Autism Detection Based on Eye Movement Sequences on the Web: A Scanpath Trend Analysis Approach

Sukru Eraslan
Department of Computer Science, University of Manchester, Manchester, United Kingdom
sukru.eraslan@manchester.ac.uk

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

Victoria Yaneva
Research Group in Computational Linguistics, University of Wolverhampton, Wolverhampton, United Kingdom
v.yaneva@wlv.ac.uk

Simon Harper
Department of Computer Science, University of Manchester, Manchester, United Kingdom
simon.harper@manchester.ac.uk

ABSTRACT

Autism diagnostic procedure is a subjective, challenging and expensive procedure and relies on behavioral, historical and parental report information. In our previous, we proposed a machine learning classifier to be used as a potential screening tool or used in conjunction with other diagnostic methods, thus aiding established diagnostic methods. The classifier uses eye movements of people on web pages but it only considers non-sequential data. It achieves the best accuracy by combining data from several web pages and it has varying levels of accuracy on different web pages. In this present paper, we investigate whether it is possible to detect autism based on eye-movement sequences and achieve stable accuracy across different web pages to be not dependent on specific web pages. We used Scanpath Trend Analysis (STA) which is designed for identifying a trending path of a group of users on a web page based on their eye movements. We first identify trending paths of people with autism and neurotypical people. To detect whether or not a person has autism, we calculate the similarity of his/her path to the trending paths of people with autism and neurotypical people. If the path is more similar to the trending path of neurotypical people, we classify the person as a neurotypical person. Otherwise, we classify her/him as a person with autism. We systematically evaluate our approach with an eye-tracking dataset of 15 verbal and highly-independent people with autism and 15 neurotypical people on six web pages. Our evaluation shows that the STA approach performs better on individual web pages and provides more stable accuracy across different pages.

W4A '20 April 20-21, 2020, Taipei, Taiwan

ISBN 978-1-4503-7056-1/20/04...$15.00

DOI: 10.1145/3371300.3383340

CCS Concepts

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

Keywords

Eye tracking, Scanpath, Autism, Classification, STA, Web Pages

OPEN DATA

All the individual paths used for the evaluation of our proposed approach are available in our external repository at Zenodo [20]. The repository also includes the Python code to re-run the evaluation. Therefore, the proposed approach can be re-evaluated by other researchers with different individual paths in the future.

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by differences in communication and social interaction [3]. Obtaining an autism diagnosis is an elaborate and expensive procedure, which is also highly subjective as it entirely relies on behavioral, historical and parental report information [8, 22]. The lack of objective markers makes it especially difficult to diagnose cases of highly-able and independent adults with autism, since their symptoms may be less salient and may be further obscured by a variety of coping strategies developed in time. As a result, people with autism, and especially adults with autism, are at high risk of not receiving a diagnosis or receiving one late in life. Similar to other conditions, having a correct autism diagnosis is crucial for receiving formal support and treatment, as well as for better self-awareness, coping with the condition, and finding a peer-support community.

In addition to being subjective and challenging, the autism diagnostic procedure is also expensive. Raising awareness of the condition in recent years has strongly increased the demand for diagnostic services, but it has also significantly slowed down the process and led to an increased financial burden to families [29, 11]. In the United Kingdom alone, there has been a significant increase in the demand for autism diagnostics but usually less than half of the applicants are found to meet the criteria [34]. Importantly, the difficulties reported so far refer to the diagnosing of autism in high-income countries. People in medium- and low-income countries have even more limited access to autism screening, diagnosis, and treatment [12, 2, 40] owing to the high cost of tools and training professionals and para-professionals to use them [12]. The lack of resources for recognising and supporting the condition is one of the reasons why low autism prevalence is mostly associated with low-income countries.

Based on the information reported so far, the diagnostic procedure for autism requires significant improvements in terms of its reliability and cost. These issues can be alleviated if the procedure is aided by (or preceded by) screening tools that rely on observable and measurable behaviours, with minimal associated cost. Ideally, those tools would also be unobtrusive and rely on behaviours from highly-familiar everyday tasks as opposed to abstract tasks or self-report.

To address these challenges, in our previous work [48], we proposed an objective model that helps to detect autism based on differences in visual attention. This was done by recording the eye movements of people while searching for information on web pages and then using this behavioural data as input for a classification algorithm. This approach was inspired by previous studies which provide eye-tracking evidence that people with autism tend to follow different strategies on the web when they complete their tasks in comparison with neurotypical people [13, 14]. These studies have shown that people with autism tend to produce more fixations, more transitions between web-page elements and more frequent fixations on the elements that are not relevant to a given task but the duration of the produced fixations is comparatively shorter [14]. Our autism detection classifier achieved 75% accuracy as the best accuracy when combining data from several web pages. This led to an interesting observation: across participants, data from certain web pages and tasks were more informative than data from other pages and tasks, however, no clear pattern emerged as to what makes a web page or a task "suitable". Since the suitability of web pages and tasks (and other visual stimuli) to elicit discriminative attention-related behaviours currently cannot be known a priori, there is a need to explore approaches that may provide stable results across pages, thus reducing the number of tasks and stimuli that a participant needs to attend to receive a reliable screening result.

In this present paper, we test whether using eye-movement sequences improves the stability of autism detection, as the differences found in the eye-movement data of people with autism and neurotypical people (especially, more fixations, more transitions and more frequent fixations on irrelevant elements) potentially affect their eye-movement sequences [14]. In our previous work [13, 14], we observe that there is a difference in trending sequences between neurotypical and people with autism. Therefore, here, we hypothesise that the analysis of sequential data may provide predictions that are more robust compared to using non-sequential data as input, since such an approach takes into consideration the order of behavioural events. We use the Scanpath Trend Analysis (STA) algorithm [18, 19] to identify the trending paths of people with autism and neurotypical people and classify a person as a person with autism or neurotypical person based on the similarity of his/her path to the trending paths. The STA algorithm provides the most representative path of multiple individual paths on a web page compared to other existing algorithms [18]. It has been used for different purposes including experiential transcoding of web pages [25], the detection of common code reading patterns [42] and the generation of a feature for a classifier that correlates cognitive characteristics with interaction and visual behaviour patterns [37]. We evaluate our approach with an eye-tracking dataset of 15 people with autism and 15 neurotypical on six web pages where the data of 10 people from each group were used for the training purposes and the rest of the data was used for the testing purposes. We present an experiment that is repeated 100 times on each web page and we report the mean of precision, recall, F1-measure and accuracy.

1.1 Contributions

The contributions of this paper are as follows:

- We present a novel approach to detect autism utilising the signal from the sequences in which subjects process web-page element. This approach can be used as a potential screening tool or used in conjunction with diagnostic methods to help identify people who are at risk and thus aid established diagnostic methods. The experiments produced evidence that the sequence in which web users with and without autism approach web-page elements is sufficiently different to allow user classification with accuracy higher than chance[1].

- To evaluate the gain obtained from the sequential analysis of the eye movements, we directly compared the proposed approach with the most relevant non-sequential approach for eye-tracking data based autism detection using the same data [48]. We show that sequential data improves the stability of predictions, which is an important consideration for the practical use of screening tools; furthermore, the observation can impact the development of other eye-tracking data based tools for the detection of attention-related conditions (e.g. dyslexia).

- The code and generated sequences are made freely available for replication purposes.

2. RELATED WORK

In recent years, the accessibility community has made substantial progress in going beyond report-based assessment and in using machine learning to detect and measure a number of conditions such as dyslexia [39], level of physical capability [36], speech capability loss [35], and even emotional arousal in children with neurodevelopmental disorders [9]. Machine-learning models for autism detection using behavioural data are still not widely explored, but several of them have been trained using data types such as functional Magnetic Resonance Imaging (fMRI), electroencephalograms (EEG), speech, and gaze data obtained from children.

The accuracy of the best fMRI models varies between 79% [4] and 86% [8] with leave-one-out cross-validation (LOOCV) and between 71% [4] and 83% [44] with unseen data validation. Although these models achieve promising accuracy for autism detection, the collection of fMRI data needs a very expensive and obtrusive procedure, especially for people with sensory issues. The data collection procedure may also not be suitable for people with metal implants, claustrophobia, head trauma, etc. These circumstances limit the applicability of using fMRI data for autism detection. There have also been machine learning approaches reporting 94% accuracy with EEG data [30] and 93% accuracy with speech data [6].

[1] The binary classification (autism/non-autism) reflects the current diagnostic criteria introduced in 2013 (DSM-5), which no longer defines sub-levels of autism because of the innate heterogeneity of the condition.

These results are likely over-optimistic, as data from the same participants was split into different data segments, some of which used for training and others used for testing. This artificially increased the similarity between training and test sets, resulting in very high accuracy (in reality, no portions of one's data would be present as labeled instances in a test set). Further discussion of these machine learning approaches can be found in [48].

While currently not a formal diagnostic criterion, atypical attention patterns in people with autism are a well-known phenomenon [10, 31, 49, 28, 21, 50, 46]. The theoretical underpinnings of the differences in visual attention can be traced back to atypical information processing where "the ASD cognitive profile is biased towards local sensory information with less account for global, contextual and semantic information" [24]. Indeed, in many eye-tracking studies, participants with autism are reported to focus on specific areas as opposed to exploring a larger part of the visual scene as their neurotypical counterparts do [28, 27, 10]. Therefore, differences in visual attention are indicative of higher-order information-processing differences between people with and without autism and can thus be used as a suitable proxy to distinguishing between cognitive profiles.

Eye tracking allows observing what areas of a visual scene people fixate their gaze on, for how long, and in what order. To illustrate the nature of eye-tracking data, Figure 1 shows a gaze plot visualization of a scanpath of a single user on a web page [43]. In a gaze plot visualization, fixations are represented as circles, where the size of the circle is positively correlated with the duration of the fixation.

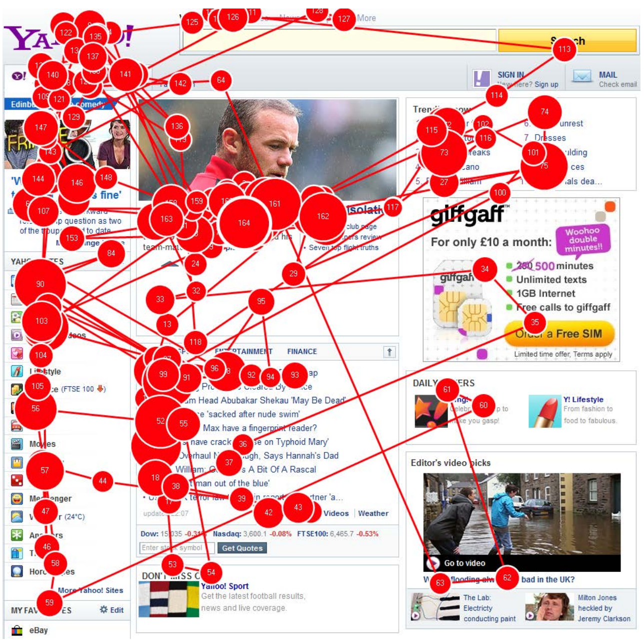

Figure 1: A gaze plot visualisation of a scanpath of a single user on a web page

Several recent models aim to detect autism based on eye-tracking data, almost exclusively focused on toddlers and young children [28, 27, 45, 32]. Many of their contributions lie in the attempt to reliably record quality data with such young subjects and the reported accuracy is between 85% and 88%. The stimuli used in these experiments include videos and faces and the level of autism severity has not been reported.

As mentioned in the Introduction section, in our previous work, we proposed a machine learning model based on eye-tracking data for autism detection in highly independent adults [48]. We first created a feature set with non-sequential eye-tracking data features for each element for each participant on each web page (time to the first fixation, time viewed, time viewed %, fixation count and revisit count) with other features (page identifier, element identifier, visual complexity of the page, participant gender, target element (yes/no)). We then trained several machine learning classifiers and achieved 75% accuracy as the best accuracy with the logistic regression approach by combining data from several web pages. However, in our previous work, when we look at the accuracy on individual pages, we can see that the results were not consistent and affected by the underlying page. The accuracy ranged from 45-63% and 39-69% on individual pages for the browsing and searching tasks respectively therefore the standard deviation between pages was high. The current paper is a continuation of our previous work and proposes the analysis of eye-movement sequences as a way to provide a more stable prediction across different web pages by capturing differences caused by eye-movement data [14].

Different kinds of algorithms are available to analyse eye-movement sequences of users on a web page to detect sequential patterns in a group of eye-movement sequences (such as eyePatterns [47] and SPAM [26]) or identify a representative sequence for a group of users (such as eMine [15, 16]). A detailed overview of these algorithms can be found in [17]. In this study, we use the STA algorithm to identify the trending paths of people with autism and neurotypical people to be used as a basis for autism detection because (i) this algorithm was found to be the most successful algorithm in identifying the most similar path to the individual paths in comparison to other algorithms [18] and (ii) we observe a difference between the trending paths of people with autism and neurotypical people in our previous work [13, 14].

3. PROPOSED APPROACH

We propose to generate the trending paths for people with autism and neurotypical people by using Scanpath Trend Analysis, and then classify a person as a person with autism or neurotypical person based on his/her path's similarity to the trending paths. An overview of this approach is illustrated in Figure 2.

3.1 STA: Scanpath Trend Analysis

Scanpath Trend Analysis (STA) is designed to summarise a group of eye-movement paths into a single representative path and it is comprised of three core stages: (1) Preliminary Stage, (2) First Pass, and (3) Second Pass. The full description of the STA algorithm can be found in [18, 19].

Preliminary Stage: A series of fixations for each person on a particular web page and the visual elements of the web page are taken as input. The individual paths are then represented in terms of the visual elements by identifying the corresponding element of each fixation. When the individual path of a person is represented as X [150 ms] Y [100 ms] Z [200 ms], it means that the person looked at the elements X, Y and Z for 150 ms, 100 ms and 200 ms respectively.

First Pass: A visual element can exist consecutively (XYYZ) and/or non-consecutively (XYXZ) in an individual path. Each non-consecutive existence of a visual element in an individual path is referred to as the instance of the element. The instances of a particular visual element are differentiated from each other with the use of numbers where the longest instance receives the first number. A visual element instance is identified as trending if it satisfies the following rules where X represents the tolerance level parameter.

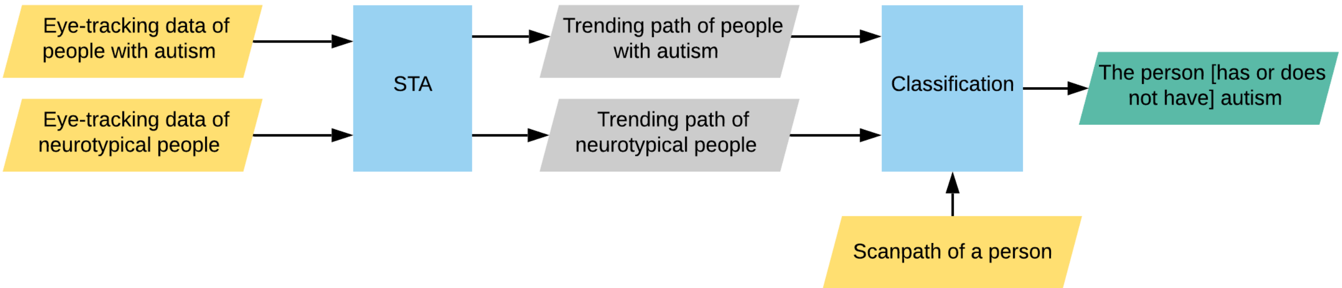

Figure 2: An overview of autism detection with STA

- The total number of the existence of the instance >= the minimum total number of the existence of the instances that are shared by X percent the individual paths.
- The total duration of the instance >= the minimum total duration of the instances that are shared by X percent of the individual paths.

If the tolerance level is set to 0.80, then the trending instances are identified based on the instances which are shared by 80% of the individual paths. The trending instances are kept and other instances are removed from the individual paths.

Second Pass: The trending instances are combined based on their overall positions in the individual paths to generate the trending path. For each individual path, the same instances are combined (e.g. X1 [150 ms] X1 [200 ms] Y2 [400 ms] → X1 [350 ms] Y2 [400 ms]) and then their sequential priority values are computed with Equation 1:

$$\psi_i = 1 - P_i \cdot \frac{\max_i - \min_i}{L-1} \tag{1}$$

where:

ψ_i = Sequential priority value of ith instance in the individual path

P_i = Position of the instance in the individual path (starting from 0)

L = Length of the individual path.

$\max_i$ = Maximum sequential priority value (default: 1)

$\min_i$ = Minimum sequential priority value (default: 0.1)

The total priority value for each trending instance is then computed with Equation 2.

$$\Psi = \sum_{i=1}^{n} \psi_i \tag{2}$$

where:

Ψ = Total priority value of an instance

n = The number of trending instances

The trending path is then generated by sorting the trending instances based on their total priority values in descending order. In case of the same priority values, the total duration and the total number of the existence of the instances are also considered. The identification numbers of the instances are then deleted (e.g. X1→X) and the consecutive repetitions are excluded (e.g. XYYZ → XYZ), and finally the trending path is represented in terms of the visual elements.

3.2 Classification

When there is an individual path on a particular page, its similarity is computed to the trending paths of people with autism and neurotypical people on that page by using the Levenshtein distance which is commonly referred to as the String-edit distance [17]. This similarity measure has widely been used for comparing two scanpaths which are represented as string sequences [17]. It represents a minimum number of editing operations (addition, deletion and/or substitution) required to transform one string to another one. For example, the String-edit distance between the paths XYZ and XYA is equal to one as one substitution operation with Z and A is able to transform one path to another. Equation 3 is then used to calculate a similarity score as a percentage.

$$S = 100 \cdot (1 - \frac{D}{L}) \tag{3}$$

where:

S = Similarity score

D = String-edit distance

L = The length of the longer path

If the path is more similar to the trending path of neurotypical people, the person of the path will be classified as a neurotypical person. Otherwise, s/he will be classified as a person with autism. If there is no trending path identified for neither people with autism nor neurotypical people (which could happen with very high tolerance levels, such as 1.00), the person of the path will be classified as a person with autism. This classification algorithm is also illustrated in Algorithm 1.

Algorithm 1: Classification of an individual scanpath based on STA

```
Data: Individual path (P), Trending path of people with autism
      (TP1), Trending path of neurotypical people (TP2)
Result: Class of P
D1 = LevenshteinDistance(P, TP1)
S1 = 100 * (1 - D1 / max(length(P), length(TP1))
D2 = LevenshteinDistance(P, TP2)
S2 = 100 * (1 - D2 / max(length(P), length(TP2))
if S1 <= S2 then
  | return Person with Autism
else
  | return Neurotypical Person
end
```

4. EVALUATION

We aim to evaluate our approach by comparing it with the most relevant approach which has been proposed in our previous work [48]. Therefore, we use the same dataset and follow the same methodology.

4.1 Dataset

The dataset includes a series of fixations for each participant on six web pages for two different kinds of tasks. We briefly explain the dataset below. The full description of the dataset can be found in [48].

Participants: The dataset was constructed with 15 verbal and highly-independent people with autism (Female: 6, Male: 9) and 15 neurotypical people (Female: 7, Male: 8). The group of people with autism was referred to as the ASD group and the group of neurotypical people was referred to as the control group. The participants with autism were recruited through a UK charity organisation, while the control-group participants were recruited through snowball sampling. The participants in both groups were from the West Midlands area of the UK.

The inclusion criteria for the ASD group was a formal diagnosis of autism (Asperger's syndrome and High-functioning autism were also acceptable diagnoses for those participants that were diagnosed before the introduction of DSM-5), being over 18 years of age and being able to use a computer. The inclusion criteria for the control group were similar, except for the presence of autism. To ensure that no participants with a high incidence of autistic traits were included in the control group, participants were screened using the 50-item Autism Quotient test [7]. Exclusion criteria included the presence of any degree of intellectual disability, disorders that may affect reading (other than autism) and any conditions affecting vision that could not be corrected with glasses or lenses.

The mean age of the ASD group was 37 (SD: 9.14) whereas the mean age of the control group was 33.6 (SD: 8.6). In addition, the mean of the number of years spent in formal education was 16 (SD: 3.33) for the participant with autism and 18.35 (SD: 2.47) for the neurotypical participants. All of these participants were daily web users and were highly independent (i.e. none of them relied on a caregiver in their day-to-day life).

Materials: The participants viewed the home pages of six websites which were randomly selected from the top sites listed by Alexa.com by ensuring that their home pages had varying levels of visual complexity determined with the ViCRAM tool [33]. The websites along with their visual complexity levels were as follows: Apple (Low), Babylon (Low), AVG (Medium), Yahoo (Medium), GoDaddy (High) and BBC (High). The participants were asked to specify how often they visit these pages (daily, weekly, monthly, less than once a month, never) and their answers show that these web pages were not frequently visited by the majority of the participants in both of the ASD and control groups (monthly, less than once a month or never), apart from the BBC page which was visited by the majority of the participants from the ASD group either daily or weekly (see the details in [14]).

Similar to our previous work [48], the web pages were divided into their elements by using the extended VIPS algorithm [1] as STA is conducted based on visual elements of web pages. The VIPS algorithm has widely been used to divide web pages into their elements [1]. It provides the elements in a hierarchical form where deeper levels include more and smaller elements. The fifth level was used as it was found to be the most preferred level by users based on a study conducted by [1].

Tasks: The participants completed two different types of tasks for each page.

- Browse task: The participants were instructed to freely browse the web pages for a maximum of 120 seconds without being required to answer any specific questions. This involved the selective inspection of page elements in a spontaneous order. The participants were free to proceed to the next page once they were satisfied with their familiarity with the current page. If they did not want to stay on the web page for 120 seconds, they were allowed to continue with other pages.
- Search task: The participants were asked to find specific information within each page. For example, the participants were asked to find a telephone number for technical support and also the text box where they could search for a new domain. There were two questions per page and the time limit for answering both was 30 seconds. After completing the tasks, the participants could proceed to the next page.

Equipment: The eye movements of the participants were recorded with a Gazepoint GP3 video-based eye-tracker on a 19" LCD monitor with 1440 x 900 screen resolution [23]. The degree of accuracy of this eye tracker was given as 0.5-1 degree. The distance between the participants and the eye tracker was approximately 65 cm.

Procedure: The participants firstly read the information sheet and signed the consent form to accept for taking part in the study. The neurotypical participants were also asked to complete the Autism Quotient test to ensure they do not have a high level of autistic traits. The demographic data was then collected, including age, gender, education, and web usage. After that, they started their eye-tracking sessions and they viewed all the pages twice in a counter-balanced random order for both types of tasks.

4.2 Methodology

We first divided the participants in each group into two sets called training (10 participants) and testing sets (5 participants) which is a typical hold-out method [5] (we used 2/3 of the dataset as training set). We used the training sets to discover the trending paths of the participants with autism and the neurotypical participants. We then combined the testing sets of the two groups. After that, we compared each participant's path in the combined testing set (10 participants in total) with the trending scanpaths of the participants with autism and neurotypical participants to classify the participant. Precision, recall, F1-score, and accuracy were then computed. This hold-out method was repeated for each page 100 times and we computed the mean of precision, recall, F1-score, and accuracy for each page. We followed this methodology with each tolerance level of the STA algorithm [0.01-1.00] and selected the the case with the highest F1-score for each page.

We then compared the results of our proposed STA approach with the results of our previous logistic regression approach [48]. In order to do a fair comparison, we used the same dataset and same methodology for evaluation of the results. Since only the accuracy values are reported for the logistic regression approach, we compare the logistic regression approach with the STA approach based on the accuracy values instead of F1-scores. As the STA algorithm is designed to identify a trending path of multiple users on a particular web page, we compare these approaches based on individual web pages instead of different combinations of web pages. Besides, we consider the accuracy values of the logistic regression approach achieved with only eye-tracking data for the comparison purpose because other features do not improve its classification accuracy in general, and we want to compare it with our approach based on how they work with a single source of data.

Furthermore, we also investigated the effect of the tolerance parameter of the STA algorithm on the results. We calculated the standard deviation of precision, recall, F1-score and accuracy values achieved by our approach for each web page with all tolerance levels of the STA algorithm for the browsing and searching tasks.

5. RESULTS

To illustrate how the trending paths of people with autism and neurotypical could differ from each other, Figure 3 and Figure 4 show two trending paths identified for 10 people with autism and 10 neurotypical people respectively on the home page of the Yahoo website with the tolerance parameter 0.90 for the browsing task[2]. In this case, the trending path of people with autism consists of more and repeated elements and has more transitions between these elements.

Figure 3: The trending path of 10 people with autism on the Yahoo page with the tolerance level 0.90 for the browsing task

Table 1 shows the mean values of precision, recall, F1-score and accuracy for each web page achieved by our approach for autism detection for the browsing and searching tasks. These values show that the sequential analysis of eye movements allows to detect autism with accuracy higher than chance.

Table 2 shows the comparison of the accuracy values achieved by the logistic regression approach and the STA approach based on individual pages. In addition, Figure 5 and Figure 6 visualises these comparisons by column charts. The mean accuracy values of the STA and logistic regression approaches are 60.00% (SD: 3.2) and 54.67% (SD: 6.15) for the browsing tasks and 58.20% (SD: 3.12) and 56.00% (SD: 10.84) for the searching tasks respectively. These results show that when individual pages are taken into account, higher accuracy and lower standard deviations are observed when using the STA approach with sequential data compared to using the

[2]STA does not provide the exact locations of fixations and therefore the locations of fixations in the figures are for illustration purposes only.

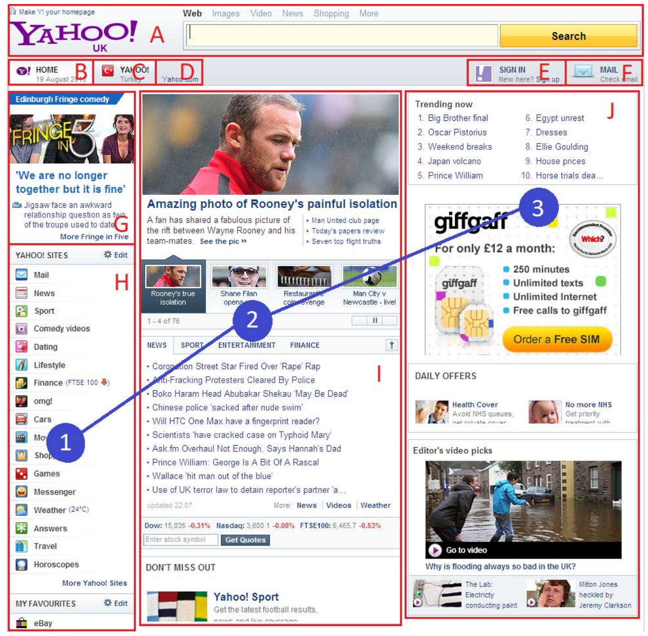

Figure 4: The trending path of 10 neurotypical people on the Yahoo page with the tolerance level 0.90 for the browsing task

Task	Page	Precision	Recall	F1	Accuracy
Browse	Apple	0.63	0.61	0.60	61.00%
	Babylon	0.61	0.58	0.55	57.90%
	AVG	0.67	0.65	0.63	64.80%
	Yahoo	0.63	0.61	0.57	60.90%
	GoDaddy	0.62	0.60	0.58	60.30%
	BBC	0.55	0.55	0.52	55.10%
Search	Apple	0.59	0.59	0.57	58.60%
	Babylon	0.67	0.63	0.60	63.10%
	AVG	0.53	0.57	0.50	56.50%
	Yahoo	0.52	0.55	0.48	55.10%
	GoDaddy	0.52	0.56	0.47	55.50%
	BBC	0.63	0.60	0.57	60.40%

Table 1: The mean values of precision, recall, F1 and accuracy for each web page achieved by our approach for the browsing and searching tasks

logistic regression approach with non-sequential data. Therefore, the advantage of the sequential approach is the stability of the predictions, as shown by the lower standard deviation compared to the non-sequential approach.

To illustrate the effect of the tolerance level of the STA algorithm, Table 3 shows the standard deviation of precision, recall, F1-score and accuracy for each web page for the browsing and searching tasks for all the tolerance levels [0.01-1.00]. The mean values of the standard deviation of the F1-scores and accuracy are 0.06 and 3.5% respectively for the browsing tasks and 0.05 and 3.6% respectively for the searching tasks. Therefore, these suggest that the tolerance level has minor effect on the overall results.

6. DISCUSSION

The results of this study show that the analysis of eye-movement sequences from web-page processing allows autism detection with an above-chance accuracy, owing to differences in the sequence in which the two groups attend to the elements.

In comparison to the most relevant non-sequential approach, the

Task	Page	Logistic Regression [48]	STA
Browse	Apple	63 %	61.00 %
	Babylon	54 %	57.90 %
	AVG	59 %	64.80 %
	Yahoo	52 %	60.90 %
	GoDaddy	55 %	60.30 %
	BBC	45 %	55.10 %
Search	Apple	69 %	58.60 %
	Babylon	63 %	63.10 %
	AVG	39 %	56.50 %
	Yahoo	48 %	55.10 %
	GoDaddy	60 %	55.50 %
	BBC	57 %	60.40 %

Table 2: The comparison of the accuracy values achieved by the logistic regression approach and the STA approach based on individual pages

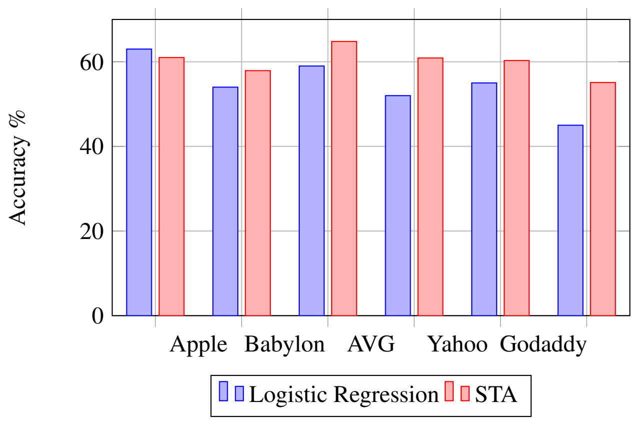

Figure 5: The comparison of the accuracy values achieved by the logistic regression approach and the STA approach for the browsing tasks

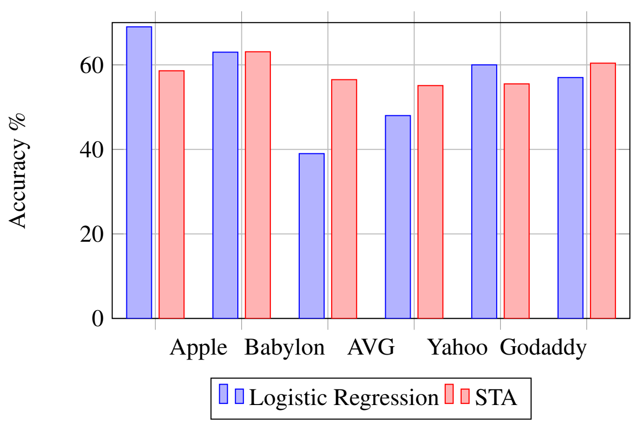

Figure 6: The comparison of the accuracy values achieved by the logistic regression approach and the STA approach for the searching tasks

analysis of sequential data using the STA algorithm provides not only more stable results but also higher accuracy when taking individual pages into consideration. This finding has direct implications for the development of autism-detection screening tools, as it is a step towards achieving stable results across stimuli, thus reducing the number of tasks and pages the participants need to complete

Task	Page	Precision	Recall	F1	Accuracy
Browse	Apple	0.06	0.05	0.05	4.57
	Babylon	0.07	0.05	0.05	4.85
	AVG	0.10	0.02	0.06	2.03
	Yahoo	0.05	0.04	0.03	3.51
	GoDaddy	0.05	0.04	0.04	3.73
	BBC	0.09	0.03	0.06	3.00
Search	Apple	0.12	0.04	0.08	4.09
	Babylon	0.08	0.03	0.05	2.88
	AVG	0.11	0.04	0.08	4.03
	Yahoo	0.08	0.03	0.05	3.23
	GoDaddy	0.10	0.04	0.07	4.33
	BBC	0.05	0.02	0.04	2.50

Table 3: The standard deviation of precision, recall, F1-score and accuracy values for each web page achieved by our proposed approach for the browsing and searching tasks for all tolerance levels of the STA algorithm [0.01-1.00]

and attend to in order to obtain a reliable screening result. The impact of these results is also relevant to screening tools for the detection of other attention-related conditions, such as the screening methods for dyslexia proposed by [38]. Additionally, the produced evidence that the two groups attend to page elements in order sufficiently different to allow consistent above-chance user classification has implications for the development of autism-friendly web pages and websites.

There are several possible factors that may have influenced the results that need further investigation. For example, in our study, we directly use the entire scanpaths to identify the trending scanpaths without any pre-processing. However, some pre-processing steps can be conducted to deal with any possible noisy data. For example, the fixations made after the completion of searching tasks could be considered noise and excluded in order to contain signals from the search tasks alone. One may also consider ignoring the very first few fixations as there may be some possible contamination from the preceding stimulus, i.e. these fixations could potentially represent the position of the eyes when leaving the previous web page. If these assumptions were true, these pre-processing steps could serve as a noise-removal procedure and result in increased overall accuracy. Besides the noise elimination techniques in data, we can also apply different *windowing*[3] techniques to the sequence data – in order words, instead of using the full sequences, we can slice the sequence data into different window sizes and try to see if the generated scanpaths are more accurate in prediction. We can also explore windowing time frames with respect to task completion or reaching certain parts of the page.

Another factor that may have had a direct impact on the results is the web page segmentation. We use the VIPS algorithm to segment the web pages into their elements due to its popularity in the literature for web page segmentation and due to the fact that our previous work has suggested that it provides higher accuracy compared to other existing approaches [48]. We can also explore how other web page segmentation approaches would affect the overall results with STA. For example, an alternative approach to VIPS is Block-o-Matic (BOM) which can also be used to segment web pages [41]. Besides these approaches, we could also divide the pages into different sizes of grids and then explore the effect as we have done in our previous work [48]. Therefore, future work may include repeated evaluations using other web page segmentation algorithms including grid segmentation with different granularity levels.

The proposed approach is a foundation step towards improving

[3] https://towardsdatascience.com/ml-approaches-for-time-series-4d44722e48fe

the stability of autism detection across stimuli and tasks, however, it has the limitation that it currently does not allow the combination of signals from different web pages. Therefore, future work would aim to combine the stability of the results achieved by this approach with the ability to use the signals from multiple pages, as in [48]. Furthermore, as the dataset consists of six web pages, we are not able to investigate how the features of web pages such as visual complexity and text density affect autism detection and the order to element processing. Additional studies with more web pages will allow us to investigate the effects of these features.

Finally, in this research we use real web pages that were not particularly designed for aiding autism detection. However, as it is common in the related work (see Section 2), one can also create synthetic pages that are mainly designed in aiding autism detection. In order to do this, of course further studies are needed to better understand what kind of structural elements and formulation of those elements can better guide this process. It is for example critical to understand what kind of elements or structures cause more difficulties to people with autism. One can argue that this would be very challenging but if we conduct more and different eye-tracking studies with varying complexity of web pages with different formulation of structural contents and tasks, we will be able to have very large dataset that can help us better understand the phenomena. Using synthetic web pages would also allow to deal with any possible familiarity issues.

7. CONCLUSION

This study shows that eye-movement sequences can be used as an indicator of detecting autism. The main contribution is the finding that the analysis of sequential data provides more stable results compared to non-sequential data, which can help overcome drawbacks in stimulus selection. By using the STA algorithm, we could achieve approximately 60% accuracy for both the browsing and searching tasks with very minor variations among the web pages. These results were achieved with verbal and highly-independent people with autism who are difficult to diagnose and we do not make claims for people with autism who are non-verbal, have intellectual disability and/or do not frequently use the web. We can also conduct the evaluation of our approach with a different group of people with autism to see how the accuracy will be affected.

An important direction for future work is to explore how other non-sequential features (e.g. the number of transitions, the number of irrelevant elements fixated, etc.) can be integrated into this approach for achieving higher accuracy. Although we use the STA algorithm in this study, other sequence-based algorithms can also be investigated in the future to see how they work for autism detection.

8. REFERENCES

[1] M. E. Akpinar and Y. Yesilada. Vision based page segmentation algorithm: Extended and perceived success. In Q. Z. Sheng and J. Kjeldskov, editors, *Current Trends in Web Engineering*, pages 238–252, Cham, 2013. Springer International Publishing.

[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*. American Psychiatric Publishing, Arlington, VA, 5th edition, 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] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statist. Surv.*, 4:40–79, 2010.

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

[7] 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.

[8] 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.

[9] M. A. Betancourt, L. S. Dethorne, K. Karahalios, and J. G. Kim. Skin conductance as an in situ marker for emotional arousal in children with neurodevelopmental communication impairments: Methodological considerations and clinical implications. *ACM Trans. Access. Comput.*, 9(3):8:1–8:29, Feb. 2017.

[10] 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.

[11] 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.

[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*, W4A '17, pages 20:1–20:4, New York, NY, USA, 2017. ACM.

[14] S. Eraslan, V. Yaneva, Y. Yesilada, and S. Harper. Web users with autism: eye tracking evidence for differences. *Behaviour & Information Technology*, pages 1–23, 2018.

[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] S. Eraslan, Y. Yesilada, and S. Harper. Identifying patterns in eyetracking scanpaths in terms of visual elements of web pages. In S. Casteleyn, G. Rossi, and M. Winckler, editors, *Web Engineering*, pages 163–180, Cham, 2014. Springer International Publishing.

[17] S. Eraslan, Y. Yesilada, and S. Harper. Eye Tracking Scanpath Analysis Techniques on Web Pages: A Survey, Evaluation and Comparison. *Journal of Eye Movement Research*, 9(1:2):1–19, 2016.

[18] S. Eraslan, Y. Yesilada, and S. Harper. Scanpath trend analysis on web pages: Clustering eye tracking scanpaths. *ACM Trans. Web*, 10(4):20:1–20:35, Nov. 2016.

[19] S. Eraslan, Y. Yesilada, and S. Harper. Engineering web-based interactive systems: Trend analysis in eye tracking scanpaths with a tolerance. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '17, pages 3–8, New York, NY, USA, 2017. ACM.

[20] S. Eraslan, Y. Yesilada, V. Yaneva, and S. Harper. Autism Detection Based on Eye Movement Sequences on the Web: A Scanpath Trend Analysis Approach, Feb. 2020. https://doi.org/10.5281/zenodo.3668740.

[21] T. Falck-Ytter, S. Bölte, and G. Gredebäck. Eye tracking in early autism research. *Journal of neurodevelopmental disorders*, 5(1):28, 2013.

[22] 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, Jun 2013.

[23] Gazepoint. *Open Gaze API by GazePoint (Version 2.0)*, 2013.

[24] F. Happé and U. Frith. The Weak Coherence Account: Detail-focused Cognitive Style in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 36:5–25, 2006.

[25] S. Harper, S. Eraslan, and Y. Yesilada. It's all about the message: Visual experience is a precursor to accurate auditory interaction. In *Proceedings of the 16th Web For All 2019 Personalization - Personalizing the Web*, W4A '19, pages 20:1–20:8, New York, NY, USA, 2019. ACM.

[26] P. Hejmady and N. H. Narayanan. Visual attention patterns during program debugging with an ide. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 197–200, New York, NY, USA, 2012. Association for Computing Machinery.

[27] R. S. Hessels, R. Andersson, I. T. Hooge, M. Nyström, and C. Kemner. Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, 20(6):601–633, 2015.

[28] R. S. Hessels, I. T. Hooge, T. M. Snijders, and C. Kemner. Is there a limit to the superiority of individuals with asd in visual search? *Journal of autism and developmental disorders*, 44(2):443–451, 2014.

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

[30] 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.

[31] 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.

[32] W. Liu, M. Li, and L. Yi. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, 9(8):888–898, 2016.

[33] E. Michailidou, S. Harper, and S. Bechhofer. Visual complexity and aesthetic perception of web pages. In *Proceedings of the 26th Annual ACM International Conference on Design of Communication*, SIGDOC '08, pages 215–224, New York, NY, USA, 2008. ACM.

[34] 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.

[35] T. Pellegrini, L. Fontan, J. Mauclair, J. Farinas, C. Alazard-Guiu, M. Robert, and P. Gatignol. Automatic assessment of speech capability loss in disordered speech. *ACM Trans. Access. Comput.*, 6(3):8:1–8:14, May 2015.

[36] K. J. Price and A. Sears. The development and evaluation of performance-based functional assessment: A methodology for the measurement of physical capabilities. *ACM Trans. Access. Comput.*, 2(2):10:1–10:31, June 2009.

[37] G. E. Raptis, C. Fidas, C. Katsini, and N. Avouris. A cognition-centered personalization framework for cultural-heritage content. *User Modeling and User-Adapted Interaction*, 29(1):9–65, Mar 2019.

[38] L. Rello and M. Ballesteros. Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proceedings of the 12th Web for All Conference*, page 16. ACM, 2015.

[39] L. Rello, K. Williams, A. Ali, N. C. White, and J. P. Bigham. Dytective: towards detecting dyslexia across languages using an online game. In *Proceedings of the 13th Web for All Conference*, page 29. ACM, 2016.

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

[41] A. Sanoja and S. Gançarski. Block-o-matic: A web page segmentation framework. In *International Conference on Multimedia Computing and Systems (ICMCS)*, pages 595–600, Marrakesh, Moroco, 2014.

[42] C. L. Tablatin and M. M. Rodrigo. Identifying common code reading patterns using scanpath trend analysis with a tolerance. In *Proceedings of thee 26th International Conference for Computers in Education*, ICCE 2018, pages 286–291, Metro Manila, Philippines, 2018.

[43] Tobii Technology. *Tobii Studio 2.X User Manual (version September 2010)*. Tobii Technology AB, 2010.

[44] 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.

[45] G. Wan, X. Kong, B. Sun, S. Yu, Y. Tu, J. Park, C. Lang, M. Koh, Z. Wei, Z. Feng, et al. Applying eye tracking to identify autism spectrum disorder in children. *Journal of autism and developmental disorders*, 49(1):209–215, 2019.

[46] S. Wang, M. Jiang, X. M. Duchesne, E. A. Laugeson, D. P. Kennedy, R. Adolphs, and Q. Zhao. Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron*, 88(3):604–616, 2015.

[47] J. M. West, A. R. Haake, E. P. Rozanski, and K. S. Karn. Eyepatterns: Software for identifying patterns and similarities across fixation sequences. In *Proceedings of the 2006 Symposium on Eye Tracking Research Applications*, ETRA '06, pages 149–154, New York, NY, USA, 2006. Association for Computing Machinery.

[48] V. Yaneva, L. A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov. Detecting autism based on eye-tracking data from web searching tasks. In *Proceedings of the Internet of Accessible

Things, W4A '18, pages 16:1–16:10, New York, NY, USA, 2018. ACM.

[49] 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.

[50] V. Yaneva, I. P. Temnikova, and R. Mitkov. A corpus of text data and gaze fixations from autistic and non-autistic adults. In *LREC*, 2016.