Cognitive State Measurement from Eye Gaze Analysis in an Intelligent Virtual Reality Driving System for Autism Intervention

Lian Zhang[1], Joshua Wade[1], Amy Swanson[2], Amy Weitlauf[2,3], Zachary Warren[2,3], and Nilanjan Sarkar[4,1]
[1]Electrical Engineering and Computer Science Department
[2]Treatment and Research in Autism Spectrum Disorder (TRIAD)
[3]Pediatrics and Psychiatry Department
[4]Mechanical Engineering Department
Vanderbilt University, Nashville, TN 37212

***Abstract*— Autism Spectrum Disorder (ASD) is a group of neurodevelopmental disabilities with a high prevalence rate. While much research has focused on improving social communication deficits in ASD populations, less emphasis has been devoted to improving skills relevant for adult independent living, such as driving. In this paper, a novel virtual reality (VR)-based driving system with different difficulty levels of tasks is presented to train and improve driving skills of teenagers with ASD. The goal of this paper is to measure the cognitive load experienced by an individual with ASD while he is driving in the VR-based driving system. Several eye gaze features are identified that varied with cognitive load in an experiment participated by 12 teenagers with ASD. Several machine learning methods were compared and the ability of these methods to accurately measure cognitive load was validated with respect to the subjective rating of a therapist. Results will be used to build models in an intelligent VR-based driving system that can sense a participant's real-time cognitive load and offer driving tasks at an appropriate difficulty level in order to maximize the participant's long-term performance.**

Keywords—Autism; autism intervention; virtual reality; driving; cognitive load measurement; eye gaze features

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a group of neurodevelopmental disabilities characterized by pervasive impairments in social communication and behavioral functioning. The estimated prevalence of ASD is 1 in 68 in the United States [1]. While researchers have extensively studied how to improve social skills, language development, and emotion recognition in young children with ASD[2-4], far fewer studies have focused on meaningful skills related to adaptive adult independence such as driving. Many individuals fail to achieve typical milestones related to adult independence, with the ability to drive an automobile representing a particularly important skill for individuals with ASD as well as typically developing adults.

Recent literature has shown that individuals with ASD face difficulty in learning and maintaining driving skills [5-8]. Compared with their Typically Developing (TD) peers, individuals with ASD have difficulty in identifying driving hazards and appear to exhibit differences in gaze pattern and physiology signals while driving [6, 7, 9].

A growing number of studies have investigated Virtual Reality (VR)-based intervention for children with ASD [10]. VR technology offers the potential to create an immersive, interactive, and realistic environment for behavioral learning and generalization in real-world situations. Using VR-based intervention platforms for ASD treatment has the advantage of allowing precise control of complex stimuli, providing individualized treatment, and creating a structured and safe learning environment [10-12]. We developed a novel VR-based driving system for the purpose of addressing driving skills of adolescents with ASD. A future goal is to incorporate closed-loop control into this VR-based driving system such that each individual can be trained at an optimal cognitive load by manipulating the difficulty levels of the driving tasks.

An individualized intelligent VR-based driving system, which can offer driving tasks with appropriate cognitive load for each individual, may have the ability to improve the user's long-term performance. The cognitive capacity in working memory is limited, and thus if a learning task requires too little or too much cognitive capacity, learning may not be optimal [13]. A proper cognitive load has the ability to maximize a participant's long-term performance by offering challenging tasks within his/her ability [14, 15]. The cognitive feedback has been studied in the context of intelligent systems [15-17], which provide the motivation for this work.

One of the challenges to building such an intelligent system is to measure a participants' cognitive load in real time. Eye gaze has been shown to reflect cognitive information [18-20] and thus has the potential for real-time cognitive load measurement [18]. In this paper, we explore the feasibility of eye gaze as a tool for assessing cognitive load of individuals with ASD as they participate in a VR-based driving task. Tracking eye movement is widely used and a valuable means for cognitive load measurement [19]. Eye gaze data include rich information, such as blink rate and pupil diameter, that reflects a user's cognitive state [18, 20]. Previous research documented that pupil dilation

quickly responded to changes in cognitive workload [20, 21] and that pupil diameter increased as workload increased [20, 22, 23]. However, when cognitively overloaded, the pupil diameter decreased as the workload intensified [21]. The mean pupil diameter [18] and the blink rate [24] also varied under different workload conditions. The blink duration became shorter under a higher workload [22]. Because of this variability, for this work we extracted multiple eye gaze features that are known to vary with cognitive loads and input them into classification methods for cognitive load measurement.

Perceived task difficulty is a widely used indicator for cognitive load [25] and such difficulty level rating was utilized to verify cognitive load measurement in [26, 27]. The correlation between cognitive load and perceived task difficulty has been shown in [28, 29]. It was further demonstrated that the difficulty level has the ability to modulate the cognitive load a user experiences [16, 27], which implies the option of adjusting difficulty level to optimize the cognitive load in the closed-loop intelligent system. In the current work, we compared six classification methods for cognitive load measurement. The results were verified against a therapist's ratings of perceived difficulty level.

The contributions of this paper are two-fold. By having teenagers diagnosed with ASD participating in a VR-based driving system, we first demonstrate that there are several eye gaze features that vary under different cognitive loads. We then perform a comparative study on feature-based measurement of cognitive load. This work will inform the future closed-loop intelligent driving system that we are developing, which will adapt the driving task difficulty to optimize a participant's cognitive load in order to optimize learning.

This paper is organized as follows. The VR-based driving system and the experiment protocol are described in Section II. Eye gaze feature extraction and analysis are presented in Section III. The results using eye gaze features for cognitive load measurement are discussed in the following section. Section V summarizes the contributions and limitations of the present work.

II. SYSTEM DESIGN AND EXPERIMENT SETUP

A. The VR-based drivnig system

The VR-based driving system included three components: a driving simulator, a data capture module and a therapist rating module, shown in Fig. 1. The participant operated the driving simulator to complete driving assignments. His/her peripheral physiological signals, Electroencephalogram (EEG) signals, eye gaze data and performance data were acquired via the data capture module. In the therapist rating module, a therapist observed and rated the participant's affective and cognitive state in real time. These three components of the VR-based driving system were synchronized by time stamped events from the driving simulator via a local area network (LAN).

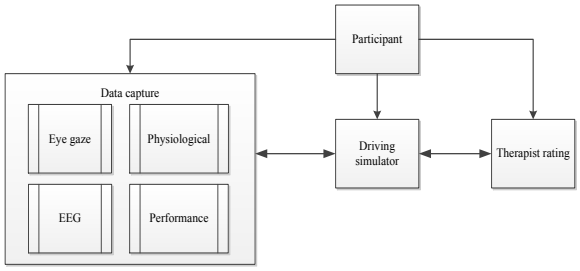

Fig. 1. The framework of VR-based driving system

The driving simulator included a VR driving environment and a Logitech G27 steering wheel controller, shown in Fig. 2. The models in the VR driving environment were built with the modeling tools ESRI CityEngine (www.esri.com/cityengine) and Autodesk Maya (www.autodesk.com/maya). Driving assignments in the VR driving environment were developed with the game engine Unity3D (www.unity3d.com). Eighteen different driving assignments belonging to six different difficulty levels were developed in the VR driving environment.

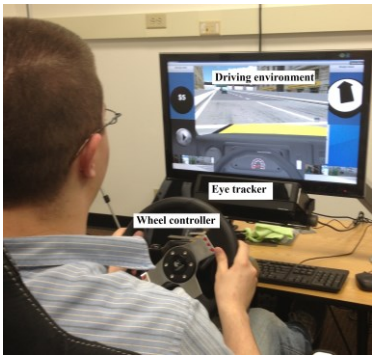

Fig. 2. The driving simulator and eye tracker

The difficulty level of a driving assignment was determined by controlling parameters of the driving simulator. Controlling parameters included the speed limit of the agent vehicle, the weather of the driving environment, and the responsiveness of the agent vehicle's brakes and accelerator. In the lowest difficulty level (level 1), the speed limit was the lowest, the vehicle fully responded to the brake and accelerator, and the weather was sunny. In the highest difficulty level (level 6), the speed limit was the highest, the responsiveness of brakes and the accelerator was only half of the level 1, and the weather was raining. The difficulty levels were tested and validated in our previous work [30]. Three different driving assignments was included in each difficulty level.

The data capture module recorded participants' driving performance, EEG, peripheral physiological, and eye gaze data. Participants' driving behaviors and task performance were automatically logged as their driving performance. The driving behaviors indicated how the participant drove the agent vehicle. The task performance reflected to what degree one participant completed the driving tasks. The data capture module also measured data from three physiology related modalities: EEG, peripheral physiological and eye gaze data, shown in Fig. 3. A Tobii X120 remote eye tracker

(www.tobii.com/) was used to track the participant's eye gaze data. Biopac MP150 (www.biopac.com) sensors recorded peripheral physiological signals wirelessly with a frequency of 1000Hz. The peripheral physiological signals included electrocardiogram (ECG), photoplethysmogram (PPG), electromyogram (EMG), respiration (Resp.) skin temperature (SKT), and galvanic skin response (GSR). An Emotiv EEG headset (www.emotiv.com) recorded 14-channel EEG signals with frequency 128Hz. The 14-channel signals were measured from positions AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, defined by the 10-20 system of electrode placement [31].

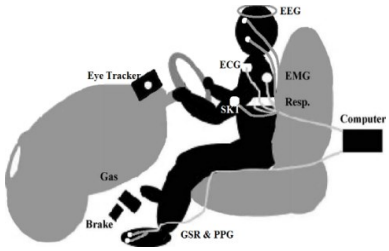

Fig. 3. The data capture module

In the rating module, the participant's affective and cognitive states were observed and rated by a therapist on a 0-9 Likert scale in real time. The difficulty level rating was used as the ground-truth for cognitive load since difficulty level was an indicator of cognitive load [25]. The initial difficulty level ratings were clustered into two classes for classification based on the therapist's rating criterion. Ratings under 5 fell into the low cognitive load class. The ratings greater than 5 were in the high cognitive load class.

The perceived difficulty level was more flexible to reflect the cognitive load compared to the designed task difficulty level because the same task may cause different cognitive loads for different persons depending on their individual ability, state, and attitude [32]. Additionally, children with ASD commonly have challenges with recognition and communication of cognitive and affective experiences and thus self-reported cognitive and affective states were considered to be less reliable than a trained therapist's ratings [33]. As a result, the ratings from a trained research clinician with multiple years of experience in coding behavior and affective states of individuals with ASD was used for ground-truth of cognitive load in this paper.

B. Experiment setup

Twelve teenagers (ages 13-17 years) with ASD completed the driving experiments. Participants were recruited from an existing university clinical research registry. All participants had scores at or above clinical cutoff on the Autism Diagnostic Observation Schedule [34]. Their cognitive functioning was measured using either the Differential Ability Scales [35] or the Wechsler Intelligence Scale for Children [36].

Each participant completed a series of six experiments in different days. Each experiment included three driving assignments. The driving assignments in each experiment were unique except the first and the last experiments. TABLE I. shows all three assignments within each experiment. For example, in the first experiment, its three assignments were: assignment2 in level 2 (L2A2), assignment1 in level5 (L5A1), and assignment2 in level5 (L5A2).

TABLE I. THE ASSIGNMENTS FOR EACH EXPERIMENT

Experiment index	1	2	3	4	5	6
Assignment1	L2A2	L1A1	L3A1	L4A1	L6A1	L2A2
Assignment2	L5A1	L1A2	L3A2	L4A2	L6A2	L5A1
Assignment3	L5A2	L1A3	L3A3	L4A3	L6A3	L5A2

At the beginning of an experiment, device setup was carried out by three researchers followed by a nine-point eye gaze calibration. Three minutes of baseline data including peripheral physiological and EEG data was recorded in a silent environment. Then, the participant practiced driving for three minutes in a free-form practice mode. After the practice, the participant completed three specific driving assignments of that experiment. Each experimental session lasted approximately one hour.

III. EYE GAZE DATA ANALYSIS

The Tobii X120 remote eye tracker was used to track participant's eye movement with a frequency of 120Hz. The detection range of the eye tracker was 22×22×30cm. The average accuracy of the eye tracker for eye gaze position was 0.5 degree, which was less than 1cm when the average distance between the participant and the screen was 70 cm as shown in (1). The initial eye gaze data offered by the eye tracker and their descriptions are listed in TABLE II.

$$70\times\tan(0.5^{\circ}/2) = 0.305\text{cm} \quad (1)$$

TABLE II. THE INITIAL EYE GAZE DATA AND DESCRIPTION

Data	Label	Description
Gaze position	(x_{li}, y_{li}), (x_{ri}, y_{ri})	Horizontal and vertical gaze position of one eye (left or right eye) on the screen.
Pupil	p_{li}, p_{ri}	Size of pupil of one eye.
Distance	d_{li}, d_{ri}	Distance from one eye to the eye tracker.
Validity	v_{li}, v_{ri}	Validity of one eye. The value can be 0, 1, 2, 3, or 4, where 0 means that the eye was well found and 4 means that the eye was not found.

A. Preprocessing of eye gaze signal

The initial eye gaze data were preprocessed by 1) removing long duration invalid data; 2) reducing noise; and 3) mapping initial data into a quadruple data structure (x_i, y_i, d_i, v_i) for feature extraction.

The invalid data in the initial eye gaze data were generated from 1) noise; 2) participant closed his eyes for blink; 3) participant's head moved out of the eye tracker's detection range. We first removed long duration invalid data generated from head movement. The duration of invalid data from noise and the blink duration were less than 1000ms. So if the duration of a section of continuous invalid data was more than one second, the data were removed.

We then reduced noise with a gap-fill-in method and median filter. The gap-fill-in method [37] filled in the lost eye gaze data via linear interpolation. The lost eye gaze points were added along a straight line, which were calculated with two nearest valid data before and after the lost data. The gap with less than 75ms length was filled since the minimum blink duration was 75ms [38]. The median filter in Matlab was implemented to reduce the noise after the gap-fill-in method.

Initial eye gaze data were then mapped into a 4-dimensional dataset (x_i, y_i, d_i, v_i) for feature extraction. The validity of the eye gaze data (v_i) was calculated with (2). If one of the eyes was detected, that eye gaze sample was considered valid. The 2-D eye gaze position (x_i, y_i) was the average gaze position of the right eye and the left eye on the screen: $x_i = (x_{li} + x_{ri})/2$ and $y_i = (y_{li} + y_{ri})/2$ when the data was valid.

$$v_i = \begin{cases} \text{valid}, & \text{if } v_{ri} < 2 \text{ or } v_{li} < 2 \\ \text{invalid}, & \text{otherwise} \end{cases} \quad (2)$$

The averaged initial distance between the eye and the eye tracker (d_{ii}) was calculated from initial distances from one eye to the eye tracker: $d_{ii} = (d_{li} + d_{ri})/2$. The distance between eye and eye gaze position on the screen (d_i) was approximated using the averaged initial distance between the eye and the eye tracker: $d_i \approx d_{ii}$. All distances are shown in Fig. 4.

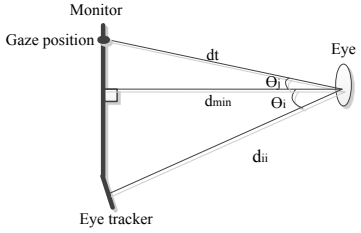

Fig. 4. visual distances

The angle θ_i equals the eye tracker angle, which was less than 30˚ during experiments. Considering the monitor size, θ_j is less than θ_i. The relation of the distance d_{ii} and actual eye and eye gaze position distance d_t satisfies (3). The equation ignores the distance between the eye tracker and the bottom of the monitor since it is much smaller compared to d_t and d_{ii}.

$$d_{ii}\cos\theta_i = d_{min} = d_t\cos\theta_j$$
$$d_t < d_{ii} = d_t\cos\theta_j / \cos\theta_i < d_t/\cos 30^\circ = 1.15 d_t \quad (3)$$

The error between d_{ii} and d_t is less than $0.15d_t$. So, using $d_i \approx d_{ii}$ is a simple and reliable way for the calculation of the distance between the eye and the eye gaze position.

B. Eye gaze features

We calculated six raw features after preprocessing, shown in TABLE III. Ten initial eye gaze features were generated from the six raw features, including the blink rate, the fixation rate, the mean and standard deviation (Std.) of blink duration, pupil diameter, fixation duration, and saccade duration.

TABLE III. THE EYE GAZE FEATURES

Raw features	Description
Blink rate	The blink frequency per minute
Blink duration	The closure time duration of a blink in ms
Pupil diameter	The pupil diameter in mm
Fixation duration	The time duration when the gaze stays at one point before shifting
Fixation rate	The frequency of looking at some point per second
Saccade duration	The time duration of a rapid movement of the eye between two fixation points.

The blink was defined with a closure duration from 75 to 400 milliseconds [39]. A blink was detected only when the eyes' closure duration was within the thresholds. The pupil diameter was the average of pupil diameters of two eyes measured in mm ($p_i = (p_{li} + p_{ri})/2$).

An adaptive algorithm for fixation and saccade detection from [39] was implemented. Compared to the algorithm with a predefined static velocity threshold [40], this adaptive algorithm calculated a dynamic threshold from the eye gaze data, which made the fixation and saccade detection more robust. The adaptive algorithm included three steps: 1) angular velocity (unit deg/sec) calculation; 2) iterative determination of the threshold between the fixation and the saccade; and 3) detection of the fixation and saccade with the calculated threshold.

Savitzky-Golay filter in Matlab was used to reduce the noise for visual angle calculation. The changed visual angle (unit in degree) from the $(i-1)^{th}$ data to the i^{th} data was calculated from (4).

$$\Delta\theta_i = 2\tan^{-1}(\| (x_i, y_i), (x_{i-1}, y_{i-1}) \| / 2d_i) \quad (4)$$

The visual angle velocity θ_i' was calculated with the changed visual angle in (5), where f = 120 was the eye tracking frequency.

$$\theta_i' = f\Delta\theta_i \quad (5)$$

The threshold was calculated with an iterative function as shown in (6). The initial peak velocity threshold was set to PT_1 = 100 degree/second. The iteration stopped when $|PT_n - PT_{n-1}| < 1$degree/second.

$$PT_n = \mu_{n-1} + 4\,\sigma_{n-1} \quad (6)$$

The fixation was detected when the visual angular velocity (θ_i') was less than the calculated threshold. The saccade was the eye gaze movement with angular velocity larger than the calculated threshold.

Baseline features were considered in order to reduce the effects of individual differences in eye gaze pattern and experiment environment differences. However, it was very difficult to capture baseline eye gaze data in an ideal condition for teenagers with ASD. Many participants did not look at all towards the screen during the baseline time (and one even fell asleep). Instead, the eye gaze data before the driving assignment excluding the practice time were used as non-ideal condition baseline data. The baseline data were

selected only if they included more than one minute of valid data.

The ten initial eye gaze features (named features without baseline) were extracted from eye gaze data during driving assignments. Ten baseline features were extracted from the non-ideal condition baseline data. Features with baseline were calculated by subtracting the baseline features from the initial features. Because of non-ideal conditions (e.g. the noise and the data loss due to head movements) for baseline recording we were not sure whether and to what extent the baseline features would be useful. We have, therefore, decided to include features with and without baseline for the cognitive load measurement.

IV. RESULTS

Each participant completed a total of eighteen assignments during his/her six experiments. Each assignment lasted two to five minutes. We got one data sample from each assignment. There were a total of 216 data samples from twelve participants. Because of unwanted movements of the teenagers with ASD during experiments, 31 data samples were removed. 185 data samples were remained for data analysis, among which 89 and 96 data samples belonged to the high and low cognitive load classes, respectively.

One way Analysis of variance (ANOVA) was used to test whether the means of the eye gaze features under different cognitive loads were significantly different. The mean features, which were significantly different under different cognitive loads ($p<0.05$), were: the blink rate with baseline ($p = 0.0401$), pupil diameter with baseline ($p=0.0295$), pupil diameter without baseline ($p=0.0013$), fixation duration mean without baseline ($p=0.00185$), fixation rate without baseline ($p=0.0066$), and saccade duration mean without baseline ($p=0.0019$). Fig. 5 shows the box plot of these eye gaze features in cognitive load measurement. The left box with number 1 represented the data from the high cognitive load condition; while the right box with number 2 was data from the low cognitive load condition.

Most of the features without baseline outperformed the corresponding features with baseline. The blink rate was an exception. However, the difference in the blink rate with baseline under different cognitive loads was small as its p value was close to 0.05. It seems that the usefulness of the non-ideal condition baseline data was limited. One possible reason why the blink rate with baseline performed better than the blink rate without baseline was that the influence of the individual differences of blink rate was large [41].

The blink rate with baseline increased when the cognitive load was high, Fig. 5 (a). A similar variation of blink rate for TD population was found in [24]. The pupil diameter (Fig. 5 (b) and Fig. 5 (c)) increased as the cognitive load increased. The change of pupil diameter with cognitive load was consistent with results in [21] for a TD population. The fixation duration in this experiment was mainly in the range of 250 ms to 400ms. The saccade duration, on the other hand, was in the range of 45ms to 55ms. The fixation duration decreased (Fig. 5 (d)), the fixation rate decreased (Fig. 5 (e)), and the saccade duration increased (Fig. 5 (f)) under a higher cognitive load.

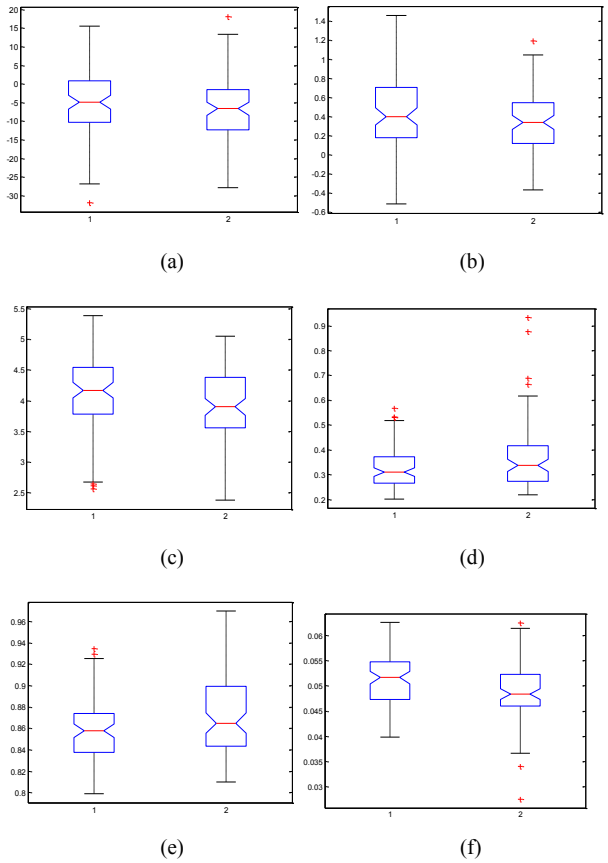

Fig. 5. The box plot of (a) the blink rate with baseline, (b) the pupil diameter mean with baseline, (c) the pupil diameter mean without baseline, (d) the fixation duration mean without baseline, (f) the fixation rate without baseline, and (g) the saccade duration mean without baseline.

TABLE IV. THE CLASSIFICATION RESULTS

Method	Parameters	Accuracy
NaiveBayes		76.22%
SVM	Kernel degree = 1	77.84%
Logistic regression	Update Method = conjugate gradient descent	72.97%
KNN	K = 1	78.38%
Neural network	HiddenLayer = 7	72.43%
Decision tree	Search Method = GreedyStepWise	76.22%

The Waikato Environment for Knowledge Analysis (WEKA)[1] was used for cognitive load classification. The important components of the eye gaze features were extracted using the Principal Component Analysis (PCA) before classifying. PCA output thirteen principal components from twenty features. The thirteen principal components were input into six selected machine learning methods for cognitive load measurement. The selected classification

[1] http://www.cs.waikato.ac.nz/ml/weka/

methods included Naïve Bayes, Support Vector Machine (SVM), Logistic regression, K-nearest-neighbors (KNN), Neural Network and Decision tree. A 7-fold cross validation method was used to verify the classification method. TABLE IV. shows parameters and the numerical results of all methods.

Naïve Bayes used supervised discretization to convert numeric features to nominal ones for preprocessing, which increased the accuracy. The SVM used linear kernel (kernel degree = 1). An increased kernel degree did not increase the accuracy any further. The conjugate gradient descent [42] was selected as the update method for Logistic regression. The KNN method classified each new data sample based on one nearest data sample with known class. For the neural network, the number of hidden layers was calculated by the function: (feature number + class number)/2. We had 13 features after PCA and 2 classes. As a result, 7 hidden layers were used for neural network. In order to create the decision tree, a greedy stepwise algorithm [42] was used. The accuracies of all methods were above 72%, which are comparable to some of the existing results [19, 43]. The KNN method achieved the best result with 78.38% accuracy.

V. CONCLUSION

A VR-based driving system was designed to train and improve the driving skills of teenagers with ASD. Twelve teenagers with ASD completed eighteen driving assignments in six experiments. Their eye gaze data were analyzed to measure their cognitive loads. The data analysis will be used to build models for an intelligent driving system, which can offer driving assignments with proper difficulty level tailored to maximize users' performance.

ANOVA test indicated that several eye gaze features were important for cognitive load measurement. In the literature [21, 24] TD individuals reported a higher blink rate and increased pupil diameter under high cognitive load. We observed a similar variation of these features for teenagers with ASD. Other findings included decreased fixation duration and increased saccade duration with increased cognitive load for the teenagers with ASD during driving. No comparison of features between the ASD and the TD groups was included since no TD individuals were involved in the current experiment, which was one of the limitations of this current work.

The accuracies for binary cognitive load measurement using eye gaze features were above 72% for all six classification methods selected in this paper, Naïve Bayes, SVM, Logistic Regression, KNN, Neural Network and Decision Tree. The best accuracy was 78.38% with the KNN method, which was comparable to existing results for TD individuals [19, 43].

Most existing studies of eye gaze analysis examine TD individuals under ideal conditions. This eye gaze data analysis for teenagers with ASD under driving conditions was difficult due to participants' frequent head movements while driving. The processing of data with a large quantity of invalid eye gaze data and noise was proved to be useful in this paper.

These preliminary results in this paper will be used to adjust task difficulty level for an optimized cognitive load in an intelligent driving system for individuals with ASD to maximize their performance in the future.

References

[1] M. Wingate, R. S. Kirby, S. Pettygrove, C. Cunniff, E. Schulz, T. Ghosh, C. Robinson, L.-C. Lee, R. Landa, and J. Constantino, "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010," *MMWR Surveillance Summaries,* vol. 63, 2014.

[2] M. L. Sundberg and J. W. Partington, "Teaching language to children with autism and other developmental disabilities," *Pleasant Hill, CA: Behavior Analysts,* 1998.

[3] N. Bauminger, "The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes," Journal of autism and developmental disorders, vol. 32, pp. 283-298, 2002.

[4] O. Golan, E. Ashwin, Y. Granader, S. McClintock, K. Day, V. Leggett, and S. Baron-Cohen, "Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces," Journal of autism and developmental disorders, vol. 40, pp. 269-279, 2010.

[5] N. B. Cox, R. E. Reeve, S. M. Cox, and D. J. Cox, "Brief Report: Driving and young adults with ASD: Parents' experiences," Journal of autism and developmental disorders, vol. 42, pp. 2257-2262, 2012.

[6] E. Sheppard, D. Ropar, G. Underwood, and E. van Loon, "Brief report: Driving hazard perception in autism," Journal of autism and developmental disorders, vol. 40, pp. 504-508, 2010.

[7] B. Reimer, R. Fried, B. Mehler, G. Joshi, A. Bolfek, K. M. Godfrey, N. Zhao, R. Goldin, and J. Biederman, "Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm," Journal of autism and developmental disorders, vol. 43, pp. 2211-2217, 2013.

[8] B. P. Daly, E. G. Nicholls, K. E. Patrick, D. D. Brinckman, and M. T. Schultheis, "Driving behaviors in adults with Autism Spectrum Disorders," Journal of autism and developmental disorders, vol. 44, pp. 3119-3128, 2014.

[9] J. Wade, D. Bian, L. Zhang, A. Swanson, M. Sarkar, Z. Warren, and N. Sarkar, "Design of a Virtual Reality Driving Environment to Assess Performance of Teenagers with ASD," in Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge, ed: Springer, 2014, pp. 466-474.

[10] M. Wang and E. Anagnostou, "Virtual reality as treatment tool for children with autism," in Comprehensive guide to autism, ed: Springer, 2014, pp. 2125-2141.

[11] T. R. Goldsmith and L. A. LeBlanc, "Use of technology in interventions for children with autism," Journal of Early and Intensive Behavior Intervention, vol. 1, p. 166, 2004.

[12] D. Strickland, "Virtual reality for the treatment of autism," Studies in health technology and informatics, pp. 81-86, 1997.

[13] T. De Jong, "Cognitive load theory, educational research, and instructional design: some food for thought," Instructional Science, vol. 38, pp. 105-134, 2010.

[14] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," Interacting with computers, vol. 24, pp. 154-172, 2012.

[15] G. F. Wilson and C. A. Russell, "Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding," Human factors: the journal of the human factors and ergonomics society, vol. 49, pp. 1005-1018, 2007.

[16] A. Koenig, D. Novak, X. Omlin, M. Pulfer, E. Perreault, L. Zimmerli, M. Mihelj, and R. Riener, "Real-time closed-loop control of cognitive load in neurological patients during robot-assisted gait training," Neural Systems and Rehabilitation Engineering, IEEE Transactions on, vol. 19, pp. 453-464, 2011.

[17] M. Mihelj, D. Novak, and M. Munih, "Emotion-aware system for upper extremity rehabilitation," in Virtual Rehabilitation International Conference, 2009, 2009, pp. 160-165.

[18] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman, "Estimating cognitive load using remote eye tracking in a driving simulator," in Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, 2010, pp. 141-144.

[19] E. Haapalainen, S. Kim, J. F. Forlizzi, and A. K. Dey, "Psycho-physiological measures for assessing cognitive load," in Proceedings of the 12th ACM international conference on Ubiquitous computing, 2010, pp. 301-310.

[20] M. Pomplun and S. Sunkara, "Pupil dilation as an indicator of cognitive workload in human-computer interaction," in Proceedings of the International Conference on HCI, 2003.

[21] E. Granholm, R. F. Asarnow, A. J. Sarkin, and K. L. Dykes, "Pupillary responses index cognitive resource limitations," Psychophysiology, vol. 33, pp. 457-461, 1996.

[22] U. Ahlstrom and F. J. Friedman-Berg, "Using eye movement activity as a correlate of cognitive workload," International Journal of Industrial Ergonomics, vol. 36, pp. 623-636, 2006.

[23] T. de Greef, H. Lafeber, H. van Oostendorp, and J. Lindenberg, "Eye movement as indicators of mental workload to trigger adaptive automation," in Foundations of augmented cognition. Neuroergonomics and operational neuroscience, ed: Springer, 2009, pp. 219-228.

[24] D. E. Irwin and L. E. Thomas, "Eyeblinks and cognition," Tutorials in visual cognition, pp. 121-141, 2010.

[25] F. G. Paas, "Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach," Journal of educational psychology, vol. 84, p. 429, 1992.

[26] S. Kalyuga, P. Chandler, and J. Sweller, "Managing split-attention and redundancy in multimedia instruction," Applied cognitive psychology, vol. 13, pp. 351-371, 1999.

[27] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in Proceedings of the 24th Australian Computer-Human Interaction Conference, 2012, pp. 420-423.

[28] S. Chen, J. Epps, and F. Chen, "A comparison of four methods for cognitive load measurement," in Proceedings of the 23rd Australian Computer-Human Interaction Conference, 2011, pp. 76-79.

[29] A. Girouard, E. T. Solovey, L. M. Hirshfield, K. Chauncey, A. Sassaroli, S. Fantini, and R. J. Jacob, "Distinguishing difficulty levels with non-invasive brain activity measurements," in Human-Computer Interaction–INTERACT 2009, ed: Springer, 2009, pp. 440-452.

[30] D. Bian, J. W. Wade, L. Zhang, E. Bekele, A. Swanson, J. A. Crittendon, M. Sarkar, Z. Warren, and N. Sarkar, "A novel virtual reality driving environment for autism intervention," in Universal access in human-computer interaction. User and context diversity, ed: Springer, 2013, pp. 474-483.

[31] G. H. Klem, H. O. Lüders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation," Electroencephalogr Clin Neurophysiol, vol. 52, p. 3, 1999.

[32] F. A. Muckler and S. A. Seven, "Selecting performance measures:" Objective" versus" subjective" measurement," Human factors: the journal of the human factors and ergonomics society, vol. 34, pp. 441-455, 1992.

[33] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Online affect detection and robot behavior adaptation for intervention of children with autism," Robotics, IEEE Transactions on, vol. 24, pp. 883-896, 2008.

[34] C. Lord, S. Risi, L. Lambrecht, E. H. Cook Jr, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism," Journal of autism and developmental disorders, vol. 30, pp. 205-223, 2000.

[35] C. D. Elliott, Differential Ability Scales-ll: San Antonio, TX: Pearson, 2007.

[36] D. Wechsler, "Wechsler intelligence scale for children," 1949.

[37] A. Olsen, "The Tobii I-VT fixation filter," Copyright© Tobii Technology AB, 2012.

[38] O. V. Komogortsev, D. V. Gobert, S. Jayarathna, D. H. Koh, and S. M. Gowda, "Standardization of automated analyses of oculomotor fixation and saccadic behaviors," Biomedical Engineering, IEEE Transactions on, vol. 57, pp. 2635-2645, 2010.

[39] M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," Behavior research methods, vol. 42, pp. 188-204, 2010.

[40] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in Proceedings of the 2000 symposium on Eye tracking research & applications, 2000, pp. 71-78.

[41] M. Ingre, T. ÅKerstedt, B. Peters, A. Anund, and G. Kecklund, "Subjective sleepiness, simulated driving performance and blink duration: examining individual differences," Journal of sleep research, vol. 15, pp. 47-53, 2006.

[42] C. M. Bishop, Pattern recognition and machine learning vol. 4: springer New York, 2006.

[43] S. Chen and J. Epps, "Automatic classification of eye activity for cognitive load measurement with emotion interference," Computer methods and programs in biomedicine, vol. 110, pp. 111-124, 2013.