Turkish Journal of Psychology, November 2024, *39*, 20-40 DOI: 10.31828/turkpsikoloji.1515430

Summary
Machine Learning in Classification of Autism Spectrum Disorder: Visual Attention Directed to the Emotional States

Işık Akin-Bülbül[1] İbrahim Kök[2] Selda Özdemir[3]
Hacettepe University Ankara University Hacettepe University

Facial emotion recognition, a component of social cognition, is a fundamental skill for effective social communication and interaction (Adolphs, 2002; Wang et al., 2004). The development of social cognition starts early in typically developing (TD) infants, and infants have a strong visual preference for people and their faces from the earliest developmental periods compared to other visual stimuli. In addition, babies can distinguish qualitative differences in the emotions displayed on people's faces from an early age (Begeer, Koot, Rieffe, Terwogt, and Stegge, 2008; Schultz, 2005). When TD babies focus on the internal features of a face, they have a strong preference for the eyes area, and TD babies receive social information about external events (e.g., speech) and internal processes (e.g., emotions) from the faces of their communication partners (Sterling et al., 2008).

Children with autism are characterized by diminished eye contact and diminished attention to human faces since the date they were first described by Kanner (1943) in the literature. Currently, similar diagnostic criteria to the first autism definition have been utilized in autism diagnosis. Diagnosing autism and improving observation protocols related to early markers of autism has been a difficult task that requires consistent effort and time (Crane et al., 2016; Vabalas et al., 2020; Zhao et al., 2021). Although the physiological and clinical features accompanying autism cannot be determined based on behavioral observation from the earliest period, many research findings showed that some early markers are highly effective in determining autism (Frazier et al., 2018; Schaller, et al., 2021). Therefore, more objective, accessible, and faster screening methods are needed, especially in young children, to improve the early autism screenings (Bolte et al., 2016; Elsabbagh and Johnson, 2016). The physiological evidence gathered from visual attention parameters of children with autism can be used as promising biomarkers in autism screening practices.

In various current studies, eye tracking and machine learning algorithms were successfully applied to the classification of individuals with autism (Canavan et al., 2017; Kennedy ve Adolphs, 2012; Liu, Li, and Yi, 2016; Özdemir, Akın-Bülbül, Kök, Özdemir, 2022). An accurate algorithm that combines visual attention patterns directed to social stimuli can provide a more reliable and faster screening by gathering physiological data that are not possible with the human eye (Canavan et al., 2017). Within the scope of this research, we plan to take the above literature discussion one step further. We propose using machine learning algorithms to investigate the visual attention data gathered watching emotions to distinguish children with autism from their TD peers. Thus, this study examines the predictive power of eye-tracking data collected from videos reflecting three basic emotions to explore the usability of atypical visual attentional orientations of children with autism as a biomarker.

Method

Machine Learning

Machine learning (ML) is a sub-field of artificial intelligence that focuses on learning from data, identifying patterns, classifying, clustering, and predicting by imitating the way humans learn. ML uses a variety of learning approaches to deal with complex problems. These approaches are basically supervised learning, unsupervised learning and reinforcement learning. Supervised learning is used in solving classification problems using labeled datasets, while unsupervised learning is used in clustering problems with unlabeled datasets. On the other hand, reinforcement learning learns by interacting

Address for Correspondence: [1]Işık Akin-Bülbül, Hacettepe University, Education Faculty, Department of Special Education

Tablo 1. Participant demographic and clinical characteristics

Variable		Autism group (n = 54, range 18–36)	TD group (n = 70, range 18–36)	*p*
Age in months		28,20	27,07	0,87
Male/Female		10/44	33/37	0,00*
Bayley Cog. Score		77,04	100,00	0,00*
Bayley Lang. Score		67,69	99,89	0,00*
Bayley Motor Score		67,89	97,81	0,00*
M-Chat Class	No risk	-	-	-
	Medium risk	7	-	-
	High risk	47	-	-

Note: *p < 0.05. TD = Typically Developing; M-Chat = Modified Checklist for Autism in Toddlers.

with the environment without data at the beginning of the process (Shalev-Shwartz ve Ben-David, 2014).

Participants
This study was carried out within the scope of an international research project funded by the Scientic and Technological Research Council of Turkey (#115K459). The study participants consisted of children with autism with a mean age of 28.20 and TD children with a mean age of 27.07. Based on the DSM-V diagnostic criteria, children with autism were diagnosed by the child psychiatry clinics of universities or public hospitals. See Table 1 for participant characteristics of children with autism and TD children.

Instruments

Early Childhood Autism Screening Scale M-CHAT-Revised
The Early Childhood Autism Screening Scale (M-CHAT-F), developed to identify autism symptoms in early childhood, was developed by Robins, Fein, Barton, and Green (2021). M-CHAT-F is used to evaluate whether a child is suspected of autism. The current study used the scale to display the autism severity levels of children with autism.

Bayley-III Developmental Scale for Infants and Young Children
Bayley-III Infants and Toddlers Developmental Scale, which has high test-retest reliability and internal consistency scores, was developed to assess the developmental functioning levels of infants and children between 1and 42 months (Bayley, 2006). Bayley-III was used in the current study to determine the developmental levels of the participants in the Language, Cognitive, Motor, Social-Emotional, and Adaptive Subscales.

The Emotional States and Identified Area of Interests
The emotional states video set consists of videos that reflect male and female models' happy, neutral, and sad moods. There are two videos for each emotional state in the dataset, and each video lasts approximately 10 seconds. Six areas of interest were determined in all sets, namely, the face reflecting the mood, the face of the other person, the bodies of the persons, the object used, and the external space (see Figures 1, 2, and 3.).

Eye-Tracking and Procedure
In this study, the data gathering procedure used the SMI-Red 250 eye tracker fixed under a 17-inch LCD monitor with 1680 × 1050 pixels resolution. The participants sat alone in front of a computer screen, either on their preferred parent's lap or in an adjustable-height chair. A five-point calibration phase was completed with a crying baby animation on the screen. Next, the data sets were tracked in random order.
The participants' social attention parameters were recorded and stored using the Experiment Center™ 3.6 software. In the analysis of the recorded data, the SMI BeGaze™ (The Behavioral and Gaze Analysis) software was utilized. The eye movement sampling rate of the SMI-Red250 eye tracker was 250 Hertz (Hz).
Analysis
Monitoring parameters for three emotional states were recorded using an eye-tracking device, and study data were analyzed by implementing data mining methods. 12 features were extracted from each area of interest in the emotional states dataset. These attributes are numbered as follows. 1: [Net Dwell Time, ms], 2: [Dwell Time ms], 3: [Glance Duration, ms], 4: [Diversion Duration, ms], 5: [First Fixation Duration, ms], 6: [Glances

Table 2. Happy dataset features

	Happy-Face											Happy-Body	Object	Outside	Other Person Face		Other Person Body
	1	2	3	4	5	7	8	9	10	11	12	7		10	6	12	
RF	x	x	x	x	x	x	x	x	x	x	x				x		
IG	x	x	x	x	x	x	x	x	x	x	x					x	
GR	x	x	x	x	x	x	x	x	x	x	x					x	
W-NB	x														x		
W-J48									x								
W-KNN				x				x				x		x	x		

RF = ReliefF; IG = InfoGain; GR = Gain Ratio; W-NB = Wrapper Naive Bayes; W-J48 = Wrapper Decision Tree; W- KNN = Wrapper- K-Nearest Neighbors

Table 3. Neutral dataset features

	Neutral Face												Neutral Body		Object	Outside	Other Person Face		Other Person Body
	1	2	3	4	5	6	7	8	9	10	11	12	6	11	7		7	12	
RF	x	x	x	x		x	x	x	x	x	x				x			x	
IG	x	x	x	x		x	x	x	x	x	x						x	x	
GR	x	x	x	x	x	x	x	x	x	x	x							x	
W-NB			x					x					x		x				
W-J48	x		x			x						x		x	x				
W-KNN	x		x			x						x		x	x				

RF = ReliefF; IG = InfoGain; GR = Gain Ratio; W-NB = Wrapper Naive Bayes; W-J48 = Wrapper Decision Tree; W- KNN = Wrapper- K-Nearest Neighbors

Table 4. Sad dataset features

	Sad Face										Sad Body		Object	Outside		Other Person Face				Other Person Body
	1	2	3	4	5	8	9	10	11	12	3	6	7	6	7	4	9	10	11	
RF	x	x	x	x	x	x	x	x	x	x		x	x							
IG	x	x	x	x		x	x	x	x	x							x	x	x	
GR	x	x	x	x	x	x	x	x	x	x				x	x					
W-NB										x						x				
W-J48											x								x	
W-KNN										x										

RF = ReliefF; IG = InfoGain; GR = Gain Ratio; W-NB = Wrapper Naive Bayes; W-J48 = Wrapper Decision Tree; W- KNN = Wrapper- K-Nearest Neighbors

Count], 7: [Fixation Count], 8: [Net Dwell Time, %], 9: [Dwell Time, %], 10: [Fixation Time, ms], 11: [Fixation Time, %], and 12: [Average Fixation Duration, ms]. (See Appendix-1 for definitions of specified attributes.).
Datasets with a total of 72 features were created for the six identified areas of interest. ReliefF, Information Gain, Gain Ratio, and Wrapper feature selection methods were applied to the features determined for each dataset set. The attributes selected for the happy, neutral, and sad datasets are presented in Table 2, Table 3, and Table 4, respectively.

Table 5. Classification success rates for all datasets

	Happy Dataset				Neutral Dataset				Sad Dataset			
	ReliefF	Info Gain	Gain Ratio	Wrapper	ReliefF	Info Gain	Gain Ratio	Wrapper	ReliefF	Info Gain	Gain Ratio	Wrapper
DT	68,54	73,38	73,38	73,38	79,03	75,8	75,8	75,0	68,54	70,16	66,93	74,19
NB	63,70	62,9	62,9	69,35	78,22	76,61	78,22	80,64	75,00	74,19	74,19	76,61
KNN	69,35	57,25	57,25	76,61	71,77	75,00	63,7	81,45	73,38	71,77	66,12	74,19

DT = Decision Tree; NB = Naïve Bayes; KNN = K-Nearest Neigbour

Results

The feature selection methods applied to the current study features obtained based on the area of interest were determined as highly distinctive in the faces of the model reflecting the emotional states. When the machine learning algorithm results were examined, the KNN classification algorithm applied to the features determined by the Wrapper method in the happy data set distinguished children with autism from their peers with TG with a classification success of 76.61%. The Wrapper feature selection method showed a classification success of 81.45% with the KNN classification algorithm from the neutral mood data set. See Table 5 for the aggregated results of machine learning algorithms.

Discussion and Conclusion

This study's results showed that the faces of the models, which reflect the emotional states, were highly distinctive when machine learning algorithms were applied to identify the distinctive features. Children with autism and TD differed significantly in the facial area of interest, expressing happy and sad neutral moods. Many studies in the literature showed that children with TD are highly motivated to direct their visual attention to human faces during social interactions with an interaction partner from the earliest periods of development. Another well-established finding in the autism literature is that children with autism exhibit diminished visual attention to faces, which is an essential source of information in understanding the other person's emotional state. Although the source of this limited motivation toward human faces cannot be clearly identified in the current literature, studies report that diminished visual attention to human faces was observed in infants younger than six months. This early marker was considered one of the first signs of the atypical developmental course of autism (Shic, Macari, and Chawarska, 2014).

When the current study findings are evaluated as a whole, we observed that the highest success rate was achieved with the features determined by the Wrapper method. The highest classification success was performed with the KNN classification algorithm in the set, reflecting the neutral emotional state. The KNN classification algorithm classified children with autism with a success rate of 81.45%. Considering how the emotions displayed in the videos related to the emotional states, it is possible to say that the strongest expression was the sad emotion. For example, the female model cried quite strongly in the video, which reflects a sad mood. The happy data set takes second place in terms of the clarity of how emotions are displayed.

On the other hand, the most significant differentiation between groups was reached in the neutral data set. The clarification of the emotional expressions may have increased the motivation of children with autism to direct visual attention to the face area, resulting in a decrease in the difference between children with autism and TD children. Therefore, the use of neutral faces in future studies aiming to distinguish children with autism from TD children can better identify children with autism from TD children. The current study was designed based on a growing machine learning literature that was previously carried out using video and/or visual material that reflects emotional states (Jiang et al., 2019; Kennedy & Adolphs, 2012; Król & Król, 2019; Liu, Li, and Yi, 2016). At the same time, the fact that this research was conducted with very young children provides a strong advantage compared to previous study results. Although the use of eye-tracking technologies with machine learning algorithms to identify autism biomarkers is promising, one criticism directed at studies is that most studies were conducted with children older than four years old (Liu et al., 2016; Yaneva et al., 2020; Wan et al., 2019; Zhao et al., 2021). The mean chronological age of the participants of this study is 28.20 for children with autism and 27.07 for TD participants. This is a critical study in demonstrating the usability of visual attention directed to emotional states as a biomarker in children affected by moderate and severe autism symptoms from the earliest developmental stages. Although autism -related symptoms are present in children with autism from the earliest stages, unfortunately, the diagnosis of autism is usually made two to three years after the first symptoms appear, usually at the age of 4 (Dawson, 2012). On the other hand, a late diagnosis causes a delay in the early intervention implementations, which subsequently affects the prognosis of the diagnosis (Dawson et al., 2012). Overall, the study findings indicated the KNN algorithms produced the highest performance rate of 81.45%, when children watched the videos that reflect the neutral emotional state. Finally, our study results suggested that eye tracking data of visual attention directed towards neutral faces can be used as a biomarker for autism screening in young children.

Ek 1. Göz İzleme Parametreleri ve Tanımları

Parametre/Parameter	Birim /Unit	Tanım	Description
Sabitleme Sayısı Fixation Count	[sayı] [count]	Denemedeki sabitlenme sayısı.	Number of fixations in the trial.
Net Bekleme Süresi Net Dwell Time	[ms]	AOI'ye isabet eden tüm bakış veri örnekleri için örnek sürelerinin toplamı.	Sum of sample durations for all gaze data samples that hit the AOI.
Bekleme Süresi Dwell Time	[ms]	Toplam (seçilen tüm katılımcılar için bir AOI içindeki tüm sabitlemeler ve sakkadlar).	Sum (all fixations and saccades within an AOI for all selected participants).
Bakış Süresi Glance Duration	[ms]	Nesneye girmek için sakkad süresi + gözler AOI'den ayrılmaya başlamadan önce tüm sabitleme süreleri ve sakkad sürelerinin toplamı = bekleme süresi + AOI'ye giren sakkad süresi.	Saccade duration for entering the object + sum of all fixation durations and saccade durations before the eyes begin to leave the AOI = dwell time + duration of saccade entering AOI.
Yönlendirme Süresi Diversion Duration	[ms]	Tüm katılımcıların sapma süresinin toplamının katılımcı sayısına bölümü.	Sum of diversion duration of all participants divided by number of the participants.
İlk Sabitleme Süresi First Fixation Duration	[ms]	Bir AOI'daki ilk sabitlenme süresi.	The duration of the first fixation in an AOI.
Bakış Sayısı Glances Count	[sayı] [count]	Belirli bir süre içinde bir hedefe bakış sayısı	Number of glances to a target within a certain period
Net Bekleme Süresi Net Dwell Time	[%]	Net bekleme süresi (ms) / (bitiş zamanı - başlangıç zamanı)	Net dwell time (ms) / (end time - start time)
Bekleme Süresi Dwell Time	[%]	Bekleme süresi ortalaması % = bekleme süresi ortalaması * 100 / (geçerli zaman - başlangıç zamanı)	Dwell time average % = dwell time average * 100 / (current time - start time)
Sabitleme Süresi Fixation Time	[ms]	AOI içindeki sabitleme sürelerinin toplamı	Sum of the fixation durations inside the AOI
Sabitleme Süresi Fixation Time	[%]	Sabitleme süresi (ms) / (bitiş zamanı - başlangıç zamanı)	Fixation time (ms) / (end time - start time)
Ortalama Sabitleme Süresi Average Fixation Duration	[ms]	Bir AOI içindeki sabitleme sayısına bölünen sabitleme sürelerinin toplamı.	The sum of fixation times divided by number of fixations inside an AOI.

Not: Tablo 26.04.2022 tarihinde http://www.humre.vu.lt/files/doc/Instrukcijos/SMI/BeGaze2.pdf adresinden erişilen SMI BeGaze Manual (Versiyon 3.7) isimli kullanım kitapçığından yararlanarak oluşturulmuştur. Daha ayrıntılı açıklamalar için kullanım kitapçığına bakınız.