed age as a feature in both of the sets. As can be seen in Table 17, age is a factor in classification, however, the gaze-based features, excluding age, are still able to classify ASD risk. When age was removed as a feature, across the 4 tested classifiers, the average classification accuracy decreased by 5.23% and 1.42% for the raw gaze and gaze pattern feature sets, respectively. These results suggest that age is a factor, although the gaze-based features are still able to accurately classify risk for ASD. It is also important to note that when age was removed as a feature in the raw gaze data, the feed-forward neural network accuracy increased by 1.74%. Overall, these results suggest that the fea-

tures from gaze patterns are a more robust solution to classifying risk of ASD from gaze, as the overall decrease in accuracy was less compared to the raw gaze feature set.

5. Discussion

In this paper we have proposed a new method for classification of ASD risk based on hand-crafted features and raw gaze data. This work has the potential to aid medical practitioners in early diagnosis of ASD. To facilitate this classification, we have investigated the potential use of four different machine learning classifiers (PART, C4.5, random forest, and a deep feedforward neural network). We validated the use of these hand-crafted features by showing that low, medium, high, and ASD classes display different patterns that can be used for classification of ASD risk. Our analysis reflects the studies conducted in medical and psychology literature which show gaze fixation being important as well as where and at what the subject is looking, further validating their use.

We have shown that while age is an important factor in classifying risk of ASD, in the ETS-E dataset, the proposed gaze-based features can still accurately classify risk when age is not included as a feature. Using raw gaze points, we achieved a max classification accuracy of 90.83% using both PART and C4.5 classifiers. When age was removed as a feature, the max accuracy decreased to 86.03%, however, this time with a random forest. Using features from gaze patterns we achieved a max classification accuracy of 93.45% and 92.58%, using a random forest, with and without age as a feature, respectively. These results suggest that hand-crafted features may have more discriminative power to classify risk of ASD compared to using raw gaze points. It is interesting to note that medium and ASD classes were successfully classified 100% of the time, with at least one machine learning approach, using both raw gaze and gaze patterns, however, low and high each had instances that were misclassified as the other. These inaccuracies require further investigation, and some interesting questions have developed from these observations such as: Do certain classes display characteristics of other classes (e.g. are there instances in the high class where the subject focuses in a similar manner as low)? If so, can more advanced features be used to accurately split these classes? We did not look at how age affects gaze (and class distribution) over time. This leaves the question: Will a subject's gaze change over time, ultimately classifying them differently (i.e. medium class today to high class in a years' time)? This question could lead to strong gaze patterns that can more accurately classify, and ultimately predict risk of ASD.

To the best of our knowledge the database used in this study is the only publicly available gaze database for subjects with ASD and ultimately, we want to develop a general set of features that cannot only be used to classify risk, but more importantly be used to predict ASD risk. From this work, we are especially encouraged by the results of a deep feedforward neural network. Considering this, we will collect a large corpus of multimodal data for classifying and predicting risk of ASD. This corpus of data will include gaze, facial affect, and sound collected from children up to 36 months of age (i.e. early diagnosis).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Allie, et al., Analysis of eye gaze pattern of infants at risk of autism spectrum disorder using Markov models, WACV (2011).

[2] Autism Brain Imaging Data Exchange I. [Online] Available: http://fcon_1000.projects.nitrc.org/indi/abide/abide_1.html. Accessed: 2-1-2020.

[3] J. Baio, et al., Prev. of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, *MMWR* 67 (6) (2018) 1–23.

[4] E. Bekele, et al., Understanding how adolescents with autism respond to facial expressions in VR environments, *IEEE Trans. Visual. Comput. Graphics* 19 (4) (2013) 711–720.

[5] D. Bone, et al., Use of ML to improve autistic screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion, *JCPP* 57 (8) (2016) 927–937.

[6] Leo. Brieman. Random forests, *Mach. Lea.*, 45(1): pp. 5–32, 2011.

[7] S. Canavan, et al., Combining gaze and demographic feature descriptors for autism classification, *ICIP* (2017).

[8] K. Dalton, et al., Gaze Fix. and the Neural Circuitry of Face Proc. in Autism, *Nat. Neurosci.* 8 (4) (2005) 519–526.

[9] M. Duda, et al., Crowdsourced validation of a machine-learning classification system for autism and ADHD, *Transl. Psychiatry* 7 (5) (2017) e1133–e1133.

[10] A. Estes, et al., Long-term outcomes of early intervention in 6-year-old children with ASD, *J. Am. Acad. Child Adolesc. Psychiatry* 54 (7) (2015) 580–587.

[11] T. Falck-Ytter, et al., Eye tracking in early autism research, *J. Neurodev. Disord.* 5 (1) (2013) 28.

[12] G. Fanelli, et al., Real Time Head Pose Estimation With Random Regression Forests, *CVPR*, 2011.

[13] E. Frank, I. Witten, Generating accurate rule sets without global optimization, *Mach. Learn.* (1998).

[14] T. Frazier, et al., A meta-analysis of gaze differences to soc. and nonsocial info. between ind. with and without autism, *J. Am. Acad. Child Adolesc. Psychiatry* (2017).

[15] D. Gangi, et al., Gaze to faces across interactive contexts in infants at heightened risk for autism, *Autism* (2017).

[16] A. Georgescu, et al., Machine learning approaches to study social interaction difficulties in ASD, *Front. Robot. AI* 6 (2019) 132.

[17] Q. Guillon, et al., The importance of networking in autism gaze analysis, *PLoS ONE* 10 (10) (2015).

[18] Paul Hagermann, The fragile x prevalence paradox, *J. Med. Genet.* (2008).

[19] A. Hamilton, Gazing at me: the importance of social meaning in understanding direct-gaze cues, *Phil. Trans. Royal Soc. B* 371 (1686) (2016) 20150080.

[20] M. Hayhoe, D. Ballard, Eye movements in natural behavior, *Trends Cogn. Sci.* 9 (4) (2005) 188–194.

[21] A. Heinsfeld, et al., Identification of autism spectrum disorder using deep learning and the ABIDE dataset, *NeuroImage* 17 (2018) 16–23.

[22] T. Helminen, et al., Atypical psychological orienting to direct gaze in low-functioning children with autism spectrum-disorder, *Autism Res.* 10 (5) (2017) 810–820.

[23] G. Hinton et al. Improving neural nets. by preventing co-adaption of feature detectors, *arXiv preprint arXiv:1207.0580*, 2012.

[24] A. Ip, et al., Factors influencing autism spectrum disorder screening by community pediatricians, *Pediatric Child Health* 20 (5) (2015) 20–24.

[25] I. Kamp-Becker, et al., Diagnostic accuracy of ADOS and ADOS-2 in clinical practice, *ECAA* (2018) 1–15.

[26] T. King, et al., Implementing devel. screen. and ref: lessons learned from a national proj, *Pediatrics* 125 (2) (2010) 350–360.

[27] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[28] Y. Kong, et al., Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier, *Neurocomputing* 324 (2019) 63–68.

[29] W. Liu, et al., Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework, *Autism Res.* 9 (8) (2016) 888–898.

[30] Y. Nakai, et al., Detecting abnormal word utterances in children with autism spectrum disorders: machine-learning-based voice analysis versus speech therapists, *Percept. Mot. Skill.* 124 (5) (2017) 961–973.

[31] National Database for Autism Research. [Online] Available: <https://ndar.nih.gov/>, 2017. [Accessed 18-Sept-2017].

[32] K. Pearson, Note on Reg. and Inheritance in the Case of Two Parents, in: *Proc. of the Royal Soc. of London*, 1895, pp. 240–242.

[33] K. Pierce, et al., Pref. for geometric patterns early in life as a risk factor for autism, *Arch. Gen. Psych* 68 (1) (2011) 101.

[34] G. Puisol, et al., Vision-based class. of devel. disorders using eye movements, in: *Intl. Conf. on Medical Image Computing and Computer-assisted Intervention*, 2015, pp. 317–325.

[35] J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publisher, 1993.

[36] J. Rehg, in: *Behavioral Imaging and Autism*, Pervasive Computing, 2014, pp. 84–87.

[37] J. Rehg, et al., Decoding children's social behavior, *Comput. Visi. Patt. Recogn.* (2013) 3414–3421.

[38] R. Rhoades, et al., The importance of physician knowledge of ASD: results of a parent survey, *BMC Pediatr.* 7 (37) (2007).

[39] U. Rosenhall, et al., Oculomotor Findings in Autistic Children, *J. Laryngol. Otol.* 102 (5) (1988) 435–439.

[40] N. Sasson, et al., Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays, *Autism Res.* 1 (1) (2008) 31–42.

- [41] B.J. Tonge, et al., A review of evidence-based early int for behav prob in chil with ASD: the core compq of effective programs, child-focused interventions and comprehensive treatment models, *Curr. Opin. Psychiatry* 27 (2) (2014) 158–165.
- [42] M. Tran, et al. Are you really looking at me? A Framework for Extracting Interpersonal Eye Gaze from Conventional Video. *arXiv preprint arXiv:1906.12175*, 2019.
- [43] D. Wacker, et al., Conducting functional communication training via telehealth to reduce the problem behavior of young children with autism, *J. of Dev. and Phys. Dis.* 25 (1) (2013) 35–48.
- [44] Z. Ye, et al., Detecting eye contact using wearable eye-tracking glasses, in: *ACM Conf on Ubiq. Computing*, 2012, pp. 699–704.
- [45] X. Zhao, et al., Real-time hand Gesture Detection and Rec. by Random Forest, in: *Comm. and Info. Processing*, 2012, pp. 747–755.