Detecting Autism Based on Eye-Tracking Data from Web Searching Tasks

Victoria Yaneva
University of Wolverhampton
Wolverhampton, UK
v.yaneva@wlv.ac.uk

Le An Ha
University of Wolverhampton
Wolverhampton, UK
Ha.L.A@wlv.ac.uk

Sukru Eraslan
Middle East Technical University, Northern Cyprus Campus, Kalkanli, Guzelyurt, Mersin 10, Turkey
seraslan@metu.edu.tr

Yeliz Yesilada
Middle East Technical University, Northern Cyprus Campus, Kalkanli, Guzelyurt, Mersin 10, Turkey
yyeliz@metu.edu.tr

Ruslan Mitkov
University of Wolverhampton
Wolverhampton, UK
r.mitkov@wlv.ac.uk

ABSTRACT

The ASD diagnosis requires a long, elaborate, and expensive procedure, which is subjective and is currently restricted to behavioural, historical, and parent-report information. In this paper, we present an alternative way for detecting the condition based on the atypical visual-attention patterns of people with autism. We collect gaze data from two different kinds of tasks related to processing of information from web pages: Browsing and Searching. The gaze data is then used to train a machine learning classifier whose aim is to distinguish between participants with autism and a control group of participants without autism. In addition, we explore the effects of the type of the task performed, different approaches to defining the areas of interest, gender, visual complexity of the web pages and whether or not an area of interest contained the correct answer to a searching task. Our best-performing classifier achieved 0.75 classification accuracy for a combination of selected web pages using all gaze features. These preliminary results show that the differences in the way people with autism process web content could be used for the future development of serious games for autism screening. The gaze data, R code, visual stimuli and task descriptions are made freely available for replication purposes.

CCS Concepts

•Social and professional topics → User characteristics; People with disabilities; •Human-centered computing → User studies; Web-based interaction; Empirical studies in accessibility; Accessibility design and evaluation methods;

W4A '18, April 23–25, 2018, Lyon, France

ISBN 978-1-4503-5651-0/18/04…$15.00

DOI: https://doi.org/10.1145/3192714.3192819

Keywords

Autism, Eye Tracking, Web, Screening, Diagnostic Classification

Open Data

The eye-tracking data, code, and materials are available in our external repository at: https://github.com/victoria-ianeva/Gaze-Data-from-Web-Searching-Tasks.

1. INTRODUCTION

The early diagnosis of most conditions is crucial for their treatment and for the success of therapeutic interventions. One such neurodevelopmental condition is Autism Spectrum Disorder (ASD), which affects communication and social interaction, as well as the processing of sensory information and attention shifting [3]. Different therapies, when applied early enough, have been shown to improve cognitive performance, language skills, and adaptive behavior skills in people with autism and thus to contribute to a better life for them and their families [37, 22].

The ASD diagnosis requires a long, elaborate, and expensive procedure, which is subjective and is currently restricted to behavioural, historical, and parent-report information [7, 16]. This is because to date there is no biomarker which could objectively indicate the presence or absence of autism. Different levels of autism severity are exhibited through variation in developmental level, language ability and IQ and further complicate the presentation of symptoms [23]. As a result of the subjectivity of the diagnosis, many people who are on the spectrum either never receive a diagnosis or receive a diagnosis very late in their lives (e.g. in their fourties, fifties or sixties). In addition, screening for autism in adults has been identified as a particularly challenging area [29].

Due to the diagnosis being based on observing behaviour which can have a plethora of root causes, many a time people on the autism spectrum have initially received a wrong diagnosis (e.g. intellectual disability, hyperactivity, etc) or no diagnosis at all. Both these scenarios can significantly damage a person's self-perception and diminish their chances of benefiting from suitable intervention. For example, a study by

Davidovitch *et al.* observed a cohort of 221 patients whose initial comprehensive developmental evaluations were negative for ASD but who were subsequently diagnosed after the age 6 [10]. The authors state that the cohort "underwent a total of 1028 developmental evaluations before the age of 6" and conclude that "subsequent late diagnosis of ASD after an initial ASD-negative comprehensive assessment is a common clinical experience" [10]. This observation illustrates the difficulty of receiving an ASD diagnosis early in life even in cases where the child's behaviour has been brought to the attention of clinicians.

Last but not least, the cost of the diagnostic procedure is high both for families and for the economy. In terms of health care resources, Murphy *et al.* point out that, as of 2011, clinical services for autism in the United Kingdom are experiencing "very significantly increased demand; but that just over 50% of people seeking a diagnosis from one expert service do not have ASD" [30]. At the same time a delay in diagnosis is associated with an indirect increased financial burden to families [20, 9]. While this is the case in high-income countries, people in developing countries have even more limited access to autism screening, diagnosis, and treatment [12, 2, 34]. There are a number of barriers perpetuating this imbalance, including high cost of proprietary tools for diagnosing autism and for delivering evidence-based therapies, and the high cost of training of professionals and para-professionals to use the tools [12].

One way to change this situation is to provide an accessible and unobtrusive screening method for autism which relies on behavioural data from everyday tasks. In this paper we use a machine-learning classifier trained on gaze data obtained from two different tasks where a group of people with ASD (N = 15) and a control group of neurotypical participants (N = 15) were asked to search for information within web pages. The rationale behind this approach is that the different attention-shifting mechanisms of the two groups are revealed through the gaze data and thus attention differences are used as a marker of the condition. Our method is based on logistic regression with 100-fold cross validation, where data from 10 random participants per group is used for training and then the unseen data from the remaining 5 participants per group is used for testing. We used a number of gaze-based features as well as other features related to the web pages and the areas of interest. We explore the effects of two independent tasks on eliciting attention-shifting differences between the two groups, namely "browsing" and "searching" of the web pages. We also explored different approaches to defining the Areas of Interest (AOIs) on the classification performance, where we distinguish between task-specific and task-independent AOIs. Furthermore, we investigate the effects of participant gender and the type of information contained in the AOI.

Contributions: The main contributions of this study are that:

- To the best of our knowledge, it is the first study to investigate whether gaze data could be successfully used to screen for risk of autism.
- We explore the effects of task type, approaches to defining areas of interest, page visual complexity, and participant gender on the classification performance.
- The experiments are fully replicable as the gaze data, task materials, and R code are made freely available.

The rest of this paper is organised as follows. The next Section 2 discusses visual attention in autism and related work on training classifiers for detecting autism. Section 3 outlines the main idea of our approach and Section 4 describes the process of data collection. After that Section 5 presents the set up for the classification experiments and Section 6 presents the results, which are then discussed in Section 7. Section 8 contains a summary of the conclusions.

2. BACKGROUND

This section presents related work on visual attention in people with autism and on previous studies using machine learning to classify people with and without autism.

2.1 Visual Attention in Autism

Autism is a heterogeneous condition. While people who are on the autism spectrum may largely differ in their levels of ability and the types of difficulties they encounter, one of the commonalities between people from various degrees of autism severity is that their attention patterns often differ from the attention patterns of people without autism [17].

Differences in attention shifting are used as a core component in some of the theories aiming to explain autistic behaviour. The Weak Central Coherence Theory (WCCT) for example, posits that "the ASD cognitive profile is biased towards processing local sensory information with less account for global, contextual and semantic information" [19]. This means that people who are on the autism spectrum tend to have a bottom-up approach to information processing, whereby they would focus on small details, often to the exclusion of the "bigger picture". Another explanation of certain autistic traits known as the stimulus overselectivity phenomenon [27] refers to the observation that in many situations autistic individuals may rely on only one sensory modality, while several are relevant to a task. Other autism scholars call this neglect of part of the sensory information "tunnel vision" [32].

In terms of visual attention in particular, several research studies indicate that: i) differences in visual attention between people with autism and neurotypical control groups can be detected for a variety of tasks, and that ii) these differences can successfully be captured using eye tracking [8, 25, 38]. Eye tracking is a process where an eye tracking device measures the point of gaze of an eye (fixation) or the motion of an eye (saccade) relative to the head and a computer screen [11]. The durations of fixations and their positioning could then be used as a proxy for identifying the focus of attention and for measuring cognitive load [24]. In [8] the authors investigate joint attention, which is described as the "ability to coordinate visual attention with another person and then shift the gaze toward a shared object or event" and is known to be deficient in many people with autism. The results show that these differences had neural basis, however, brain activity could be changed after rehabilitative treatment and that these changes correlated with changes in the eye-tracking data. Another study used both MRI data and eye-tracking data to investigate attention to the eye region (known to be avoided by people with autism) in a group of 294 adults with autism [25]. The results indicated that the differences in the visual attention to the eye region could be due to a decrease of the volume of the right anterior cerebellum in subjects with ASD compared to controls. The study also found significant correlations between the gaze fixations to the eyes and the volume of the relevant part of the brain, suggesting that "eye tracking may be a promising neuro-anatomically based stratifying biomarker of ASD" [25]. This finding is particularly important for our study, suggesting that differences in attention recorded through the use of eye-tracking data correlate to differences in the brain morphology of the two groups.

Finally, some differences in the attention of people with and without autism to text documents and images have been reported. In [38] the authors investigate the processing of easy-to-read documents by adults with and without autism. The documents contain both text paragraphs and images, used as comprehension cues. The results show that the group with autism would spend longer focusing on the images in the text-image pairs compared to the control group and that they would have different preferences to the use and positioning of the images in the easy-to-read documents. In another study by [35] the authors use gaze data from a reading task and show that the lexical properties of words have an effect on different durations of the viewing time per word in autistic and neurotypical readers.

The next section provides a more detailed look into the use of fMRI and other data in combination with machine learning approaches in order to differentiate between people with and without autism.

2.2 Classifying Autism based on fMRI and EEG Data

Several attempts to classify people with and without autism using functional Magnetic Resonance Imaging (fMRI) have been reported. Many of them use data from resting-state fMRI recordings, meaning that the participants are not completing a given task but instead the data is recorded while they are resting inside the fMRI machine. This allows for comparability of results and for using data collected in different research centers as an unseen evaluation set. The majority of these works use the topology of networks as features to train classifiers. The experiments achieve varying accuracy such as 79% using leave-one-out cross-validation and 71% on a validation set in [4] for 40 ASD and 40 Control participants. Other studies report 76.7% for cross-validation in [31], and 78% for cross-validation and 83% for a validation data set in [36]. Another study described in [7] uses the temporal aspects of fMRI data as features instead of their topology and achieves 86.7% accuracy using leave-one-out cross-validation and 80% when classifying an independent data set. The number of participants in that study is 12 participants with ASD and 12 Control participants for the training data set and 12 ASD and 18 Control participants for the test set. It is interesting to note that the authors of this study report 100% sensitivity (recall) and 66.7% specificity (precision) for the classification of the independent data set.

Other types of data previously used for training autism-screening classifiers is data from electroencephalograms (EEG). In a study described in [21], the authors use EEG data to distinguish between neurotypical control participants, participants with epilepsy and participants with autism. The reported classification accuracy for autism versus control is 94%, however, it is crucial to note that the evaluation technique used in this study was different from the one used in our experiment and the fMRI studies described above. In these experiments, no data from the participants from the test set was used for training (i.e. the test and training sets were divided on the basis of participants). This is not the case with the EEG study where the authors used 10-fold cross-validation where "the data is randomly partitioned, into 10 subsets with the same size and with the same number of EEG segments" [21]. Dividing the training and tests sets in this way makes it possible for different data segments from the same participant to be used for training and testing, allowing for higher accuracy owing to similarity between the data in the two sets. A similar case was described in another paper [5], where the authors aim to classify French children with and without autism based on their speech (N = 12). In that study, the corpus "treats sentences read by the same speaker as independent samples partitioned randomly in test, development, and training sets" and the classifier achieves 93.5% accuracy. The evaluation procedures used in these two studies report potentially overoptimistic results owing to the partitioning of the training and test sets. This partitioning does not account for real-world scenarios where the algorithm will have the task of correctly classifying a new person, portions of whose data have not been used for training.

This section presented related work on training classifiers distinguishing between people with and without autism. As evidenced by the studies described above, eye tracking has not been used to investigate whether it is suitable to detect autism by considering atypical visual attention patterns of people with autism. The next sections describe our approach to collecting the eye tracking data and training a classifier based on gaze data from web-related tasks.

3. OUR APPROACH

In this study we propose the use of eye-tracking data from web-related tasks for the classification of people with and without autism. Obtaining such data is significantly easier, cheaper and less-obtrusive compared to the recording of fMRI data. While still widely in the realm of research, eye tracking is already making its way into everyday use. For example, laptops and smartphones including eye-tracking navigation features have already been on the market for a few years (e.g. Samsung Galaxy S4[1], Tobii Eye Trackers for PC Gaming[2] and Acer Consumer Notebook with Tobii Eye Tracking[3], among others). Another part of our motivation is that if this preliminary study shows that attention differences between the two groups are large enough to produce a classifier with high level of accuracy, these differences could be captured using different methods that are not reliant on an eye tracker (e.g. logs of mouse clicks or other behaviour).

An important consideration in our study was to use an evaluation technique that emulates a real-world scenario where the model would have to diagnose a new person, i.e. no data from the participants in our test set has been used for training. To achieve this, we perform 100-fold cross-validation, where data from 10 randomly selected participants per group is used for training and the data from the remaining 5 participants per group is used for testing.

[1]BBC News. Samsung Galaxy S4 eye-tracking smartphone unveiled. [online] Last accessed: 8 January 2018. http://www.bbc.co.uk/news/technology-21791023

[2]https://venturebeat.com/2017/01/21/hands-on-with-tobiis-eye-tracking-laptop-controls/

[3]https://www.tobii.com/group/news-media/press-releases/2017/1/beyond-gaming-acer-launches-consumer-notebook-with-tobii-eye-tracking/

The tasks that our participants perform are highly familiar to them, as all of them are regular web users (Section 4). Using data from everyday tasks for autism screening has the potential to unveil differences between the two groups that could be captured by means which are independent of laboratory equipment and might highlight the importance of already existing data for the task of diagnostic screening. In that sense, owing to the information processing differences of people with autism, the Web might have the potential to reveal autistic traits at a much larger scale than smaller samples of data collected in an experimental setting.

Worldwide access to web pages in the native languages of many people is a lot more common than access to diagnostic tests for autism. This means that an approach based on screening for visual attention differences when using the Web could more easily scale up to larger groups of people from low and medium income countries compared to current diagnostic practices.

Finally, it is important to note that such as screening tool of the future is not meant to substitute clinical diagnostic procedures where these are available. Rather, its purpose is to be used to identify autistic traits at a wider level (e.g. to be offered as a screening tool in schools) and refer for further assessment those people who might be at risk. Another benefit of this approach is that it could help researchers get a better understanding of autism.

4. DATA COLLECTION

The data used in this study was initially collected for the purpose of investigating any potential web-searching difficulties that web users with autism might experience[4]. The results of that experiment are presented in both [13] and [14]. The preliminary analysis presented in [13] focused on the differences in the success scores of the two groups for the web searching tasks, showing that the group with autism was slightly less efficient in finding relevant information under limited time constrains. The further analysis presented in [14] showed that the participants with autism exhibited a tendency towards longer scan paths when searching for information and they tended to make more transitions between the elements of web pages [14]. The analysis also showed that they had a tendency to make shorter fixations, but more fixations on irrelevant elements in comparison with the control group [14]. The experiments presented in the current paper utilize some of these differences for the purpose of automatically distinguishing between the two groups.

The next paragraphs describe the tasks, materials, participants, apparatus and procedure used for the data collection.

Tasks: Participants from two groups (ASD and Control) were shown 6 web pages with varying visual complexity and their eye movements were recorded using an eye-tracking device. The participants were given two tasks: *browsing* and *searching* the web pages. The instruction for the browsing task was that the participants could spend up to two minutes per web page to look for any information that could be of interest to them. They were free to read or ignore certain parts of the web pages and process the visuals up to their preference and move on to the next page whenever they were ready. The search task was a lot more intense, as the

Figure 1: The Apple web page (low visual complexity)

participants had 30 seconds per web page to locate specific information in order to answer two questions per web page. The questions were asked verbally by the researcher (first author) and the participants were required to read aloud the answer. The full list of the questions is provided in Table 1.

Materials: We used the screen shots of six web pages that were previously used in [15]. The web pages had varying visual complexity, as measured by the ViCRAM tool [28]: Apple (Low), Babylon (Low), AVG (Medium), Yahoo (Medium), Godaddy (High) and BBC (High). The rationale behind using web pages with varying visual complexity is to see whether the level of complexity has an effect on the way the participants complete the tasks. Figures 1 and 2 give examples of two of the web pages with low and high visual complexity, respectively. The rest of the web pages can be found in our repository.

Participants: The eye tracking data was collected from 18 adult volunteers diagnosed with high-functioning autism or Asperger's syndrome (12 male and 6 female) and 18 non-autistic control participants (10 male and 8 female). Subsequently, data for 3 participants from the ASD group and 2 participants from the control group was discarded due to difficulties with calibration and missing data because of head movements. Furthermore, in order to have balanced classes, we excluded the data of 1 random participant from the control group[5]. As a result, the data included in this analysis was obtained from 15 participants with autism (9 male and 6 female) and 15 control neurotypical participants (8 male and 7 female). The frequency of using the web for all participants was assessed through the question *"How often do you browse the web?"* with five possible answers: *Daily, Weekly, Monthly, Less than once a month, Never.* All 30 participants selected the option *"Daily"*.

All participants with autism were recruited through a charity organisation based in the West Midlands area in the UK.

[4]Full ethical approval was obtained by the University of Wolverhampton Faculty of Arts Ethics Committee prior to collecting the data

[5]For this reason the number of participants whose data was retained for analysis in this study was different from the one reported in [13]

Table 1: The searching tasks used in the eye tracking study

Page	Tasks
Apple	(a) Can you locate the link that allows to watch the TV ads relating to iPad mini? (b) Can you locate a link labelled iPad on the main menu?
Babylon	(a) Can you locate the link that you can download the free version of Babylon? (b) Can you find and read the names of other products of Babylon?
AVG	(a) Can you locate the link which you can download the free trial of AVG Internet Security 2013? (b) Can you locate the link which allows you to download AVG Antivirus Free 2013?
Yahoo!	(a) Can you read the titles of the main headlines which have smaller images? (b) Can you read the first item under the News title?
Godaddy	(a) Can you find a telephone number for technical support and read it? (b) Can you locate the text box where you can search for a new domain?
BBC	(a) Can you read the first item of Sport News? (b) Can you locate the table that shows market data under the Business title?

Figure 2: The BBC web page (high visual complexity)

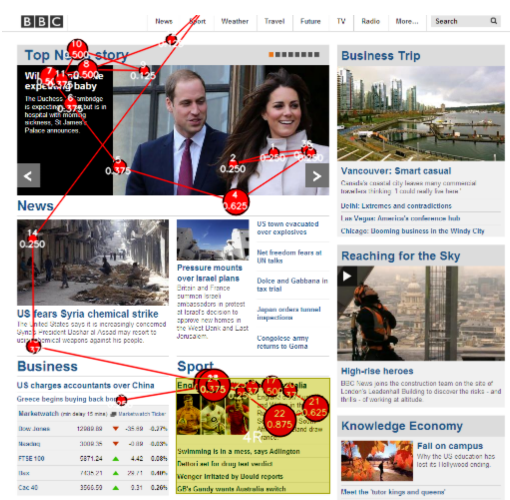

Figure 3: The scan path of an ASD participant answering the question "Can you read the first item of Sport News" for the BBC page. The answer is contained in the green area.

They all had a confirmed clinical diagnosis of autism, as verified by the charity organisation. Exclusion criteria was any presence of formally diagnosed intellectual disability, being under the age of 18 and not being able to use a computer. The mean age in years for the ASD group was $\mu = 37$, $SD = 9.14$ and mean number of years spent in formal education was $\mu = 16$ with $SD = 3.33$

The participants without autism were recruited through open advertisement across the West Midlands, UK. In order to rule out the presence of autism in the control group participants, all of them were asked to fill in an Autism Quotient (AQ) screening test [6] to make sure that none of them had a high incidence of autistic traits without having received a formal diagnosis. This test is widely used by general practitioners in the UK as an initial screening before a formal diagnostic referral. None of the control participants included in the study had a AQ score higher than 32, which meant that they all had "little or no autistic traits". The mean age of the control group was $\mu = 33.66$, $SD = 8.6$ and mean number of years spent in formal education was $\mu = 18.35$ with $SD = 2.47$.

Apparatus: The device used for recording the gaze of the participants during task performance was a Gazepoint GP3 video-based eye tracker[6] (60Hz sampling rate and accuracy of 0.5-1 degree of visual angle). The screen shots of the web pages were presented on a 19" LCD monitor. The distance between each participant and the eye tracker was controlled by using a sensor integrated within the Gazepoint software, and was roughly 65 cm.

Procedure: After getting familiar with the purpose and procedure of the experiment, all the participants signed a consent form. The participants from the control group were first asked to fill in the Autism Quotient test in order to rule out control-group participants with high level of autistic traits. The demographic data about age, gender and diagnosis was collected and a nine-point calibration of the eye tracker was performed. After the successful calibration, the participants were presented with the six web pages in a randomised order to deal with the memory effect (specifically, the pages were randomised for each participant). The two tasks (browsing and searching) were presented in counterbalanced order for each participant and web page. After the completion of the eye tracking part, the participants

[6]https://www.gazept.com/

were also asked to fill in a short survey including questions about how often they use the web and how often they have visited the six websites where the web pages were from (1: Daily, 2: Weekly, 3: Monthly, 4: Less than once a month, 5: Never). The latter was done to control for the potential situation where a participant was overly familiar with the stimuli. The results indicated that none of them was. After the completion of the experiment, the participants were debriefed.

5. CLASSIFICATION EXPERIMENTS

This section describes the experimental setting, features and parameters used for training and evaluating of the the autism screening classifier.

5.1 Defining the Areas of Interest

The areas of interest (AOI) are visual elements or regions in which the raw eye tracking data is analysed. Inspired by [18], we define AOIs using different AOI identification approaches, calculate statistics on basic gaze measures restricted to these AOIs, and explore the effects of the different identification approaches on the prediction accuracy. To test how much knowledge is necessary for the accurate classification of our participants, we process the raw gaze data in two different AOI set ups: page-specific AOIs and generic AOIs, which are shown in Figures 4 and 5, respectively. The page-specific AOIs were defined by [15] based on the extended and improved version of the Vision-Based Page Segmentation (VIPS) algorithm, which segments web pages by using their source code and visual representations based on different granularity levels [1] (Figure 4). The generic AOIs consisted of a 2 x 2 grid as shown in Figure 5.

Figure 4: Page-specific Areas of Interest for the Apple web page

5.2 Features

A number of gaze and non-gaze features were used to train the classifiers. The features and their definitions are presented below.

Gaze features

- **Time to First View:** The time in seconds when an AOI was first fixated measured from the moment the web page was presented on screen.
- **Time Viewed (sec):** The sum of the total durations of all fixations per participant per AOI measured in seconds.

Figure 5: Generic Areas of Interest for the Apple web page

- **Time Viewed %:** The sum of the total durations of all fixations per participant per AOI as percentage of the total time spent looking at the web page for the specific task.
- **Fixations:** The number of gaze fixations for each participant in a given AOI.
- **Revisits:** The number of times a participant went back to a previously viewed AOI.

Non-gaze features

- **Media ID:** The identification number of each web page.
- **AOI ID:** The identification number of each area of interest.
- **Correct Answer AOI:** the area of interest which contains the correct answer to one of the search tasks. The rationale behind including this information as a feature is that it would reflect the time each participants has spent recognising the correct answer.
- **Participant Gender:** Previous research has shown that gender may have an impact on the way people look for information on the Web [33, 26], which is why we included it as a variable in our study.
- **Level of Visual Complexity:** As explained in the description of the materials, the web pages had three levels of visual complexity, low, medium and high, which were also included as features.

5.3 Experimental Setting

We performed several sets of classification experiments depending on the task (searching vs. browsing) and the AOI identification approaches (page-specific vs. generic). After several classification algorithms were tested, best performance was achieved by the logistic regression algorithm in R. We performed 100-fold cross validation where data from 10 randomly selected participants per group was used for training and the data from the remaining 5 participants per

Feature set	Pages	Result
Gaze features	all	.66 (.64 - .68)
Gaze + AOI ID	all	.64 (.62 - .66)
Gaze + Media ID	all	.65 (.63 - .68)
Gaze + selected media	Apple + AVG	**.71 (.69 - .73)**

Table 2: Results for all web pages for the Browse task with 95% confidence intervals (page-specific AOIs)

Feature Set	Apple	Babylon	AVG	BBC	GoDaddy	Yahoo!
Gaze features	.63 (.61-.66)	.54 (.52-.56)	.59 (.57-.62)	.45 (.42-.47)	.55 (.53-.58)	.52 (0.50-.55)
Gaze features + AOI ID	.47 (.46-.48)	.5 (.48-.52)	.57 (.54-.60)	.49 (.48-.50)	.47 (.46-.48)	.52 (.50-.53)

Table 3: Results for individual web pages for the Browse task with 95% confidence intervals (page-specific AOIs)

	Control	ASD
Control	0.722	0.294
ASD	0.278	0.706

Table 4: Confusion matrix for the Browse task best result: All gaze features + Selected media (Apple + AVG). The number is calculated through dividing 500 (100 test trials x 5 participants per group) by the number of correctly labeled participants for each class.

group was used for testing. The model is first trained to predict which group a behaviour would belong to. To predict which group a participant belongs to, we compare the number of times a behaviour from that participant is classified as belonging to each group and consider the group with the higher number as the predicted group. The accuracy of each fold is then averaged and reported together with the 95% confidence interval (CI).

6. RESULTS

This section present the results of the classification experiments and the effect that different variables had on the prediction accuracy. We also report the 95% confidence intervals for each classifier, as well as the confusion matrices for the two best results.

6.1 Classification performance for the Browse task

We trained various classifiers using combinations of feature sets. We used all gaze features together with non-gaze ones for i) all pages and ii) selected combinations of pages. The classification performance results for the Browse task are presented in Table 2. In addition, Table 3 presents the classification performance of the classifiers employing different combinations of features for each individual web page. These results refer to the data for the page-specific AOIs.

As can be seen from the tables, best performance for the Browse task was achieved when using all gaze features for a combination of the two most predictive web pages for this task: Apple and AVG. The classification accuracy for those was 0.71 with 95% CI (0.69;0.73). The confusion matrix presented in Table 4 indicates that both classes (ASD and Control) are predicted with a similar level of accuracy. The Apple and AVG web pages were the ones with largest between-group differences for the Browse task, which resulted in them having a highest classification performance. The result was similar for other combinations including these pages (e.g. 0.702 for Apple + AVG + Yahoo!). When not all of these pages were included, the best performance of 0.662 was achieved by combining AVG + GoDaddy + BBC. The results for all the rest of the web page combinations can be found in our repository.

The analysis of the results for the individual web pages (Table 3) shows that the level of visual complexity of the web pages does not seem to have an effect on the classification performance. The case was similar with the participant gender variable which did not have a significant positive or negative influence on the results.

The results from classifiers using the data from the 2 x 2 grid AOIs were comparable to the best results for the page-specific AOIs with respect to the Browsing task, with an accuracy of 0.615 for all media and 0.7 for selected media.

6.2 Classification performance for the Search task

Similar to the Browse task, for the Search task we also trained classifiers using different features and different combinations of web pages. The results presented in Table 5 show that the data obtained through the Search task is more suitable for classifying people with autism, since the best classifier achieved an accuracy of 0.75 with 95% CI (0.73; 0.78). While outperforming the classifiers trained on the data from the Browse task, this result confirms that, just like with the Browse task, best performance is achieved when using gaze data for selected web pages (in this case Apple + Babylon, as they had largest between-group differences for the Search task). The confusion matrix presented in Table 7 again indicates that both classes (ASD and Control) are predicted with a similar level of accuracy, where the ASD class is predicted slightly better than the Control one (0.808 and 0.698, respectively). The result was similar for other combinations including these pages (e.g. 0.754 for Apple + Babylon + AVG or 0.756 for Apple + Babylon + GoDaddy). When none of these pages were included, the best performance of 0.611 was achieved by combining AVG + Yahoo! + GoDaddy. The results for all the rest of the web page combinations can be found in our repository.

Feature set	Pages	Result
Gaze features	all	.70 (.68 - .72)
Gaze + AOI ID	all	.58 (.56 - .60)
Gaze + Correct Ans. AOI	all	.69 (.67 - .71)
Gaze + Media ID	all	.66 (.63 - .68)
Gaze + selected media	Apple + Babylon	**.75 (.73 - .78)**
Gaze + Correct Ans. AOI + selected media	Apple + Babylon	.73 (.70 - .75)

Table 5: Results for all web pages for the Search task with 95% confidence intervals (page-specific AOIs)

Feature Set	Apple	Babylon	AVG	BBC	GoDaddy	Yahoo!
Gaze features	.69 (.67-.72)	.63 (.60-.65)	.39 (.37-.42)	.57 (.55-.59)	.60 (.58-.63)	.48 (.46-.50)
Gaze features + AOI ID	.51 (.50-.52)	.49 (.48-.50)	.48 (.47-.50)	.50 (.50-.50)	.48 (.47-.49)	.50 (.49-.52)
Gaze + Correct Ans. AOI	.69 (.67-.72)	.61 (.58-.63)	.43 (.41-.46)	.55 (.52-.57)	.58 (.55-.60)	.47 (.46-.49)

Table 6: Results for individual web pages for the Search task with 95% confidence intervals (page-specific AOIs)

	Control	ASD
Control	0.698	0.192
ASD	0.302	0.808

Table 7: Confusion matrix for the Search task best result: All gaze features + Selected media (Apple + Babylon). The number is calculated through dividing 500 (100 test trials x 5 participants per group) by the number of correctly labeled participants for each class.

Similar to the Browse task, the analysis of the results for the individual web pages (Table 7) shows that the level of visual complexity of the web pages does not seem to have an effect on the classification performance and nor does the participant gender. However, unlike the Browse task, the classifiers trained on the data from the 2 x 2 grid for the Search task were significantly less successful than the ones trained on the page-specific data, with accuracy of 0.57 for all media and 0.56 for selected media.

7. DISCUSSION

As can be seen from the results, the best performance achieved for the Search task was 0.75 accuracy and 0.71 for the Browse task, when training on selected media. The 95% confidence intervals show that the models are robust in their prediction. Examination of the confusion matrices indicates that both classes (ASD and Control) are predicted with a similar level of accuracy. These results indicate that visual attention differences captured through gaze data from web-related tasks are a promising direction in autism screening.

Task Effects The type of the task that the participants performed had an effect on the prediction accuracy: the models trained on the data from the Search task were more successful than the ones trained on the data from the Browse task. A possible explanation for this effect could be that the Search task revealed larger differences in the visual attention patterns of the two groups. However, the free scanning of the web pages also shows that the participants with and without autism spread their attention differently when presented with stimuli such as web pages. Further analysis is needed to identify whether this difference was in terms of text versus images, as previously suggested by [39], or with respect to other phenomena. Another question to investigate further is why the results were better for particular web pages, where potential reasons could be related to granularity of the segmentation, task complexity or due to some design elements in the web page. In any case, both tasks showed that visual processing differences could be successfully used in classification experiments even in a scenario like this one, where the primary objective behind collecting the data was related to a different task on improving web accessibility for people with autism. A question of paramount importance for future research in this direction would be to design tasks which amplify the visual processing differences between the two groups and thus enhance the classification accuracy.

AOI Identification Effects It is also interesting to note the effects that the different AOI identification approaches had on the results. The 2 x 2 grid resulted in a fairly similar classification accuracy when compared to the page-specific set up for the Browse condition (0.7 and 0.71, respectively). By contrast, it resulted in a much lower accuracy for the Search condition when compared to the page-specific set up (0.56 and 0.75, respectively). This indicates that extracting gaze data from generic AOIs is suitable for tasks which are not related to looking at particular areas of the screen and tasks where all regions of the screen have similar importance. However, the Search condition required the participants to identify specific information and thus certain elements of the page containing this information were more important than others and consistently featured longer fixations and a higher number of fixations. Hence, for this condition it was important that the AOIs are defined following a meaningful pattern, in this case the elements of the page. These results show that the AOIs should not be defined without taking into consideration the specificities of the task.

Effects of Other Variables Variables such as participant gender and the visual complexity of the web pages did not seem to influence the results. As expected, the variable related to whether or not an AOI contained the correct answer to a search task had a slightly positive effect on the Search task classifiers (0.69). The Media ID variable behaved slightly differently depending on the task but in general lowered the accuracy compared to models trained only on the gaze data.

Limitations While this study showed the potential of using gaze data from web searching tasks for autism screening, it has several limitations, which need to be addressed in future work. First and foremost, our best accuracy of 0.75 compares to accuracy in the range of 71% to 86.7% for the various classifiers trained on fMRI data presented in Section 2.2. One way to close this gap in future experiments would be to design such visual tasks that would result in larger between-group differences. As mentioned previously, the data used in this study was not collected with this objective in mind. It is also worth noting that the presented experiment resulted in a more balanced prediction for the two classes compared to the one achieving 86.7% accuracy [7]. Another limitation of our study is that web searching tasks are not suitable for very young children where the early detection of autism would be most helpful. In order to include very young children as a potential group of beneficiaries, visual attention differences, if existent, will have to be captured through a different task. Last but not least, our participant sample is not very large, however, it is, in fact larger than the benchmark study presented in [7], which was conducted with a sample of 12 participants per group.

Impact and Future Work In spite of its limitations, this study revealed the promising potential of visual searching tasks for differentiation between people with and without ASD. In particular, the vast amount of behavioral data obtained from an everyday task such as using the Web, could be the next stepping stone to autism screening at a larger scale. To achieve this, our future work has two main objectives: i) to design tasks resulting in larger between-group differences and improved classification accuracy, and ii) to explore ways to capture these differences through means which do not require eye tracking (e.g. operation logs, mouse clicks, accuracy, miss rate, etc). The results and insight obtained from such experiments could then be used to develop a serious game for autism screening at a large scale.

The significance of this study also lies in the idea that gaze data could be used to train classifiers and detect clinical conditions which exhibit differences in visual attention. This information opens the door to the detection of other disorders or diseases such as dementia, schizophrenia, ADHD, etc. by using similar means. Finally, it is important to note that while this work is still in its infancy, potential future work towards using web searching behaviour for the profiling for certain conditions will raise ethical considerations which should not be treated lightly.

8. CONCLUSION

This paper presented a set of experiments where gaze data from two web-related tasks (Browse and Search) was used to train classifiers to distinguish between people with and without autism. The results showed that the Search task elicited larger between-group differences and hence, the classifier trained on this data achieved 0.75 accuracy compared to 0.71 for the Browse task. We also explored the effects of different approaches to defining the areas of interest by extracting the data from two different conditions: generic AOIs (2 x 2 grid) and page-specific AOIs (the elements of the web page). According to the results,the use of generic AOIs is suitable for tasks such as the Browse one, where all elements of the screen are equal. For tasks such as the Search one, much better accuracy was achieved when defining the areas in a meaningful way and taking into consideration the task being performed. Other variables such as participant gender, the visual complexity of the web pages or the positioning of the right answer to the Search tasks did not have a significant influence on the classification performance. To the best of our knowledge, this is the first study to use gaze data for the detection of autism and it shows that: i) visual attention could potentially be used as a marker of autism, ii) web page processing tasks are a good stimulus set, and that iii) performance on such tasks could be used to develop an affordable and accessible serious game for the detection of autism at a large scale.

9. REFERENCES

[1] M. E. Akpınar and Y. Yeşilada. Vision Based Page Segmentation Algorithm: Extended and Perceived Success. In Q. Z. Sheng and J. Kjeldskov, editors, *Current Trends in Web Engineering*, volume 8295 of *LNCS*, pages 238–252. Springer, 2013.

[2] Y. M. Al-Farsi, M. M. Al-Sharbati, O. A. Al-Farsi, M. S. Al-Shafaee, D. R. Brooks, and M. I. Waly. Brief report: Prevalence of autistic spectrum disorders in the sultanate of oman. *Journal of Autism and Developmental Disorders*, 41(6):821–825, 2011.

[3] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th ed.), 2013.

[4] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. DuBray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, P. T. Fletcher, et al. Functional connectivity magnetic resonance imaging classification of autism. *Brain*, 134(12):3742–3754, 2011.

[5] M. Asgari, A. Bayestehtashk, and I. Shafran. Robust and accurate features for detecting and diagnosing autism spectrum disorders. In *Interspeech*, pages 191–194, 2013.

[6] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley. The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1):5–17, 2001.

[7] A. Bernas, A. P. Aldenkamp, and S. Zinger. Wavelet coherence-based classifier: a resting-state functional mri study on neurodynamics in adolescents with high-functioning autism. *Computer methods and programs in biomedicine*, 154:143–151, 2018.

[8] L. Billeci, A. Narzisi, A. Tonacci, B. Sbriscia-Fioretti, L. Serasini, F. Fulceri, F. Apicella, F. Sicca, S. Calderoni, and F. Muratori. An integrated eeg and eye-tracking approach for the study of responding and initiating joint attention in autism spectrum disorders. *Scientific Reports*, 7(1):13560, 2017.

[9] A. V. Buescher, Z. Cidav, M. Knapp, and D. S. Mandell. Costs of autism spectrum disorders in the united kingdom and the united states. *JAMA pediatrics*, 168(8):721–728, 2014.

[10] M. Davidovitch, N. Levit-Binnun, D. Golan, and P. Manning-Courtney. Late diagnosis of autism spectrum disorder after initial negative assessment by

a multidisciplinary team. *Journal of Developmental & Behavioral Pediatrics*, 36(4):227–234, 2015.

[11] A. T. Duchowski. Eye tracking methodology. *Theory and practice*, 328, 2007.

[12] M. S. Durkin, M. Elsabbagh, J. Barbaro, M. Gladstone, F. Happe, R. A. Hoekstra, L.-C. Lee, A. Rattazzi, J. Stapel-Wax, W. L. Stone, et al. Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. *Autism Research*, 8(5):473–476, 2015.

[13] S. Eraslan, V. Yaneva, Y. Yesilada, and S. Harper. Do web users with autism experience barriers when searching for information within web pages? In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work*, page 20. ACM, 2017.

[14] S. Eraslan, V. Yaneva, Y. Yesilada, and S. Harper. Web users with autism: Eye tracking evidence for differences. *Behaviour & Information Technology*, In Submission.

[15] S. Eraslan and Y. Yesilada. Patterns in eyetracking scanpaths and the affecting factors. *Journal of Web Engineering*, 14(5-6):363–385, 2015.

[16] T. Falkmer, K. Anderson, M. Falkmer, and C. Horlin. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *European child & adolescent psychiatry*, 22(6):329–340, 2013.

[17] U. Frith. *Autism. Explaining the enigma*. Blackwell Publishing, Oxford, UK, second edition, 2003.

[18] M. Gingerich and C. Conati. Constructing models of user and task characteristics from eye gaze data for user-adaptive information highlighting. In *AAAI*, pages 1728–1734, 2015.

[19] Happe, F., and Frith, U. he weak coherence account: Detail focused cognitive style in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 36:5–25, 2006.

[20] C. Horlin, M. Falkmer, R. Parsons, M. A. Albrecht, and T. Falkmer. The cost of autism spectrum disorders. *PloS one*, 9(9):e106552, 2014.

[21] S. Ibrahim, R. Djemal, and A. Alsuwailem. Electroencephalography (eeg) signal processing for epilepsy and autism spectrum disorder diagnosis. *Biocybernetics and Biomedical Engineering*, 38(1):16–26, 2018.

[22] E. B. Itzchak and D. A. Zachor. Who benefits from early intervention in autism spectrum disorders? *Research in Autism Spectrum Disorders*, 5(1):345–350, 2011.

[23] R. M. Jones and C. Lord. Diagnosing autism in neurobiological research studies. *Behavioural brain research*, 251:113–124, 2013.

[24] M. A. Just and P. A. Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4):329, 1980.

[25] C. Laidi, J. Boisgontier, M. M. Chakravarty, S. Hotier, M.-A. d'Albis, J.-F. Mangin, G. A. Devenyi, R. Delorme, F. Bolognani, C. Czech, et al. Cerebellar anatomical alterations and attention to eyes in autism. *Scientific reports*, 7(1):12008, 2017.

[26] A. Large, J. Beheshti, and T. Rahman. Gender differences in collaborative web searching behavior: an elementary school study. *Information Processing & Management*, 38(3):427–443, 2002.

[27] Lovaas, O. I. and Schreibman, L. Stimulus Overselectivity of Autistic Children in a Two Stimulus Situation. *Behavior Research and Therapy*, 9:305–310, 1971.

[28] E. Michailidou. *Visual complexity rankings and accessibility metrics*. PhD thesis, The University of Manchester, 2010.

[29] C. M. Murphy, C. E. Wilson, D. M. Robertson, C. Ecker, E. M. Daly, N. Hammond, A. Galanopoulos, I. Dud, D. G. Murphy, and G. M. McAlonan. Autism spectrum disorder in adults: diagnosis, management, and health services development. *Neuropsychiatric disease and treatment*, 12:1669, 2016.

[30] D. G. Murphy, J. Beecham, M. Craig, and C. Ecker. Autism in adults. new biologicial findings and their translational implications to the cost of clinical services. *Brain research*, 1380:22–33, 2011.

[31] M. Plitt, K. A. Barnes, and A. Martin. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage: Clinical*, 7:359–366, 2015.

[32] B. O. Ploog. Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of autism and developmental disorders*, 40(11):1332–1349, 2010.

[33] M. Roy and M. T. Chi. Gender differences in patterns of searching the web. *Journal of educational computing research*, 29(3):335–348, 2003.

[34] S. A. Samadi and R. McConkey. Autism in developing countries: Lessons from iran. *Autism research and treatment*, 2011, 2011.

[35] S. Štajner, V. Yaneva, R. Mitkov, and S. P. Ponzetto. Effects of lexical properties on viewing time per word in autistic and neurotypical readers. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 271–281, 2017.

[36] L. Q. Uddin, K. Supekar, C. J. Lynch, A. Khouzam, J. Phillips, C. Feinstein, S. Ryali, and V. Menon. Salience network–based classification and prediction of symptom severity in children with autism. *JAMA psychiatry*, 70(8):869–879, 2013.

[37] Z. Warren, M. L. McPheeters, N. Sathe, J. H. Foss-Feig, A. Glasser, and J. Veenstra-VanderWeele. A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*, 127(5):e1303–e1311, 2011.

[38] V. Yaneva, I. Temnikova, and R. Mitkov. Accessible texts for autism: An eye-tracking study. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 49–57. ACM, 2015.

[39] V. Yaneva, I. Temnikova, and R. Mitkov. Accessible texts for autism: An eye-tracking study. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '15, pages 49–57, New York, NY, USA, 2015. ACM.