

One size does not fit all: detecting attention in children with autism using machine learning

Bilikis Banire¹ · Dena Al Thani¹ · Marwa Qaraqe¹

Received: 28 April 2022 / Accepted in revised form: 15 May 2023 / Published online: 17 June 2023 © The Author(s) 2023

Abstract

Detecting the attention of children with autism spectrum disorder (ASD) is of paramount importance for desired learning outcome. Teachers often use subjective methods to assess the attention of children with ASD, and this approach is tedious and inefficient due to disparate attentional behavior in ASD. This study explores the attentional behavior of children with ASD and the control group: typically developing (TD) children, by leveraging machine learning and unobtrusive technologies such as webcams and eye-tracking devices to detect attention objectively. Person-specific and generalized machine models for face-based, gaze-based, and hybrid-based (face and gaze) are proposed in this paper. The performances of these three models were compared, and the gaze-based model outperformed the others. Also, the person-specific model achieves higher predictive power than the generalized model for the ASD group. These findings stress the direction of model design from traditional one-size-fits-all models to personalized models.

Keywords Attention · Autism · Face-tracking · Eye-tracking · Machine learning

1 Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by a deficit in social communication and repetitive behavior (American Psychiatric Association 2013). Attentional deficit is one of the conspicuous deficits of ASD, hence, children with ASDrequire extra attentional support during learning. Teachers commonly use subjective attentional assessment to monitor the learning progress of their students. The subjective approach is usully through observational method which is tedious, time-consuming, and requires a long year of experience as a child with ASD

Bilikis Banire banire.bilikis.o@gmail.com

¹ Information and Computing Technology Division, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

exhibits unique attentional behavior (Marcu et al. 2013; Douglas et al. 2022; Wood et al. 2016). For example, a child with ASD may prefer to look at stimuli and listens simultaneously while another child may choose to look at the stimuli and not listen. It is, therefore, challenging for teachers to manage both the learning content and the diverse attentional behavior of their students concurrently (Wood et al. 2016). Research has shown that there is a lack of resources and experienced teachers for children with ASDdue to the increased rate of ASDprevalence in the United States of America, as reported by the Center for Disease Control and Prevention (CDC) (CDC 2022; Lee and Meadan 2021;), The increase in ASDpopulation is not limited to the USA alone but globally including Qatar, the country y of the current study (Alshaban et al. 2019). The experience and feelings of families and friends who have children with ASDshow that the children require a great deal of educational and social support from teachers, parents, siblings, relatives, and friends (Sharabi and Marom-Golan 2018). Sometimes, the support needed by children with ASDspans from childhood to adulthood, which is usually overwhelming and psychologically challenging (Russa et al. 2015). Hence, understanding how stakeholders can benefit from objective attention assessment of children with ASDcan support them in managing their learning experience effectively.

Parents do agree that attention impairment is the root cause of the core deficits in children with ASD(Kinnealey et al. 2012; Ridderinkhof et al. 2020). Thus, existing technological interventions for children with ASDoften focus on using technology to gain the attention of children with ASDto improve core ASDimpairments, such as social and communication skills (Almumen and Almuhareb 2020). While some studies have assessed the attention of children with ASD using these interventions, very few studies, have explored how technology can be utilized in assessing attention objectively. Most importantly, understanding how technologies can be used to measure attention from different perspectives. The common form of attention assessment is mostly through cognitive assessment when students respond correctly to a given attention task. According to the work of Fredricks et al. (Fredricks 2011), engagement or attention in our context can be viewed from three dimensions: behavioral, emotional and cognitive. Several studies have shown that attentional context, especially behavior and emotional engagement are measurable with various kinds of sensors (camera, pressure sensor, eye tracker, etc.) Dewan et al. (Dewan et al. 2019). Emotion is a way of expressing inner feelings toward an action. Individuals express emotions in different ways such as facial expressions, body language or vocalizations. Yet most prominently emotions are express through facial expression (Tsiourti et al. 2019). There are seven basic emotions; anger, sadness, disgust, joy, surprise, fear and contempt (Rosenberg and Ekman 2020). Clusters of facial muscle activities, also known as facial action units (FACS) describes emotions. Categories of emotions include positive, negative and neutral. Emotional states that depict enjoyment or good states such as joy and surprise falls under positive emotion. Negative emotions imply displeasure in something or people such as anger, sadness, disgust, fear and contempt. The neutral emotional state represents indifferent emotion. Educators and psychologists suggest that children's emotions can affect their ability to focus on a task (Escobedo et al. 2014; Neuhaus et al. 2019).

The advent of sensing technologies and machine learning has enhanced intelligent systems that monitor and respond to human behavior. An intelligent system using a machine learning algorithm is being applied by experts in different domains like education (Hutt et al. 2021), transportation (Palazzi et al. 2018), and healthcare (Chen et al. 2021). The development of a machine learning model for detecting the attention in children with ASD is still at infancy due to the difficulty in their generalizing attentional behaviors (Jacob et al. 2019). However, there are existing studies investigating on application of machine learning for detecting autism (Alvari et al. 2021; Carpenter et al. 2021; Thabtah and Peebles 2020). Machine learning methods have revolutionized many research fields such that they are capable of automatically detecting human behavior based on extracted features describing their reaction (Zanker et al. 2019). It is evident from previous research that attentional behavior of typically developing (TD) population can be measured using multiple behavioral data such as physiological behaviors including brain signals and heart rate (Belle et al. 2011; Yulita et al. 2019), brain signals, body movement and heart rate (Sonne et al. 2015; Mamun et al. 2019), facial expression and eye gaze (Aslan et al. 2014) to train a machine learning model. The objective of combining different attentional behavior is to compare the performance of a hybrid method to individual methods of measuring attention. However, the assessment of attention in children with ASDis often done subjectively using video data analysis where researchers code frames of recorded activity sessions as attention or inattention (Kinnealey et al. 2012; Banire et al. 2015; Higuchi, et al. 2018; Lee and Schertz 2020), while very few studies used automated attention detection using a computer vision algorithm to capture facial movement (Egger et al. 2018) and machine learning model built with eye-tracking data as a biomaker of detecting autism rather than detecting their attention (Yaneva 2020). These studies show the potential of facial and eye-tracking features for behavioral assessment. However, to the best of our knowledge, no study has investigated the performance of facial and eye-tracking features for developing a machine learning model capable of detecting the attention of children with ASD.

Thus, this current study proposes a behavioral-based attentional model using a bimodal approach: face and gaze-based models for children with ASD. The performance of each model type will be evaluated and compared. The model comparison can inform the direction of future studies on prominent attentional behavior in children with ASD. Toward the goal of this study, three major research questions are proposed: (1) Is the attentional unimodal sufficient for detecting attention in children with ASD? (2) what are the prominent facial and eye-gaze features for detecting attention? (3) What differentiates the attention model of children with ASD from TD?

In this paper, a novel machine learning model for detecting the attention of children with ASD using facial and eye-tracking features is proposed. Also, the performance of individual channels and hybrid channels are compared to identify the prominent features for detecting attention in children with ASD. Lastly, the machine learning models using these two channels are compared in children with ASD and TD children.

The structure of this paper is as follows: Sect. 2 describes the works that have been done to detect attention in children with ASD. Section 3 describes the methodology that has been used in this paper. The experimental results are discussed in Sect. 4, and Sect. 5 presents the discussion and the direction of future work. Lastly, Sect. 6

summarizes the findings of the paper and Sect. 6 presents the limitation of this study. Lastly, Sect. 7 summarizes the findings of the paper.

2 Related works

This section describes machine learning and its application in attention detection also referred to attentional model in this paper. The attentional model leverages machine learning to predict attentional behavior of students. Researchers support that attentional model can be developed using different physiological behaviors (multimodal methods) such as brain signals and heart rate (Belle et al. 2011) brain signals, body movement, and heart rate (Sonne et al. 2015); and facial expression and eye gaze (Aslan et al. 2014) while other studies use only one physiological behavior (unimodal) such as brain signals (Yang et al. 2015; Ghassemi et al. 2009). The objective of combining different physiological behavior is to compare the performance of a using unimodal and multimodal-based models for purpose of identifying a better approach for attention detection tasks.

2.1 Unimodal attentional model

The unimodal attentional model applies a single channel of physiological behavior to annotate attention. Bosch, D'mello (Bosch et al. 2016) apply facial features using appearance-based methods and body posture to detect learners' engagement. The authors gave 137 students in 8th and 9th grade an educational game about physics and coded the actions of the participants for on-task and off-task behaviors. The findings from their study show that the model performance using the AUC for off-task detection was above chance i.e., greater than 0.5 for a generalized model (AUC = 0.816). Additionally, the authors found that the face-based model could generalize across temporal, ethnicity and gender.

Similarly, (Whitehill et al. 2014) used the appearance-based method on facial features to recognize students' engagement. The students played a cognitive skills training software, and a Logitech web camera was used to record the entire session. The authors used a team of labelers consisting of undergraduate and graduate students from computer science, cognitive science, and psychology. These labelers viewed the recorded videos and labeled the learner's facial appearance from 34 participants into four levels of engagement: not engaged, nominally engaged, engaged, and very engaged. The findings in this study revealed that the SVM classifier had the highest performance score among the classifiers for the generalized model (AUC = 0.729). The authors evaluated how a face-based model generalizes across demographic factors, such as ethnicity. The model was trained with Caucasian and Asian-American populations and tested on an African American dataset. They found that the model generalized across ethnicity, and the prediction were above chance (AUC = 0.691).

Aside from facial features, other studies have used eye-tracking features to predict attentional-related behaviors. Bixler and D'Mello (2015) used a classification approach to predict mind wandering in a reading task among college students. This study used fixation, saccade, and pupil diameter features with four different algorithm classifiers. The best classifier algorithm was SVM with 74% accuracy. Lallé and Conati (2018) developed a gaze-based classification model to predict confusion during interactions with a visualization task. The study used gaze patterns, pupil size, and head distance as measurements. The authors achieved an accuracy of 61% using the RF classifier algorithm. They identified variations in user pupil size and head distance as good predictors of attentional state. Chen and Chen (2017) examined the states of attention and inattention of 15 participants during mathematics, continuous performance tasks (CPT), and reading tasks using SVM classifier algorithms. The gaze-based attentional model used achieved 93.1% accuracy. Finally, Shojaeizadeh and Djamasbi (2019) predicted task demand in a visualization task using the RF classifier algorithm using the eye-tracking features such as fixations, saccades, blinks, and pupillary responses. This study achieved an accuracy of 79%, with pupillary responses being the best task predicting feature.

Overall, these studies adopting unimodal attentional model supports the evidence that machine leaning algorithms are effective for automated attention assessment even with one channel of features such as gaze and facial features. The current study explores the two different channels both facial and gaze features independently. The outcome of machine learning model performance will reveal best channel of festures for detecting attention in children with ASD.

2.2 Multimodal attentional model

In multimodal attentional models, two or more sensors are used to record behavioral measures during attention tasks. Asteriadis and Tzouveli (2009) investigated facial and head poses using a webcam to predict the attentive and frustrated state of learners. This study applied the fuzzy rule classification method and achieved an accuracy of 80–88%. In 2010, D'mello & Graesser utilized a multimodal method of engagement recognition and applied it to facial features, body posture, and communication clues. (D'mello and Graesser 2010). Through the model evaluations using kappa scores, it was determined that a two-model combination performs similarly to a three-model method.

Another study on multimodal methods exploited facial features and interaction data to predict engagement and frustration. The authors used a multilayer perceptron (MLP) and achieved an accuracy of 78% (Shaker et al. 2013). A similar study by Monkaresi et al. (2016) utilized facial features and heart rate to detect engagement. This study applied RF and achieved a prediction power above the chance level (AUC = 0.758). The study by Chen and Tsai (2017) used SVM to train head pose and eye movement from 10 participants to develop an attention recognition model. The accuracy achieved in this study was 93.1%. Another multimodal model study used SVM to develop an engagement recognition system and achieved 85% accuracy (Aslan et al. 2014). The hybrid-based model in our research is based on multimodal methods used in the reviewed studies.

Aside studies adopting unimodal attentional models, multimodal models are also effective. Thus, the comparison of multimodel model with the unimodal model is one of the objectives of the current study.

3 Research gap

Related work on the single and multiple attentional models indicates that both methods have the potential to differentiate attention and inattention. However, the multimodal attentional model has a better advantage over the unimodal methods as it relies on multiple attentional behaviors. It is also important to mention that the related work relied on the subjective labeling of students' attentional behaviors who do not have the neurocognitive disorder. One of the drawbacks of the subjective method of labeling attention is inconsistency due to differences in the background experiences of the experts and ambiguity (Dewan et al. 2019), especially in children with ASD who exhibit unique attentional behavior. However, combining observable attentional behavior with cognitive processing assessments such as performance scores can reduce the ambiguity of annotating attentional behavior in children with ASD. Additionally, the generalizability of the model across demographic information was also explored to evaluate the robustness of the attentional model. The generalizability of the behavioral-based attentional model across groups of children with ASD, TD, and attention task types is worth investigating.

4 Methodology

The framework of the hybrid-based model for attention recognition used in this study is shown in Fig. 1. It consists of the face and eye-tracking data extracted from the video frames captured during the experiment. These two sets of data were adopted based on the evidence from the unimodal studies reviewed in the current paper. Furthermore, the methods of obtaining the data are unobtrusive which prevents children with ASD from experiencing sensory issues due to body contact with the tools. We utilized the geometric distance between 34 facial landmarks and exploited the eye-tracking data of children with ASD for feature extraction. The features were fed into the SVM algorithm to predict attention and inattention. Using four significant steps depicted in Fig. 1, three attentional models: face-based, gaze-based, and hybrid-based (face and gaze), were developed to classify attention and inattention in both children with ASD and TD. In the first step, a series of attention task experiments was conducted to collect and annotate facial and eye-tracking data that describes attention. In the second step, feature extraction and selections were carried out to identify prominent features for training the attentional models. In the third step, the selected facial and gaze features were fed separately and combined into a binary classifier to train three types of attentional models: face-based, gaze-based, and hybrid-based. The six binary classifiers considered in this study are: SVM, RF, KNN, LR, CART and GBC. Lastly, the three attentional models developed: face-based, gaze-based, and hybrid-based, were evaluated and compared to identify the best attentional model for children with



Fig. 1 Methods of face, gaze and hybrid-based attentional models

ASD as well as its robustness for generalizing it across the autism spectrum. Also, the model for ASD and TD were compared.

4.1 Data collection and annotation

To collect facial and eye-tracking data describing attentional behavior, an experiment was conducted with children with both ASD and TD where they took a series of attention tasks. Prior to the experiment, approval was obtained from the institutional review board committee of the Qatar Biomedical Research Institute-Institutional Review Board approval. A total of forty-six children between the age of seven and eleven years participated in the study. Twenty children with ASD (ASD n = 20, M = 8.57, SD = 1.40) and twenty-six TD children from the same age range (TD n = 26, M = 8.58, SD = 1.36). The ASD group had sixteen boys and four girls with mild to moderate ASD, while the TD group had eighteen boys and eight girls, as shown in Table 1. The ASD participants were recruited through a local autism school in Doha and from known families residing in Qatar. All the ASD participants were clinically diagnosed by medical practitioners using the DSM IV-TR criteria (Huang et al. 2010). The TD participants were recruited from mainstream schools. All the parents of the participants read and signed an informed consent form to allow their children to participate in the study. Further experimental validation steps we took to achieve high data quality were to prevent the participants from eating or drinking during the experiment. The experiment was conducted in a quiet and dimly lit room to avoid distraction and reduce illumination, which may affect the eye-tracking sensor.

The attention tasks adopted in this study were based on the context of CPT (Rosvold et al. 1956). The CPT test is a computer-based assessment that presents stimuli to test users in a repetitive and boring pattern to measure how long they can maintain

Parameters	Description
f _{1n}	Frame by frame detection
f'1n	Frames annotated as attention
f'' _{1n}	Frames annotated as inattention
gf′ _{1n}	Geometrical information of attention
gf =	Geometrical information represented the mean value of landmark coordinates
$g{f^{\prime\prime}}_{1n}-g{f^\prime}_{1n}$	The difference between the mean value of attention and inattention frames

Table 1 Description of parameters used in the feature selection (Fig. 4)

their attention. The CPT attention tasks were simulated in a virtual classroom with distractions to represent a naturalistic learning environment (Parsons 2014; Rizzo et al. 2009). The target stimuli for the tasks are represented with random alphabets displayed on the board. The tasks had four levels of distractions: level 1-baseline (no distractions), level 2-minimal, level 3-medium, and level 4-extensive. The higher the level, the more the distractions. The distractions used include students raising hands, coughing, chatting, and dropping books. The essence of introducing distractions is to capture robust attentional behavior in different learning scenarios. The participants were instructed to press a clicker on the table when letter X appeared and ignore other letters. The number of correct clicks gives the performance score. A perfect performance score for each level was 40, as the letter X appeared 40 times, along with 214 random letters. Participants took a two-minute break after each attention task level. Two participants could not click and see the screen simultaneously; and thus the experiment was altered such that the letters were verbally called outto them as they randomly appeared. During the task, their facial and eye movements were captured in real-time with a Logitech C920 webcam attached to the top of the smaller screen in Fig. 2, and an eye-tracking device was placed at the base of the same screen. The real-time data capturing was monitored and recorded by the researcher on the bigger monitor (Fig. 3).



Fig. 2 Real-time eye and face tracking analysis

Fig. 3 Thirty-four facial landmarks with labels



After each participant completed all the attention tasks (levels 1 to 4), the recorded session for each level was divided into separate time slots for annotation. Each time slot consists of a random letter displayed for the duration of 1400 ms. To annotate the video, each slot was labeled as attention for valid response (letter X) or inattention for invalid response (missed letter X). In the ASD group, 18 out of 20 took all the attention tasks except for three who did not attempt level 4 as they wanted to end the experiment and 26 TD children completed level 1 of the attention task. Overall, 95 videos of the attention task sessions for all the participants were annotated and each video length was 5 min long.

4.2 Feature extraction and selection

This section discusses first, the feature extraction and selection methods for face tracking. Second, the feature extraction of the eye tracking data is also reported.

4.2.1 Facial features

In the facial feature extraction, 34 facial landmarks were extracted frame-by-frame from the recorded session for each participant through the webcam. These facial landmarks cover five regions of the face: eyes, eyebrows, nose, lips, and jaw, represented as a pool of feature vectors consisting of x and y coordinates represented in Eq. (1). Suppose f_n^i denotes each landmark in the nth video frame, starting with the ith frame.

$$f_n^i = \begin{bmatrix} x_{0,y_0}^i & x_{1,y_1}^i & \cdots & x_{33}^i y_{33}^i \\ \vdots & \vdots & \ddots & \vdots \\ x_{0,y_0}^n & x_{1,y_1}^n & \cdots & x_{33}^n y_{33}^n \end{bmatrix}$$
(1)

To explore the temporal variation of muscular activity across landmarks, we estimated the lengths of distances from all pair distances using the Euclidean distance formula in Eq. (2). This method is common in the literature for exploring differences in posed emotion and neutral face (Huang et al. 2010; Sariyanidi et al. 2014; Chu et al. 2018; Ghimire and Lee 2013).

$$[(x_1, y_1), (x_2, y_2)] = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$
(2)

where x_1 , y_1 and x_2 , y_2 are representing two different facial landmarks. The estimated geometrical information generated was between one landmark to other landmarks and these sum up to 561 geometric-based features distances. These geometric features were reduced from 561 to 20 features by applying the feature selection method (Jain and Singh 2018). Feature selection reduces training samples to those with the best features while maintaining the efficiency of the model. The objective of feature selection is to reduce computational costs. In this model, we selected the best features from the 561 pair distances. Data samples were normalized to ensure comparable data sample range. This normalization is achieved as follows (Eq. 3).

$$Z = \frac{x_i - \text{mean}(x)}{\text{stdev}(x)}$$
(3)

where Z is the standardized score, and stdev is the standard deviation of the data samples. In particular, the standardization subtracts the mean value of the samples and divides their value by the standard deviation.

0-Right Top Jaw, 1-Right Jaw Angle, 2-Gnathion, 3-Left Jaw Angle, 4-Left Top Jaw, 5-Outer Right Brow, 6-Right Brow Corner, 7-Inner Right Brow Corner, 8-Inner Left Brow Corner, 9-Left Brow Center, 10-Outer Left Brow Corner, 11-Nose Root, 12-Nose Tip, 13-Nose Lower Right Boundary, 14-Nose Bottom Boundary, 15-Nose Lower Left Boundary, 16-Outer Right Eye, 17-Inner Right Eye, 18-Inner Left Eye, 19-Outer Left Eye, 20-Right Lip Corner, 21-Right Apex Upper Lip, 22-Upper Lip Center, 23-Left Apex Upper Lip, 24-Left Lip Corner, 25-Left Edge Lower Lip, 26-Lower Lip Center, 27-Right Edge Lower Lip, 28-Bottom Lower Lip, 29-Top Lower Lip, 30-Upper Corner Right Eye, 31-Lower Corner Right Eye, 32-Upper Corner Left Eye, 33-Lower Corner Left Eye.

To explore the prominent facial features for differentiating attention and inattention, the geometrical information was estimated from all points pairwise using the Euclidean distance formula (Eq. 4) as follows

Euclidean distance =
$$\sqrt{((x^2 - x^1)^2 + (y^2 - y^1)^2)}$$
 (4)



Fig. 4 Facial feature selection process

where x_1 , y_1 , and x_2 , y_2 are representing two different facial landmarks.

To select the best feature from the facial features, threshold distance was used. The threshold distance is an estimation measurement that describes the changes between facial expression at a neutral frame and expression frame (Al-agha et al. 2017). The threshold distance value is an established method for revealing the information embedded in a dataset. This approach has been successfully applied in differentiating posed emotions from neutral emotions (Asteriadis et al. 2009; Hulse et al. 2012). In this current study, the threshold distance between attention and inattention with higher values represented in Fig. 4 were selected to train several binary classifier algorithms. The parameters describing the feature selection process are described in Table 1.

4.2.2 Eye-tracking features

The gaze-based attentional model consists of six primary eye-tracking features described in Table 1: gaze position, fixation position (FixationY, FixationX), FixationDuration, Ocular distance i.e., head distance to the screen (DistanceLeft, DistanceRight), pupil size (PupilLeft, PupilRight), and interocular distance were collected. The description of these features is provided in Table 2.

Next, we identified the annotated samples labeled as attention, inattention, and unknown. The samples with unknown labels were deleted, leaving us with only samples labeled as attention and inattention. The annotation column with string values: attention and inattention were converted to integers 1 and 0, respectively. Lastly, we normalized each feature to be on the same scale using the StandardScaler library in sci-kit-learn (Eq. 5) using the following equation to achieve a relatively normalized sample distribution (Table 3).

$$Z = \frac{x_i - \text{mean}(x)}{\text{stdev}(x)}$$
(5)

The best features were selected using an embedded method that uses the inherent characteristics of decision tree algorithms such as random forest, and CART (Guyon and Elisseeff 2003) and it is especially encouraged for imbalanced datasets (Liu et al. 2019). We applied an embedded feature selection method using feature permutations

Gaze features	Gaze sub features	Description
1. Pupil size	PupilLeft	The pupil size of the left eye
	PupilRight	The pupil size of the right eye
2. Ocular distance	DistanceLeft	The distance of the participant's left eye to screen
	DistanceRight	The distance of the participant's right eye to screen
3. Fixation duration	_	The duration of time participant spent looking at the stimuli
4. Fixation position	FixationX	The <i>x</i> -coordinate value of the position of the eye to the stimuli on the screen
	FixationY	The <i>y</i> -coordinate value of position of the eye to the stimuli on the screen
5. Gaze position	GazeLeftx	The <i>x</i> -coordinate value of the participant's gaze on the screen through the left eye
	GazeLefty	The <i>y</i> -coordinate value of the participant's gaze on the screen through the left eye
	GazeRightx	The <i>x</i> -coordinate value of the participant's gaze on the screen through the right eye
	GazeRighty	The <i>y</i> -coordinate value of the participant's gaze on the screen through the right eye
6. Interocular distance	-	The distance between the left pupil and the right pupil

Table 2 Description of gaze-based features

of randomly selected samples and by calculating the percentage increase in the misclassification rate to select the best individual eye-tracking features (Lu 2019). (Fig. 5).

4.3 Binary classification

Three different attentional models were developed using six different classifier algorithms: SVM, CART, LOR, RF, GBC, and KNN. These algorithms were implemented in Scikit-Learn. The performances of these models were compared using tenfold crossvalidation to select the best classifier. The performances of these algorithms were compared using tenfold cross-validation, and SVM outperformed the other classifiers with the highest AUC value. Next, we applied hyper-parameter tuning to optimize the SVM parameters; cost (C) and gamma (γ). A large value for C indicates the model will be stricter on classification errors. γ influences the sophistication of the decision boundary. Small values of γ will lead to an increasingly sophisticated boundary that correctly classifies a higher percentage of training data. Thus, inappropriate value selection for these parameters, also known as hyper-parameter tuning, may lead to the poor performance of a model on a new dataset (i.e., overfitting). The parameter values selected for C and γ are from the following sets of values C's = (Marcu, et al. 2013; Douglas et al. 2022; Wood et al. 2016; Lee and Meadan 2021; CDC 2022; Alshaban et al. 2019; Sharabi and Marom-Golan 2018; Russa et al. 2015; Kinnealey et al. 2012;

Name	Facial features				Gaze features			
	Total	Attention	Inattention	Unknown	Total	Attention	Inattention	Unknown
P1	35,886	974	173	34,739	68,871	79	137	68,655
P2	40,211	1,571	71	38,569	80,627	282	8	80,337
P3	38,917	920	1,669	36,328	61,565	760	273	60,532
P4	49,409	2,046	0	47,363	84,537	3,759	0	80,778
P5	48,391	1,437	67	46,857	80,621	224	14	80,383
P6	48,729	1,631	111	46,987	81,698	651	86	80,961
P7	41,631	226	209	41,196	68,176	156	3	68,017
P8	48,549	1,518	920	46,111	80,630	1,043	938	78,649
P9	I	I	I	I	I	I	I	Ι
P10	47,043	2,201	54	44,788	80,099	4,547	98	75,454
P11	48,267	2,385	7	45,875	80,767	8,303	23	72,441
P12	103,046	5,198	1,543	96,305	202,107	502		201,605
P13	53,646	1,291	76	52,279	80,240	411	9	79,823
P14	48,638	582	379	47,677	81,501	587	330	80,584
P15	39,053	1,369	411	37,273	62,670	714	203	61,753
P16	38,984	513	243	38,228	81,666	100	7	81,559
P17		I	I	I	I	Ι	I	I
P18	44,179	2,715	1,073	40,391	82,587	0	0	82,587
P19	38,602	46	384	38,172	60,553	0	10	60,543
P20	48,599	963	963	46,673	81,014	1,672	0	79,342

Table 3 datapoints generate ASD group in attention tasks level 1-4

271



Fig. 5 Eye-tracking feature selection process

Ridderinkhof et al. 2020; Almumen and Almuhareb 2020; Fredricks 2011; Dewan et al. 2019; Tsiourti et al. 2019; Rosenberg and Ekman 2020; Escobedo et al. 2014; Neuhaus et al. 2019; Hutt, et al. 2021; Palazzi et al. 2018; Chen et al. 2021; Jacob et al. 2019; Alvari et al. 2021; Carpenter et al. 2021; Thabtah and Peebles 2020; Zanker et al. 2019; Belle et al. 2011) and $\gamma = [0.001, 0.01, 0.1, 1, 10]$.

4.4 Attentional model

Three attentional models were developed using facial, gaze and hybrid (facial and gaze) features. Each model was trained as participant-specific and generalized. The participant-specific model was based on individual data while the generalized model was developed with the data from all the participants. In total, six attentional models were evaluated using AUC scores to determine their performance in detecting attention. The performance of the participant-dependent model shows a higher performance value than the participant-independent.

5 Results

This section presents the descriptive statistics attentional behavior of children with ASD and TD, results of the face-based, gazed-based and hybrid-based attentional models using six different classifier algorithms, followed by the features selection, participant-specific and generalized models. Among the six binary classifiers, SVM has the highest performance value of other classifiers. The facial and gaze data used for developing the attentional model were generated series from attention tasks with 46 participants (ASD = 20 and TD = 26). The attention tasks consist of four different levels and they include: baseline, easy, medium, and hard. The baseline comprises static social and nonsocial distraction; the easy level presented the target stimuli with fewer dynamic distractions, the easy level with more dynamic distractions, and the hard level with the highest distractions. The degree of distractions levels is to obtain robust forms of attentional behavior. However, the data used for th attentional model were from.

5.1 Descriptive statistics of attentional data in ASD and TD participants

The descriptive statistics of attention and inattention data for both facial and eye gaze features from ASD group for all the attention task (levels 1–4) represented in Table 4 and Fig. 6. Next, we present the differences between the two groups (ASD and TD)

Table 4 Descriptive statistics of facial and gaze data in ASD		Ν	Mean	SD
group (Attention task levels 1–4)	Inattention(Facial)	17	437.353	532.731
	Inattention(Gaze)	13	163.538	257.140
	Attention(Facial)	17	1507.118	1191.492
	Attention(Gaze)	13	1373.615	2386.929
800	3000			
Inattention(Facial)	Inattention(Gaze)	Att	ention(Facial)	Attention(Gaze)
(a)Inattentio	on Data		(b)Attention	Data

Fig. 6 Descritive plots for attention and inattention data generated from facial and gaze features

Measure 1	Measure 2	t	df	р
Inattention (Facial)	Inattention (Gaze)	1.368	12	0.196
Attention (Facial)	Attention (Gaze)	0.135	12	0.895

and within each group using independent samples t-test and paired t-test respectively. Despite the difference in the facial and gaze data there is no significant difference between the two data types as shown in Table 5 which shows the performance of the facial and gaze-based model can be compared. The descriptives for facial and gaze features for ASD and TD groups in Table 6 and Fig. 7 show the variation between the data from the two groups. The independent t-test between data from the two groups (Table 7) shows that there is no significant difference between the group except for the inattention data types where ASD group has significant more data sample than the TD group.

5.2 Face-based attentional model

Table 5 Paired samples T-test

Both participant-specific and generalized models were evaluated using ROC-AUC scores. The performance of the participant-specific model shows a higher performance value than the generalized.

Parameters	Group	Ν	Mean	SD
Attention (Facial)	ASD	17	503.00	414.831
	TD	24	396.46	93.535
Inattention (Facial)	ASD	17	222.41	325.473
	TD	24	6.71	9.844
Attention (Gaze)	ASD	13	478.77	786.763
	TD	24	785.58	479.007
Inattention (Gaze)	ASD	13	110.85	206.617
	TD	24	12.96	21.574

Table 6 Descriptives for facial and gaze features for ASD and TD groups (Attention task level 1 only)



Fig. 7 Descritive plots for attention and inattention facial data extracted from ASD and TD groups

Table 7 Independent samples T-test for facial and gaze data		t	df	р
between ASD and TD groups	Attention (Facial)	1.040	17.158	0.313
	Inattention (Facial)	2.732	16.021	0.015
	Attention (Gaze)	-1.171	16.940	0.258
	Inattention (Gaze)	1.713	12.133	0.112

Welch's t-test

In the evaluation phase, we compared the performances of models developed with different sets of distance-based features (i.e.,1–10, 1–20, and 1–30 features). The model with 10 features and 30 features had an accuracy score of 0.873 and that of 20 features had an accuracy of 0.889 as shown in Fig. 8. Thus, this study used a model with the best 20 distance-based features due to higher performance and fewer features. Similarly, the SVM model had the best AUC score across other models depicted in Fig. 9. Also, Table 8 describes the 20 distance-based features. Five face regions emerged as prominent facial features in recognizing attention—left jaw, right brow, right eyes, nose, and gnathion (chin).



Fig. 8 Comparison of best features and classifier algorithms



Fig. 9 Model comparison with Best 20 features

Features	Feature description	Inattention (mean values)	Attention (mean values)	Distance threshold values (mm)
D: 3–15	Left jaw angle-outer right brow corner	171.45	146.9	24.55
D: 4–5	Left top jaw- outer right brow corner	168.45	144.06	24.38
D: 4–6	Left top jaw- right brow center	149.17	125.72	23.45
D: 3–6	Left jaw angle- right brow center	158.32	135.01	23.30
D: 4–16	Left top jaw-outer right eye	148.79	126.22	22.56
D: 4–7	Left top jaw- inner right brow corner	122.89	100.99	21.89
D: 4–31	Left top jaw- lower corner right eye	132.7	110.85	21.85
D: 4–30	Left top jaw- upper corner right eye	134.16	112.37	21.79
D: 0–4	Gnathion-outer right brow corner	167.05	145.36	21.69
D: 3–16	Left jaw angle-outer right eye	147.11	125.43	21.67
D: 2_5	Gnathion-outer right brow corner	159.94	138.36	21.58
D: 3_7	Left jaw angle- inner right brow corner	136.43	114.93	21.49
D: 3_31	Left jaw angle- lower corner right eye	131.48	110.45	21.02
D: 3_30	Left jaw angle- upper corner right eye	136.74	115.77	20.96
D: 4_13	Left top jaw-nose lower right boundary	111.63	90.99	20.63
D: 4_12	Left top jaw- nose tip	95.09	74.58	20.51
D: 4_17	Left top jaw- inner right eye	117.02	96.6	20.42
D: 4_11	Left top jaw-nose root	98.58	78.51	20.06
D: 2_6	Gnathion- right brow center	154.53	134.66	19.87
D: 3_17	Left jaw angle- inner right eye	121.38	101.61	19.77



Fig. 10 Average performance of participant-specific model

5.2.2 Generalized model

In this model evaluation, the attentional model used data from 17 participants for training. Consequently, the model was tested on one participant who was not part of the training data. The average of the model performance for all participants was slightly above chance (ROC-AUC = 0.541), as shown in Fig. 10. This result implies that generalized performance is weak for a generalized attentional model among children with ASD.

5.2.3 Participant-specific model

The attentional model was trained and tested only on the data samples from each participant in the model. The training and testing data were in the ratio of 80% and 20%. The average model performance for all participants was above chance (ROC-AUC = 0.957), as shown in Fig. 11. This result illustrates that the performance of the participant-specific model was better than that of the generalized model. This finding is supported by the heterogeneity in children with ASD, where each child responds differently to a stimulus (Wetherby and Prizant 2000).

5.3 Gaze-based attentional model

This section presents the results of gazed-based attentional model using six different classifier algorithms, followed by the features selection, participant-specific and generalized models.

5.3.1 Evaluation of gaze-based model

The evaluation metrics (ROC-AUC) of tenfold cross-validation for six different classifiers: KNN, CART, GBC, SVM, LOR, and RF, are shown in Fig. 10. These classifiers were trained on the 12 eye-tracking features described in Table 1. The best model classifier was SVM(AUC = 0.990), followed by RF (AUC = 0.989). Lastly, in the feature



Fig. 11 Average performance of the generalized model

selection process, the PupilLeft emerged as the best feature among the twelve features, as shown in Fig. 12. It was determined that PupilLeft is the most significant feature, and GazeRightx is the least relevant feature. The combination of these eye-tracking features was evaluated for both the specific and generalized models in their order of importance. For example, the first feature, PupilLeft, was evaluated, then PupilLeft and DistanceLeft were evaluated, and so on until we completed all 12 features. The result shows that the specific model (AUC = 0.998) had better performance than the generalized model (AUC = 0.626). The result also shows that the combination of the first four features (PupilLeft, DistanceLeft, FixationDuration, and FixationX) achieves the best performance, as shown from Fig. 13.



Fig. 12 Model selections for attentional model



Fig. 13 Eye-tracking features in incremental order for specific and generalized models

The model performance of generalized and participant-specific models of the gazebased attentional model was compared among ASD and TD groups. The participantspecific had higher performance than the generalized model in ASD for all feature combinations. In contrast, the models for the TD group had an equal performance with the first features, but the participant-specific model had higher performance when the features were less than four.

5.3.2 Generalized model

This section presents the result of the SVM that was conducted using the best four features: PupilLeft, DistanceLeft, FixationDuration, and FixationX. We trained the SVM classifier with samples from all participants except one, which was used for testing. This training and testing approach was repeated for all participants and averaged. Five participants with only one class label i.e., only attention or inattention were exempted from testing. Using the average score from all participants, the model performance was slightly above chance (ROC-AUC = 0.626), as shown in Fig. 14.



Fig. 14 Average generalized model performances with four best eye-tracking features



Fig. 15 Average participant-specific model performances with four best eye-tracking features

5.3.3 Participant-specific model

Unlike the generalized model, the participant-specific model was only trained and tested on samples from each participant using the splitting ratio of 80% to 20%, respectively. The model performance from each participant was averaged for all participants. The average model performance was above chance (AUC = 0.998), as shown in Fig. 15. This result is far higher than the chance level, which implies that attention recognition among children with ASD can be achieved using personalized eye-tracking features.

5.3.4 Comparison of gaze-based attentional model in ASD and TD groups

The performance of the generalized and specific gaze-based attentional model was compared between the ASD and TD groups. We evaluated the incremental combination of all the features starting with combining the most relevant feature, PupilLeft, with the other 11 features. In the ASD group, the performances of the specific models are all higher than the generalized models. In comparison, in the TD group, the performance of the specific models was only higher for the first three feature combinations and was equal for the other feature combinations, as shown in Figs. 16 and 17.

5.4 Hybrid-based attentional model

This section presents the results of a hybrid-based model for participant-specific and generalized models. Also, we compared the hybrid-based, gazed-based, and face-based models across the participant groups. The results show that the hybrid-based participant-specific model performed better than the generalized model. The comparison across the three models (hybrid-based, gazed-based, and face-based) showed that the gazed-based model performed better than the face-based model in the ASD group. In contrast, the face-based model performed better with the TD group.



Fig. 16 Gaze-based model performances for generalized and specific models with children with ASD



Fig. 17 Model performances for generalized and specific models with TD children

5.4.1 Generalized model performance

This section presents the results of the hybrid-based model using the best four eyetracking features and 20 best twenty face-tracking features. Using the leave-one-out evaluation model, the average score from ASD participants gave a performance value close to chance level (ROC-AUC = 0.538), as shown in Fig. 18. The k-fold metric used a higher percentage as a testing sample, where the training and testing data were in the ratios of 78% and 22%. This evaluation was iterated 50 times to ensure performance reliability, and the average result was a little above chance as well (ROC-AUC = 0.561). This result implies that the attentional behavior is diverse and does not have a regular pattern.

5.4.2 Participant-specific model performance

The participant-specific model was trained on 80% of the data samples and tested with the remaining 20% from each participant. The model performance averaged across all participants was above chance (AUC = 0.996), as shown in Fig. 19. This result is far higher than the chance level, which implies that attention recognition among children with ASD can be achieved with personalized eye-tracking features.



Fig. 18 Hybrid-based model for generalized model (ASD)



Fig. 19 Hybrid-based model for participant-specific model (ASD)

5.5 Confusion matrix metrics for three models

This section presents the model evaluation of the attentional model types developed in the current study with using confusion matrix to compare facial and gaze features from children with ASD. The model evaluation for participant 1 is used as a sample to illustrates the model performance. Figure 20 represents the generalized and personalized models for face and gaze–based models.

5.5.1 Comparison of attentional model types in ASD and TD groups

The comparison of attentional model types between ASD and TD groups was evaluated across the face-based, gaze-based, and hybrid-based models using data from attention task level 1. Figure 21 shows that the generalized model in all model types, face-based,



Fig. 20 Confusion matrix evaluatio metric for P1



Fig. 21 Comparisons of model types in ASD and TD groups

gaze-based, and hybrid-based, works for the TD group more than the ASD group. In the ASD group, the gaze-based model had the highest performance score (AUC = 0.677) when compared to the face-based (AUC = 0.535) and hybrid-based models (AUC = 0.641). In contrast, the hybrid-based model performed better (AUC = 0.963) than the gaze-based (AUC = 0.961) and face-based (AUC = 0.958) models in TD groups. When using participant-specific models for ASD, the model performance was higher than that of the TD group. In the ASD group, the gaze-based model had the highest performance score (AUC = 0.998) when compared to face-based (AUC = 0.957) and hybrid-based models (AUC = 0.996). Conversely, the gaze-based (AUC = 0.950) performed better the hybrid-based model (AUC = 0.943) and face-based models (AUC = 0.934) for TD group.

6 Discussion

6.1 Gaze-based attentional model

This study conducted an attention task experiment with children with ASD to generate eye-tracking features. The feature extraction was based on content and physiologicalbased data that describes attentional behavior during an attention task. The extracted features were trained with the classifier algorithm to recognize attention and inattention behaviors. Two main types of models, participant-specific and generalized were evaluated between the ASD and TD groups. Results show that the participant-specific model performed better than the generalized model for the ASD group while both models had a similar performance for the TD group with the best four features. This finding aligns with the concept of heterogeneity among children with ASD (Wetherby and Prizant 2000).

The best features identified for the gaze-based attentional model using embedded feature selection include a combination of PupilLeft, DistanceLeft, Fixation Duration, and FixationX. These features showed that the two eye works differently during attention tasks. This finding supports the evidence of side glancing in children with ASD, a characteristic where they tend to look off to the side of a stimulus (Noris et al. 2012; Little 2018). Furthermore, the results of this study show that pupil size is the most significant feature for detecting attention.

Our approach to developing a gaze-based attentional model for children with ASD has added to the body of knowledge with two main contributions. First, this study showed that the personalized gaze-based attentional model is more suitable for children with ASD than the generalized model. Second, we identified that the pupil size of the left eye is the most significant feature for developing gaze-based attentional model for children with ASD.

6.2 Face-based attentional model

The face-based attentional model discussed in this study utilizes 34 facial landmarks with x and y coordinates generated in real-time during the attention task. Consequently,

the features were transformed to distance between facial landmarks. Twenty distancebased features were identified as distinctive features for differentiating attention and inattention. Also, the best facial features for recognizing attention were identified based on the distance-based features—jaw, eyebrows, eyes, nose, and gnathion. Lastly, we evaluated model generalization for ASD and TD groups and different attention tasks. According to the results of this study, the performance of the participant-specific and the generalized was above chance. However, the performance of the participantspecific model had a better performance score than the generalized model.

This shows that the participant-specific model works better for children with ASD. Similarly, previous studies also concluded that children with ASD exhibit different face-based attentional behaviors (Bieberich and Morgan 2004; Czapinski and Bryson 2003). Findings from this study show that the face-based attentional model is relatively more generalized among the TD group than in the ASD group. Overall, the participant-specific model outperformed the generalized model. Therefore, a generalized face-based model for children with ASD and different attention tasks needs to be applied cautiously.

The implication of the face-based attentional model for children with ASD showed that the personalized model supports the evidence of heterogeneity in individuals with ASD. This confirms variations in facial features describing attentional behavior in children with ASD. Thus, our result suggests a personalized face-based attentional model instead of the traditional one-size-fits-all machine learning approach. A similar finding was reported by Rudovic and Lee (2018) where the authors evaluated personalized and generalized deep learning models to detect affective states and engagement during human–robot interaction with children with ASD. Their finding showed that the personalized model outperformed the generalized model due to the cultural and individual differences among the participants. Furthermore, the direction of recent research, especially in the ASD field, is heading down the path of personalized and precise intervention for children with ASD (Stevens et al. 2018) personalized and precise intervention for children with ASD (Stevens et al. 2019), and personalized medicine in ASD. The trend of the personalized model is not limited to education and medicine but also to commerce and many other fields.

The application of a personalized face-based attentional model can enhance adaptive learning support or human intervention support. The findings from this study support the evidence that facial features have the potential of assessing attention (Nezami 2019; Dubbaka and Gopalan 2020). The three contributions of this study include 1) a distance-based feature selection method that differentiates attentional behaviors, 2) five face regions that best describe attentional behaviors among children with ASD, and 3) how the face-based attentional model supports the evidence of personalized attentional model than the generalized model.

6.3 Hybrid-based attentional model

The hybrid-based attentional model consists of both face-tracking and eye-tracking datawhich were feed into SVM classifier algorithm to predict attentional behaviors. The two main types of models, participant-specific, and generalized models were

evaluated between the ASD and TD groups. The results show that all attentional model types, hybrid-based, face-based, and gaze-based models, cannot be generalized in children with ASD. This finding aligns with the concept of heterogeneity among children with ASD (Wetherby and Prizant 2000). Also, eye-tracking measures aid in the detection of more attentional behaviors than face-tracking measures or the combination of eye-tracking and face-tracking measures. However, it is worth mentioning that our study has showed that face-tracking measure can also be sufficient for attention recognition.

There is a widespread assumption that models with multiple-sensor or multimodal methods give better accuracy than models with a few sensors or single-sensor methods (Aslan, et al. 2014; Asteriadis et al. 2009; Shaker et al. 2013). A previous study on this assumption shows that the multimodal approach is not always the best. According to D'mello and Graesser (D'mello and Graesser 2010), a multimodal approach shows that integrating facial features, body posture, and interaction dialog, had similar performance when compared with combining facial features, body posture and interactive dialog. Similarly, our results show that combining many methods does not guarantee better accuracy. For example, we expected the performance of the hybrid-based attentional model to be higher than the gaze-based or face-based models. The hybrid-based model performed better than other models in the TD group but not in the ASD groups. Instead, the best model type for the ASD group was the gaze-based model.

7 Practical implication

The practical implication of this study will benefit teachers and parents in saving their time while assessing the attention of their students and children respectively. Importantly, aside from providing the stakeholder the attention duration of children with ASD, it will also inform them of the classroom stimuli types that distracts their attention. Recently we have used the model developed in the current study to develop a web-based application platform for personalized engagement assessment using an integrated webcam for face-tracking and eye-tracking. The application has four modules. In the first module, we presented attention tasks for the initial calibration of personalized behavior during learning engagement. The attention task consisted of target stimuli (social and nonsocial visual and auditory) simulating typical classroom distractions such as social, nonsocial, audio, and audiovisual distractions. The distraction levels consists of two different levels (baseline and hard) to capture the degree of the engagement level of each user. The second module will consist of feature extraction from facial and gaze features during the attention task to train the face-based machine learning model and capture their gaze pattern. The third module has an intelligent components which consists of machine learning models that detect when the user pays attention to the target stimuli in the attention task and when they are inattentive. In the fourth module, the personalized attentional report is generated for the satekholdershwing the duration of attention throughout the learning duration as well as the gaze patterns of each student.

Despite the strength of this study, it still has several limitations. First, attentional behavior of children with severe ASD was not considered as they were not included in the data collection process because it is challenging to get sit for the attention task. Therefore, the application of this attentional model may not be generalized to children with severe ASD or attention deficit hyperactive disorder (ADHD) since neither of these variables was incorporated in the experiment. Second, the experiments were tailored to the experience of children with ASD and not adults, so our findings may not apply to techniques of measuring attention in adults with ASD. Third, not all possible eye-tracking measures were covered in this study. For example, saccades, smooth pursuit, and blink rate. There is a need for further studies on how other eyetracking metrics can influence the development of an attention recognition system for children with ASD. Four, the sample size used in this study is limited as it is challenging to get the consent of parents due to the fear of unknown reactions from their children. Lastly, the facial feature extraction was primarily done on data from children with ASD and those features were applied on facial data from TD children. Therefore, the attentional model may not give optimal performance with TD children.

9 Conclusion

This study has shown the potential of unobtrusive and non-invasive sensing technology such as webcam and eye-tracking devices for extracting facial and gaze features that best describe the attentional behavior of children with ASD. Thus, designers or researchers can either use a webcam or eye-tracking device as a tool to track the attention of children with ASD. This objective approach is a step toward solving problems of attention assessment in children with ASD. The significant advantage of the objective attention detection system over subjective assessment is the provision for individualized attention assessment which supports the heterogeneity in ASD. Further improvement in the study can include the investigation of behavioral features that describes auditory attention as facial and eye-tracking features that might be limited to visual attention. Also, this work can be extended to other populations with learning disabilities such as dyslexia, down syndrome, and others to achieve desired learning outcomes.

Author contributions Conceptualization, B.B. and D.A.; methodology, M.Q., and D.A.; validation, B.B., D.A., and M.Q.; formal analysis, M.Q., D.A., and B.B.; investigation, B.B.; resources, B.B.; data curation, B.B., M.Q., and D.A.; writing—original draft preparation, B.B.; writing—review and editing, B.B., D.A., and M.Q.; visualization, B.B., D.A., and M.Q.; supervision, D.A. and M.Q.

Funding Open Access funding provided by the Qatar National Library.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Al-agha, L.S.A., Saleh, P.H.H., Ghani, P.R.F.: Geometric-based feature extraction and classification for emotion expressions of 3D video film. J. Adv. Inf. Technol. 8(2), 74–79 (2017)
- Almumen, H., Almuhareb, K.: Technology-enriched social story intervention: engaging children with autism spectrum disorders in social communication. Int. J. Early Childhood Spl. Educ. 12(1), 96–119 (2020)
- Alshaban, F., et al.: Prevalence and correlates of autism spectrum disorder in Qatar: a national study. J. Child Psychol. Psychiatry 60(12), 1254–1268 (2019)
- Alvari, G., Coviello, L., Furlanello, C.: EYE-C: eye-contact robust detection and analysis during unconstrained child-therapist interactions in the clinical setting of autism spectrum disorders. Brain Sci. 11(12), 1555 (2021)
- Aslan, S., et al.: Learner engagement measurement and classification in 1: 1 learning. In: 2014 13th International Conference on Machine Learning and Applications. 2014. IEEE
- Asteriadis, S., et al.: Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment. Multimedia Tools Appl. **41**(3), 469–493 (2009)
- Banire, B., Jomhari, N., Ahmad, R.: Visual hybrid development learning system (VHDLS) framework for children with autism. J. Autism Dev. Disord. 45(10), 3069–3084 (2015)
- Belle, A., Hobson, R., Najarian, K.: A physiological signal processing system for optimal engagement and attention detection. In: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). 2011.
- Bieberich, A.A., Morgan, S.B.: Self-regulation and affective expression during play in children with autism or Down syndrome: a short-term longitudinal study. J. Autism Dev. Disord. 34(4), 439–448 (2004)
- Bixler, R., D'Mello S.: Automatic gaze-based detection of mind wandering with metacognitive awareness. In: User Modeling, Adaptation, and Personalization. 2015. Springer
- Bosch, N., et al.: Using video to automatically detect learner affect in computer-enabled classrooms. ACM Transactions on Interactive Intelligent Systems (TiiS) 6(2), 17 (2016)
- Carpenter, K.L., et al.: Digital behavioral phenotyping detects atypical pattern of facial expression in toddlers with autism. Autism Res. **14**(3), 488–499 (2021)
- CDC. Center for Disesase Control and Prevention, Data & Statistics on Autism Spectrum Disorder. 2022 [cited 2022 16 December, 2022]; Available from: https://www.cdc.gov/ncbdd/autism/data.html.
- Chen, I.Y., et al.: Probabilistic machine learning for healthcare. Annu. Rev. Biomed. Data Sci. 4, 393–415 (2021)
- Chen, Chen, P.-C., Tsai, Y.-T.: Attention estimation system via smart glasses. In: Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2017 IEEE Conference on2017, IEEE. pp 1–5
- Chu, H.-C., et al.: Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. Soft. Comput. 22(9), 2973–2999 (2018)
- Czapinski, P., Bryson, S.: Reduced facial muscle movements in Autism: Evidence for dysfunction in the neuromuscular pathway? Brain Cogn. 51(2), 177–179 (2003)
- D'mello, S.K., Graesser, A.: Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. User Model. User-Adapt. Interact. 20(2), 147–187 (2010)
- Dekhil, O., et al.: Using resting state functional MRI to build a personalized autism diagnosis system. PLoS ONE **13**(10), e0206351 (2018)
- Dewan, M.A.A., Murshed, M., Lin, F.: Engagement detection in online learning: a review. Smart Learn. Environ. 6(1), 1 (2019)

- Douglas, S.N., et al.: Validation of wearable sensor technology to measure social proximity of young children with autism spectrum disorders. Focus Autism Other Deve Disabil. **37**(1), 24–33 (2022)
- Dubbaka, A., Gopalan, A.: Detecting learner engagement in MOOCs using automatic facial expression recognition. In: 2020 IEEE Global Engineering Education Conference (EDUCON). 2020. IEEE
- Egger, H.L., et al.: Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study. NPJ Digit. Med. **1**(1), 1–10 (2018)
- Escobedo, L., et al.: Using augmented reality to help children with autism stay focused. IEEE Pervasive Comput. **13**(1), 38–46 (2014)
- Fredricks, J.A.: Engagement in school and out-of-school contexts: a multidimensional view of engagement. Theory Into Practice 50(4), 327–335 (2011)
- Ghassemi, F., et al.: Classification of sustained attention level based on morphological features of EEG's independent components. In: ICME International Conference on Complex Medical Engineering, 2009. CME. 2009. IEEE
- Ghimire, D., Lee, J.: Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. Sensors 13(6), 7714–7734 (2013)
- Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
- Higuchi, K., et al.: Visualizing gaze direction to support video coding of social attention for children with autism spectrum disorder. In: 23rd International Conference on Intelligent User Interfaces. 2018
- Huang, K.-C., Huang, S.-Y., Kuo, Y.-H.: Emotion recognition based on a novel triangular facial feature extraction method. In: The 2010 International Joint Conference on Neural Networks (IJCNN). 2010. IEEE
- Hutt, S., et al.: Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021
- Jacob, S., et al.: Neurodevelopmental heterogeneity and computational approaches for understanding autism. Transl. Psychiatry 9(1), 1–12 (2019)
- Jain, D., Singh, V.: Feature selection and classification systems for chronic disease prediction: a review. Egypt. Inf. J. 19(3), 179–189 (2018)
- Kinnealey, M., et al.: Effect of classroom modification on attention and engagement of students with autism or dyspraxia. Am. J. Occup. Ther. 66(5), 511–519 (2012)
- Lallé, S., Conati, C., Azevedo, R.: Prediction of student achievement goals and emotion valence during interaction with pedagogical agents. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. 2018
- Lee, J.D., Meadan, H.: Parent-mediated interventions for children with ASD in low-resource settings: a scoping review. Rev. J. Autism Dev. Disord. 8(3), 285–298 (2021)
- Lee, K., Schertz, H.H.: Brief report: Analysis of the relationship between turn taking and joint attention for toddlers with autism. J. Autism Dev. Disord. 50(7), 2633–2640 (2020)
- Little, J.A.: Vision in children with autism spectrum disorder: a critical review. Clin. Exp. Optom. **101**(4), 504–513 (2018)
- Liu, H., Zhou, M., Liu, Q.: An embedded feature selection method for imbalanced data classification. IEEE/CAA J. Automat. Sin. 6(3), 703–715 (2019)
- Lu, M.: Embedded feature selection accounting for unknown data heterogeneity. Expert Syst. Appl. 119, 350–361 (2019)
- Mamun, M.I., et al.: Autilife: a healthcare monitoring system for autism center in 5g cellular network using machine learning approach. In: 2019 IEEE 17th international conference on industrial informatics (INDIN). 2019. IEEE
- Marcu, G., et al.: Why do they still use paper? Understanding data collection and use in Autism education. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013
- Monkaresi, H., et al.: Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Trans. Affect. Comput. 8(1), 15–28 (2016)
- Neuhaus, E., Webb, S.J., Bernier, R.A.: Linking social motivation with social skill: the role of emotion dysregulation in autism spectrum disorder. Dev. Psychopathol. 31(3), 931–943 (2019)
- Nezami, O.M., et al.: Automatic recognition of student engagement using deep learning and facial expression. In: Joint european conference on machine learning and knowledge discovery in databases. 2019. Springer

- Noris, B., et al.: Investigating gaze of children with ASD in naturalistic settings. PLoS ONE **7**(9), e44144 (2012)
- Palazzi, A., et al.: Predicting the driver's focus of attention: the DR (eye) VE project. IEEE Trans. Pattern Anal. Mach. Intell. 41(7), 1720–1733 (2018)
- Parsons, T.D.: Virtual teacher and classroom for assessment of neurodevelopmental disorders. In: Brooks, A.L., Brahnam, S., Jain, L.C. (eds.) Technologies of Inclusive Well-Being: Serious Games, Alternative Realities, and Play Therapy, pp. 119–135. Springer, Berlin (2014)
- Ridderinkhof, A., et al.: Attention in children with autism spectrum disorder and the effects of a mindfulnessbased program. J. Atten. Disord. 24(5), 681–692 (2020)
- Rizzo, A.A., et al.: A virtual reality scenario for all seasons: the virtual classroom. CNS Spectr. 11(1), 35–44 (2009)
- Rosenberg, E.L., Ekman, P.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, Oxford (2020)
- Rosvold, H.E., et al.: A continuous performance test of brain damage. J. Consult. Psychol. 20(5), 343 (1956)
- Rudovic, O., et al.: Personalized machine learning for robot perception of affect and engagement in autism therapy. Sci. Robot. (2018). https://doi.org/10.1126/scirobotics.aao6760
- Russa, M.B., Matthews, A.L., Owen-DeSchryver, J.S.: Expanding supports to improve the lives of families of children with autism spectrum disorder. J. Posit. Behav. Interv. 17(2), 95–104 (2015)
- Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic analysis of facial affect: a survey of registration, representation, and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37(6), 1113–1133 (2014)
- Shaker, N., et al.: Fusing visual and behavioral cues for modeling user experience in games. IEEE Trans. Cybern. 43(6), 1519–1531 (2013)
- Sharabi, A., Marom-Golan, D.: Social support, education levels, and parents' involvement: a comparison between mothers and fathers of young children with autism spectrum disorder. Topics Early Childhood Spl. Educ. 38(1), 54–64 (2018)
- Shojaeizadeh, M., et al.: Detecting task demand via an eye tracking machine learning system. Decis. Support Syst. 116, 91–101 (2019)
- Sonne, T., Obel, C., Grønbæk, K.: Designing real time assistive technologies: a study of children with ADHD. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction. 2015. ACM
- Stevens, E., et al.: Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning. Int. J. Med. Inf. 129, 29–36 (2019)
- Thabtah, F., Peebles, D.: A new machine learning model based on induction of rules for autism detection. Health Inform. J. 26(1), 264–286 (2020)
- Tsiourti, C., et al.: Multimodal integration of emotional signals from voice, body, and context: effects of (in) congruence on emotion recognition and attitudes towards robots. Int. J. Soc. Robot. **11**(4), 555–573 (2019)
- Van Hulse, J., et al.: Threshold-based feature selection techniques for high-dimensional bioinformatics data. Netw. Model. Anal. Health Inf. Bioinf. 1(1–2), 47–61 (2012)
- Wetherby, A.M., Prizant, B.M.: Autism Spectrum Disorders: A Transactional Developmental Perspective, vol. 9. Brookes Pub, Baltimore (2000)
- Whitehill, J., et al.: The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Trans. Affect. Comput. 5(1), 86–98 (2014)
- Wood, B.K., et al.: Comparison of observational methods and their relation to ratings of engagement in young children. Topics Early Childhood Spl. Educ. 35(4), 211–222 (2016)
- Yaneva, V., et al.: Detecting high-functioning autism in adults using eye tracking and machine learning. IEEE Trans. Neural Syst. Rehabil. Eng. 28, 1254–1261 (2020)
- Yang, S.-M., Chen, C.-M., Y, C.-M.: Assessing the attention levels of students by using a novel attention aware system based on brainwave signals. In 2015 IIAI 4th International Congress on Advanced Applied Informatics (IIAI-AAI). 2015. IEEE
- Yulita, I.N., Fanany, M.I., Arymurthy, A.M.: Comparing classification via regression and random committee for automatic sleep stage classification in autism patients. J. Phys. Conf. Series. **1230**, 012010 (2019)
- Zanker, M., Rook, L., Jannach, D.: Measuring the impact of online personalisation: past, present and future. Int. J. Hum Comput Stud. 131, 160–168 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. **Bilikis Banire** is currently a Postdoctoral Fellow at the Faculty of Medicine, Department of Psychiatry, and Faculty of Computer Science at Dalhousie University, Canada, She received her M.Sc. (Hons.) degree in Software Engineering from the University of Malaya, in 2014, and her Ph.D. degree in Computer Science and Engineering at Hamad Bin Khalifa University, Qatar, in 2021. Her research areas cover persuasive and personalized application design to support children and adolescents with mental health disorders or autism using objective methods such as eye-tracking and affective computing.

Dena Al-Thani is currently an associate professor at the College of Science and Engineering and the Director of interdisciplinary graduate programs, HBKU. In addition, she is a part-time consultant for Research and Innovation in Mada Assistive Technology Centre, where she leads several projects in assistive technology. She received the M.Sc. (Hons.) degree in software engineering from the University of London, in 2009, and the Ph.D. degree in computer science, in 2016. Her academic and research vocation is to explore and demonstrate how HCI as a field of applied enquiry can contribute to building a more inclusive society.

Dr. Marwa Qaraqe is an Associate Professor at HBKU and she received her Ph.D. and Master of Science degree in Electrical Engineering from Texas A&M University in College Station, Texas, USA. Her research interests lie in the field of predictive analytics, especially in the area of machine learning and its application in the biomedical field, neurological disorders, bodyarea-networks, wearable body sensors, advanced signal and image processing of biomedical signals/images, and in the areas of information and communication technologies for health care. In addition to winning several awards throughout her career, Dr. Qaraqe has published one book and several papers in several premium peer-reviewed conferences and journals.