Journal of Ambient Intelligence and Humanized Computing (2020) 11:2701–2713
https://doi.org/10.1007/s12652-019-01329-8

ORIGINAL RESEARCH

Classification approach for understanding implications of emotions using eye-gaze

Pradeep Raj Krishnappa Babu[1] · Uttama Lahiri[2]

Received: 5 November 2018 / Accepted: 18 May 2019 / Published online: 24 May 2019

Abstract

Atypical behavioral viewing pattern is one of the core deficits of individuals with Autism Spectrum Disorder (ASD). This diminishes their ability to understand the communicator's facial emotional expression and often misinterpret one's intended emotion. Here, we investigated the feasibility of using one's gaze-related indices to estimate distinctive changes corresponding to various emotions. We designed a usability study with nine individuals with ASD and Typically Developing (TD) individuals who were exposed to Virtual Reality (VR) based social scenarios. The VR scenes presented virtual characters who narrated their social experience in the form of short stories with context-relevant emotional expressions. Simultaneously, we collected one's gaze-related physiological indices (PIs) and behavioral looking pattern indices (BIs) using a technologically-enhanced eye-tracker. Subsequently, these PIs and BIs were used to classify the implications of the emotional expressions both within and across the ASD and TD groups. Results of the usability study indicate that one's gaze-related indices can be discriminated with 97% accuracy for various emotions for intra-group analysis and 100% accuracy for inter-group analysis.

Keywords Autism · Virtual reality · Eye-tracking · Fixation duration · Pupil diameter · Blink rate · Classification · SVM

1 Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder having a prevalence rate of 1 in 150 in India (Autism cases rise in last two decades 2017) and 1 in 100 globally (Data and Statistics 2018). Also, a survey shows that 85% of children with neurodevelopmental disorders belong to low and middle-income countries like India (Silberberg et al. 2013). Profound impairments in social reciprocation, repetitive and restricted idiosyncratic behavior are some of the associated core deficits of individuals with ASD (APA 2013). Additionally, they face difficulty in understanding non-verbal cues such as emotional expression, body gestures, etc. coupled with atypical looking pattern (APA 2013). One's looking pattern on communicator's face plays a significant role in picking up cues on the communicator's emotion. Studies show that individuals with ASD, often exhibit reduced attention towards the social communicator's face and increased attention towards non-social objects (Uljarevic and Hamilton 2013). This atypical gaze pattern often deters their ability to process one's emotional expression (Uljarevic and Hamilton 2013) that adversely affects fluid interpersonal communication.

There is a consensus that appropriate intervention strategies can facilitate substantial improvement in social communication skills of these individuals with ASD. For social skill training, therapists often use photographs (My Aspergers Child 2007) to train these individuals in the art of understanding others' facial emotional expressions. This is important since these individuals face challenges in understanding emotions of social partners, as shown by a meta-analysis of 48 research studies (Uljarevic and Hamilton 2013). This is often differentiated based on the type of emotion. For example, research studies show that individuals with ASD often misinterpret one's negative emotions more than the positive emotions (Uljarevic and Hamilton 2013). This is attributed to their atypical looking pattern towards faces with negative emotions in comparison to those with positive emotions (Uljarevic and Hamilton 2013; Kliemann et al. 2010). Additionally, these individuals face milestones in making an explicit expression of their affective states

✉ Pradeep Raj Krishnappa Babu
pradeepr@iitgn.ac.in

[1] Center for Cognitive Science, Indian Institute of Technology Gandhinagar, Gandhinagar, India

[2] Electrical Engineering Department, Indian Institute of Technology Gandhinagar, Gandhinagar, India

(Begeer et al. 2008) thereby putting limitations on conventional observation-based techniques. Thus, the success of conventional techniques depends on therapists who use their expert eyes to understand the child's affect and engagement in a task and accordingly modulate their intervention strategy (Hofmann and Smits 2008). Though the conventional observation-based techniques are powerful, yet these suffer from certain limitations, such as subjectivity, the cost of one-to-one therapy sessions, and insufficient availability of trained therapists (Hofmann and Smits, 2008; Watling et al. 1999), particularly in developing countries like India. Also, given the heterogeneous nature of the autism spectrum and the core deficit of this target group in making explicit expressions of their affect, technology-assisted platforms can be a complementary tool to address at least some of the resource limitations as faced by conventional observation-based techniques. Specifically, technology can provide individualized quantitative measures of one's affect by tapping subtle behavioral and physiological signals that can be implicit manifestations (i.e., variation in physiological indices which cannot be easily picked up through conventional observation-based techniques) of one's affect. Thus, investigators have been employing technology to measure one's behavioral looking pattern quantified as Fixation Duration (FD) and gaze-related physiological indices, such as Pupil diameter (PD) and Blink rate (BR) when exposed to social components of interaction.

1.1 Related work

Literature indicates that one's FD (Kliemann et al. 2010, Bekele et al. 2014), Fixation Frequency, Average Fixation Duration (AFD) (Tsang 2018) and other gaze related parameters (Frazier et al. 2018; Bours et al. 2018) can provide quantitative estimates of one's behavioral looking pattern towards a social stimulus having positive and negative emotion-related aspects. In these studies, the authors report that such social stimuli can have different implications (that is impact) on these indices as far as ASD and their Typically Developing (TD) counterparts are concerned in both Virtual Reality (VR) (Bekele et al. 2014) and non-VR based Studies (Tsang 2018; Matsuda et al. 2015). Also, the study by Wang et al. (Wang et al. 2018) showed that even at an early age, the children with ASD display signs of avoiding negative emotional expressions as measured from their fixation pattern. Again, various investigators (Frazier et al. 2018; Bours et al. 2018) have reported the use of one's fixation pattern to assist in indexing the characteristics of Autism while the target group process emotional faces of social communicators.

In our present work, we refer to these indices as one's gaze-based behavioral indices (BIs henceforth). Additionally, the physiology-related indices (such as PD and BR) have also been shown to be differentiated between the ASD and TD groups when exposed to different emotional expressions (Aracena et al. 2015; Tharp et al. 2015; Vabalas and Freeth 2016). In this work, we refer to PD and BR as physiological indices (PIs henceforth). These previous studies show the importance of BIs and PIs as far as differentiating the implications of social stimuli such as emotional expressions. However, a question remains, whether these BIs and PIs, either in isolation or combination, can be used as quantitative metrics to understand the impact of emotional components on one's behavioral and physiological indices. Also, we wanted to explore the possibility of using the BIs and PIs to differentiate the experiences of the ASD from that of TD while being exposed to varying emotional expressions. This will be feasible if we can apply classification tools while exploiting the BIs and the PIs.

Evidence from the literature shows the importance of classifying one's gaze-related indices such as, fixation patterns (Yoon et al. 2014) and pupil dilation (Nugrahaningsih and Porta 2014) for unique identification of an individual through non-invasive soft biometrics. These studies show that one's fixation pattern (Yoon et al. 2014) and PD (Nugrahaningsih and Porta 2014) while looking towards static face images without any emotional expression can be used to identify an individual with 76% accuracy using Hidden Markov Model and ~ 97% accuracy using Support Vector Machine, Neural Networks (NN) and Bayesian Networks (BNT). Again, cognitive research studies show that one's fixation pattern (Schurgin et al. 2014) and PD (Aracena et al. 2015) in response to static faces with emotional expressions can be used as potential inputs to classifiers thereby segregating the biomarkers based on emotions. The authors used fixation pattern as input to BNT (Schurgin et al. 2014) and PD as input to combination of NN and Decision Trees (Aracena et al. 2015) to distinguish the implications of emotional expressions on one's gaze-related indices with an accuracy of ~ 26% and ~ 54%, respectively. These studies show the potential of BIs and PIs in cognitive studies. However, these studies considered only TD individuals focussing on developing applications related to person identification and not for individuals with ASD. Also, these studies have used only static 2D face images (with and without emotional expressions) presented as visual stimuli and were not embedded in realistic social contexts presented as visual stimuli. Similar is the case for another study by Wang et al. 2018 where the researchers used static 2D emotional pictures to explore the difference in one's fixation pattern (without application of classification tools on the gaze-related indices) upon exposure to the emotional stimuli for both ASD and TD groups of participants. These studies, though promising have used static stimuli that is different from real-life situations. Specifically, in reality, often the social characters (displaying emotional expressions) that are embedded in social contexts, are not static. Designing and projecting realistic social situations with social characters that can move about dynamically within a visual scene can be easily accomplished by using technology-assisted platforms. Thus, exploration of (1) technology-based social communication task platform with real-time access to one's BIs and PIs along with (2) classifiers that can use the BIs and PIs to distinguish the implications of context-specific emotional expressions within and across ASD and TD groups warrants further investigation.

1.2 Contribution and objectives

For our present work, we have used Virtual Reality (VR) based social task platform (similar to that used in our previous study (Babu et al. 2018)). Using this platform, we exposed the participants with ASD (i.e., ASD group) and their TD counterpart (i.e., TD group) to simulated social situations. In this, the participants interacted with virtual characters (avatars *henceforth*) that moved dynamically within the visual scene and demonstrated context-relevant facial emotional expressions. We used VR, since VR offers (1) flexibility in designing realistic social scenarios (2) provision of customizing the avatars (3) safe learning environment that is in ways difficult in real-world settings, etc. (Nesterova et al. 2015). While the participants interacted with the avatars, we used a technologically-enhanced eye tracker to monitor one's PIs and BIs synched with the VR-based social tasks. We were interested to explore the use of PIs and BIs since our previous study (Babu et al. 2018) had shown the differentiated implications of facial emotional expressions demonstrated by avatars on these indices for both the ASD and TD groups. However, the study did not explore the possibility of using the PIs and BIs to classify the ASD and TD groups interacting with avatars demonstrating various emotional expressions. In the present work, we have explored the potential of PIs, BIs or combinations of PIs and BIs to perform this classification.

The objectives of our current study are twofold, namely, to (1) understand the possibility of using one's PIs and BIs either in isolation or in combination to classify the effect of avatar's emotional expressions and (2) perform an intra-group and inter-group classification that can differentiate the participants' gaze-related indices within and between the ASD and TD groups.

This paper is organized as follows: Sect. 2 presents the system design. Section 3 describes the experimental setup, methodology, and the classification techniques used. Section 4 discusses the results. Section 5 summarizes our research findings, limitations, and future scope.

2 System design

Our Gaze-sensitive VIrtual-reality based Social-skill Platform (G-VISP henceforth) comprised of (A) Virtual Reality based Social Communication (ViRSC) Task Module and (B) Gaze Data Acquisition (GDA) Module.

2.1 Virtual reality based social communication (ViRSC) task module

The ViRSC tasks of G-VISP were developed using Python-based Vizard software (from Worldviz Inc.). The VR environments built and exported from Google SketchUp (http://www.sketchup.com), were designed to present daily-life social situations belonging to three categories, namely, (1) *closed environment within home* (less crowded, e.g. dinner party at home, Fig. 1a), (2) *closed environment outside home* (partially crowded, e.g. restaurant, Fig. 1b) and (3) *open environment* (crowded, e.g. cricket stadium, Fig. 1c). Additionally, we chose avatars capable of demonstrating context-relevant emotions (e.g. Happy, Angry and Neutral) similar to that used in another study (Kuriakose and Lahiri 2015) based on a survey. The survey was conducted with 14 undergraduate student volunteers (mean $\pm$ SD $=$ 21.4 $\pm$ 3.6 years) from our institute. From the survey, we chose six avatar faces (demonstrating Happy, Angry and Neutral expressions) for which the emotional expressions were successfully recognized by the volunteers matching with that as intended by the designer. The avatars could speak, blink, walk inside the VR environment, and exhibit a balanced mixture of straight and averted gaze (Babu et al. 2018).

The ViRSC tasks (nine tasks, three in each category) were presented in our usability study. Each task consisted of two sessions, namely, (1) *story* and (2) *conversation sessions*. In the *story session*, the avatar narrated his experience of a social event in the form of a short story along with context-relevant emotional expression. This was followed by a *conversation session* in which the participants were expected to respond to a set of five questions (Table 1). The participants used a keypad to respond through a menu-driven interface (Babu et al. 2018).

2.2 Gaze data acquisition (GDA) module

The GDA module of G-VISP was developed using ViewPointEyeTracker-2.9.2.5 (from Arrington Research, Inc.). We acquired the participants' (1) 2D gaze coordinates, (2) FD, (3) PD and (4) pupil major axis (PMA). Our GDA algorithm processed these data to extract *PIs*, such as Mean Pupil Diameter (MEAN_PD) and Mean Blink Rate (MEAN_BR). Also, we extracted *BIs*, such as, total fixation duration

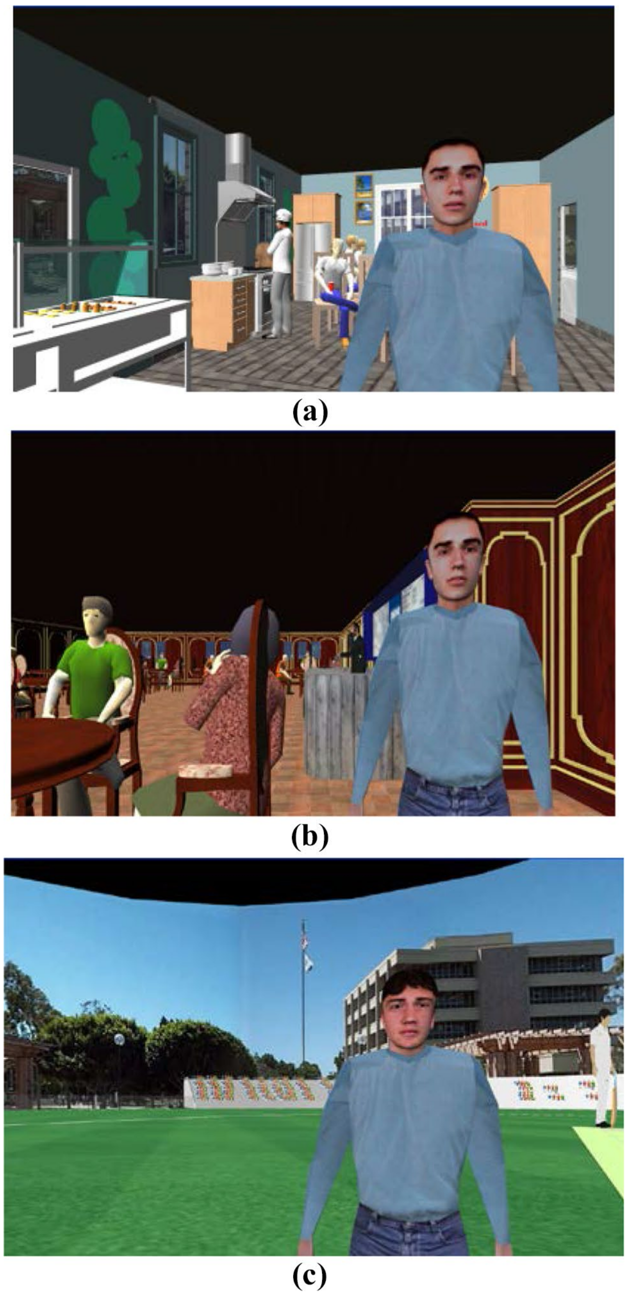

Fig. 1 VR-based social scenarios

(FD) and average FD (AFD), namely, FD_{FACE}, FD_{OTHERS}, AFD_{FACE} and AFD_{OTHERS} for each of the two regions of interest (ROIs), e.g., face region (FACE_ROI) and regions outside face (OTHERS_ROI) [for more details on the data extraction please refer (Babu et al. 2018)].

3 Procedure

3.1 Participants

For our usability study, we recruited nine individuals each from high-functioning ASD and TD groups (Table 2) with eight of them in each group (ASD1-ASD8 and TD1-TD8) being the same as that in our previous study. We used Social Responsive Scale (SRS) (Coon et al. 2010) and Social Communication Questionnaire (SCQ) (Chandler et al. 2007) to assess the participants' level of autism functioning. Additionally, we used Spence Children's Anxiety Scale (SCAS) (Spence et al. 2003) to estimate their anxiety level. We assessed the anxiety level (SCAS) of the participants before taking part in our study so as to ensure that our participants were not highly anxious. This is because, increased anxiety might adversely affect one's ability to take part in a social task. All the TD participants were well below the clinical thresholds (cut-offs) on all the three measures. All the participants with ASD (except ASD8 and ASD9) were above clinical thresholds in either/both of SRS and SCQ measures. For ASD8, we got a medical report from a reputed hospital which used Individuals with Disabilities Education Act (IDEA) scale (Brock et al. 2004) that indicated that he was in the clinical range. Also, ASD8 had the highest SCAS score. Again, ASD9 was reported to be in the Mild to Moderate Autism range by the Childhood Autism Rating Scale 2nd Edition-High Functioning (Access Policy Practice Advice: Autism Spectrum Disorder 2011). Please note that we did not have access to his SRS, SCQ and SCAS ratings. Also, all the participants with ASD were either below or marginally close to the clinical threshold of SCAS. All the participants were enrolled in special needs school. We did not have any access to their IQ scores. However, the participants included in our study possessed IQ above average as reported by their teachers based on the teachers' impression on their IQ level. Additionally, all the participants were comfortable in understanding and conversing in English.

Table 1 Categories of the Questions

Types	Description
Context relevant (CR)	Directly related to the social context directly narrated by avatar
Projected contingent (PC)	Indirect association to the situation of the story that has not been addressed directly by the avatar
Emotion recognition (ER)	Aimed at identifying the non-verbal social cues (emotions) demonstrated by the avatars

Table 2 Participant's Characteristics

ID (Gender)	AGE	SCAS Cut-off = 59	SRS Cut-off = 60	SCQ Cut-off = 15
ASD1 (male)	12	58	61	11
ASD2 (female)	15	55	75	18
ASD3 (female)	19	60	76	14
ASD4 (female)	10	55	70	8
ASD5 (male)	16	57	65	18
ASD6 (male)	19	56	75	22
ASD7 (female)	12	52	73	9
ASD8 (male)	15	61	58	10
ASD9 (male)	12	–	–	–
M (SD)	14.44 ± 3.21	56.75 ± 2.92	69.12 ± 6.96	13.75 ± 5.09
TD1 (male)	18	49	48	5
TD2 (female)	13	40	48	6
TD3 (male)	12	44	51	3
TD4 (male)	20	53	49	6
TD5 (male)	19	46	51	5
TD6 (male)	12	54	53	7
TD7 (male)	17	44	48	6
TD8 (male)	14	50	49	5
TD9 (male)	13	50	51	6
M (SD)	15.33 ± 3.16	47.78 ± 4.6	49.78 ± 1.78	5.44 ± 1.13

The values of the SRS and SCQ scores that are higher than the cut-offs indicate that one is clinically Autistic. The value of SCAS score that is higher than the mentioned cut-off indicates that one is anxious

3.2 Experimental setup

Our usability study needed a commitment of approximately 1 h from each volunteer. The ViRSC tasks were presented on a 17″ monitor. The experimenter invited the participant and introduced him to the experimental setup comprising of Eye Tracker goggles, task computer and a keypad. A visual schedule was used to give a brief description of the ViRSC tasks. Also, the participant was told that he was free to withdraw from the study at any time if he felt uncomfortable and asked to sign a consent form. In the case of individuals with ASD, the caregivers were asked to sign the consent form. Then the participant was asked to put his chin on a chin-rest (in-house built) kept at a distance of about 50 cm from the task monitor. The chin-rest height was adjusted so that his eyes were aligned with the center of the monitor. The experimenter helped the participant to wear the Eye Tracker goggles and executed 16-point calibration. Post calibration, we acquired about 3 min of baseline gaze data. During Baseline a white screen was presented on the task computer monitor accompanied with a mild music and the participants were asked to relax. Subsequently, the G-VISP presented the ViRSC tasks with the GDA module acquiring the real-time gaze data in a time-synched manner.

3.3 Computation of change in eye-physiological indices w.r.t. baseline

Our GDA module computed the participant's MEAN_PD and MEAN_BR during each *story session* of the ViRSC tasks. We calculated baseline measures of PD (BASE_PD) and BR (BASE_BR). Subsequently, we computed the %Δ in the *PIs* w.r.t. baseline using Eqs. (1) and (2).

$$\Delta MEAN\_BR\% = \frac{(MEAN\_BR - BASE\_BR)}{(BASE\_BR)} * 100 \quad (1)$$

$$\Delta MEAN\_PD\% = \frac{(MEAN\_PD - BASE\_PD)}{(BASE\_PD)} * 100 \quad (2)$$

3.4 Computation of behavioral indices

The data on fixation pattern acquired by GDA was processed to extract (1) %FD (out of the total duration of *story session*) [Eq. (3)] for each ROI and (2) Non-Face to Face Ratio (NFFR) [e.g. (4)]. The NFFR was computed since individuals with ASD often fixate more towards non-social cues (OTHERS_ROI) than communicator's face (FACE_ROI) (Kliemann et al. 2010).

$$NFFR = (\%FD_{OTHERS} / \%FD_{FACE}) * 100 \quad (3)$$

$$\%FD_i = (FD_i/Session\ duration) * 100 \tag{4}$$

where i represents the ROI.

3.5 Statistical techniques used

We used 2×3 Mixed ANOVA to the *PIs* and *BIs* of the two participant groups (ASD and TD) who were exposed to three different emotional expressions. Also, we computed Pearson's correlation coefficient (r) to estimate the effect size between the variations in independent variables such as participant groups and the emotions.

3.6 Classification tools used

We designed Gaze Classifiers (*GCs henceforth*) using popular classification tools, such as Classification Tree (CT), Regression Tree (RT), Bayesian Network (BNT) and Support Vector Machine (SVM). One of the recent studies has examined the applicability of these classification tools in detection of emotional state along with its evaluation of Model Accuracy, Training Speed, Classification Speed and Tolerance to Noise (Behoora and Tucker, 2015). The authors report that the BNT, CT and RT models have higher training and classification speeds making these suitable for online training applications with less training samples, but with some compromise in accuracy and tolerance to noise. On the other hand, SVM was reported to have higher accuracy and classification speed, moderate tolerance to noise and worst training speed. Thus, SVM could be a good choice for offline training platforms. In our present research, we have applied these different classification tools to the *PIs* (*ΔMEAN_PD%* and *ΔMEAN_BR%*) and *BIs* (AFD_{FACE}, AFD_{OTHERS}, $\%FD_{FACE}$, $\%FD_{OTHERS}$ and *NFFR*) extracted during the *story session* of the ViRSC tasks.

3.6.1 Classification trees (CT) based classifier

Generally, in tree-based models, the data samples are partitioned recursively into smaller rectangles resulting in a binary tree. The root node of the binary tree represents a single input variable (x_i) with a split point. The output variable (y_i) which acts as a predictor variable is placed at the leaf node. In this study, we have chosen a Classification and Regression Tree (CART) model (Breiman 2017), to develop a CT, which selects the split point using a greedy algorithm to minimize the cost function (Gini index). For 'N_x' number of observations for a region 'R_x', the proportion 'P_{xc}' (number of instances for training data of a class 'c' to be observed in node x) is represented as,

$$P_{xc} = 1/N_x\left(\sum\nolimits_{x_i \in R_x} I(y_i = c)\right) \tag{5}$$

Then the split criterion using the Gini index for the CT is,

$$Gini = \sum\nolimits_c P_{xc} * (1 - P_{xc}) \tag{6}$$

The resultant value of the Gini index indicates the purity of the leaf nodes. Gini = 0 indicates that the node has perfect purity with the same class type. If Gini = 0.5, then it takes an equal split of classes in a binary tree.

3.6.2 Regression trees (RT) based classifier

For the RT with node x, with region as R_x and N_x as number of observations, the predicted output is,

$$P_{xc} = 1/N_x \sum\nolimits_{i \in R_x} y_i \tag{7}$$

Here we used the Mean Squared Error (MSE) split criterion as cost function to split the training samples using Eq. (8),

$$MSE = 1/N_x \sum\nolimits_{i \in R_x} (y_i - P_{xc})^2 \tag{8}$$

where y_i is the output variable, and P_{xc} is the predicted value of the tree for a particular rectangle R_x associated with node x. The split point is chosen based on the minimum value of R_x.

3.6.3 Bayesian networks (BNT) based classifier

A BNT is a directed acyclic graph (DAG) in which each node is a random variable (R), represented by its quantitative probability information. Each parent node has a Conditional Probability Distribution (CPD), $P_i(node_i, parent(node_i))$, that quantifies the effect on its children nodes and Joint Probability Distribution (*JPD*) for a set of random variables (Russell and Norvig 2016) are computed. The BNT is given as $BN = (G, \Phi)$, where G is the DAG with a set of random variables ($R = A_1, A_2, \ldots A_k$) representing a set of nodes, having edges with direct dependencies between each other. The Φ accounts for the parameter, JPD [P_{BN} using Eq. (9)] over the set R.

$$P_{BN}(A_1, A_2, \ldots, A_k) = \prod\nolimits_{i=1}^{k} P_{BN}(A_i|\theta_i) \tag{9}$$

where, A_i (Child node) is conditioned on Θ_i (Parent node), the set of parents nodes of A_i.

3.6.4 SVM based classifier with iterative sigma tuning

SVM discriminates two categories of data by defining a hyperplane that separates them in multi-dimensional space (Cortes and Vapnik 1995). Since, SVM is a two-class classifier, in our research study we have designed *GCs* (Fig. 2) that include *n(n-1)* number of SVMs. Here the *GCs* classified *PIs* and *BIs* and were of three types, namely, (1) GC_H (Happy vs rest of the emotions), (2) GC_A (Angry vs rest of

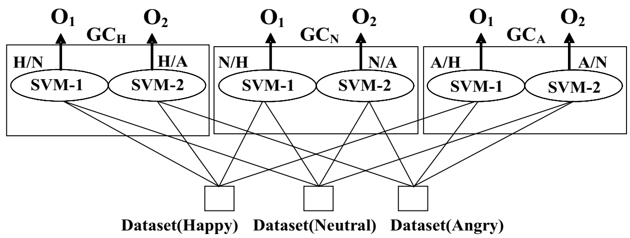

Fig. 2 Representation of Multi-class Classifier using SVM

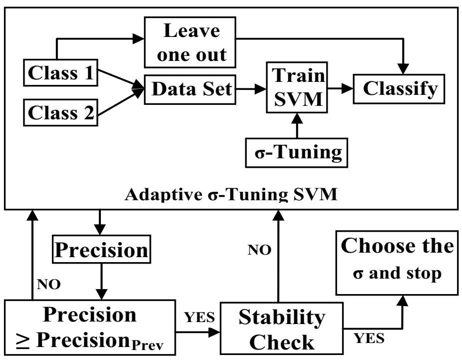

Fig. 3 Workflow of sigma tuning in Gaze Classifiers

the emotions), and (3) GC_N (Neutral vs rest of the emotions). Thus, each *GC* consisted of two SVMs (Fig. 2) to classify one's *PIs* and *BIs* corresponding to one emotion from that of the other two, that in turn acts as a multi-class SVM (Cheong et al. 2004). For example, *SVM-1* and *SVM-2* of GC_H classified participants' *PIs* and *BIs* corresponding to Happy from Angry and Neutral, respectively. Finally, the result (O_1, O_2) generated from each *GC* served as the classification accuracy of the two SVMs validated using Leave-One-Out (LOO) approach (Elisseeff and Pontil 2003).

The SVMs used Radial Basis Function (RBF) kernel [represented by Eq. (10)]. The σ value of the RBF kernel decides the possibility of classifying an unpredictable n-dimensional problem (Han et al. 2013). An optimal σ value can give better classification accuracy. When the number of training samples and features of a classification problem such as gaze data are not prefixed, the selection of optimum σ value often becomes challenging. Thus, we propose a σ-tuning approach that adapts to changes in training samples similar to the σ-tuning used in variable selection in another study (Han et al. 2013). The working principle of a σ-tuning algorithm for each SVM of *GC* is shown in Fig. 3. The initial step of the the algorithm is to find optimal σ values (say, σ_1 and σ_2) for the two SVMs. First, we trained the SVMs and validated using LOO approach and estimated the precision for one class, say, Class 1. The test data was taken only from Class 1 during the LOO validation and not from Class 2 (say) to make the SVM appropriate to Class 1. The algorithm executed iteratively (σ-tuning) with various σ values while comparing the precision of Class 1 with the previous iteration. A Stability Check was performed to verify whether the change in precision values was significant. Once the precision of Class 1 reached maximum or the difference was negligible for few subsequent iterations, the *Stability Check* stopped the iteration. Finally, the respective σ values for both the SVMs, such as σ_1 and σ_2 were chosen.

The RFB kernel function *k* of the SVM is given as,

$$k_i(v_1, v_2) = \left(-\left(\|v_1 - v_2\|^2 / 2\sigma_i^2\right)\right) \quad (10)$$

where v_1 and v_2 were the feature vectors of a particular class, k_i was the RBF kernel function for the i^{th} SVM with respective σ_i value. Then the output (O_i) of the i^{th} SVM depended on the identified supporting vectors SV_i, the weight w_i and the bias *b*, using the kernel k_i as given in (11),

$$O_i = \sum_i w_i k_i(SV_i, C) + b \quad (11)$$

where *C* is the dataset for a particular class of emotion.

Finally, the pair of accuracies (O_1, O_2) were compared to attain the minimal accuracy (to ensure maximum distinguishability) with which one class of emotion can be differentiated from the other two emotions.

4 Result and discussion

We designed a preliminary usability study with nine pairs of ASD and TD participants. Here, we wanted to understand whether the participants' looking pattern varied with avatars narrating personal experiences coupled with context-relevant emotional expressions. In case the looking patterns varied across emotional expressions for the ASD and TD groups, we wanted to study the possibility of using one's *PIs* and *BIs* either in isolation or in combination (*PIandBIs henceforth*) to classify the effect of implicit manifestation (i.e., variation in the gaze-related indices which cannot be easily picked up through manual observation) due to exposure to avatar's emotions. Subsequently, we carried out an intra-group and inter-group classification that can differentiate the participants' gaze-related indices within and between the ASD and TD groups.

4.1 System acceptability

At the end of the study, we carried out an exit survey among the participants to know their views on our system. The survey was conducted to record the participant's views about our system using a binary scale i.e., 'yes' or 'no'. All the participants expressed their interest in interacting with G-VISP.

None of the participants faced any problem using the eye tracker goggles and the chin-rest in our study. Everyone expressed that they liked interacting with our avatars in the ViRSC tasks during the story session and did not experience any difficulty in responding to the questions asked by the avatar during the conversation session. Additionally, some of the participants mentioned that the avatars looked realistic. Also they were interested in future participation in such studies. When asked regarding the reason behind their willingness for future participation in our study, some participants commented that they liked interacting with our system. This response was qualitative in nature. None of the participant reported any concerns related to their usage of our system. From this we can infer that the G-VISP has a potential of being accepted by the target group.

4.2 Analysis of participants' looking pattern

Figure 4 shows the fixation pattern of one participant each from ASD (Figs. 4a, b) and TD (Figs. 4c, d) while exposed to avatars' Happy and Angry emotions, respectively. As expected, the TD participant had fixated more towards the avatar's FACE_ROI than to OTHERS_ROI. In contrast, the ASD participant exhibited atypical gaze behavior by fixating more towards OTHERS_ROI than to FACE_ROI (similar to the reports by Vabalas and Freeth 2016). Specifically, the cumulative %FD (for FACE_ROI and OTHERS_ROI) was very less for ASD group (~31.1 ± 11.27% for Happy and 31.3 ± 9.88% for Angry) than that for TD group (~99.4 ± 0.45% for Happy and 97.3 ± 2.02% for Angry). This also infers that the participants with ASD looked outside the visual stimulus for a longer duration as compared to the TD group. Additionally, we performed an independent sample *t* test on the cumulative %FD of the ASD and TD groups while they were exposed to avatars demonstrating Happy and Angry emotions. The %FD of the groups was found to be statistically significantly different (p value $<$ 0.00001) for both the Happy and Angry Emotions.

4.3 Feasibility of PIs and BIs to classify the effect of implicit manifestation due to exposure to avatar's emotions

Previous research study (Babu et al. 2018) showed that G-VISP had a potential to have differentiated implications on participant's gaze-related indices in response to varying emotional expressions as far as the ASD and TD groups were concerned. We were interested to understand whether these distinctive implications can be harnessed through classification techniques by using one's gaze-related indices (*PIs* and *BIs*) to identify intra-group and inter-group differences of the ASD and TD groups. Also, we planned to explore whether the *PIs* and/or *BI*s individually or in combination

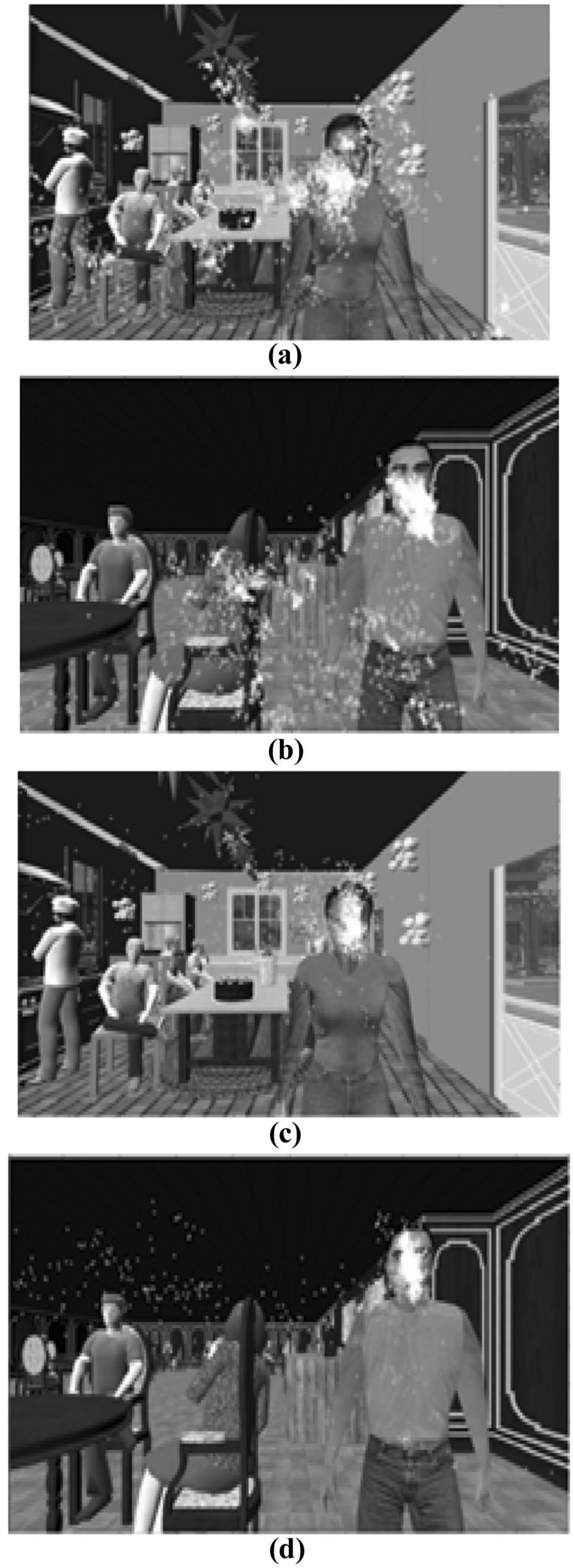

Fig. 4 Gaze pattern due to (**a**) Happy and (**b**) Angry emotions for ASD and (**c**) Happy and (**d**) Angry emotions for TD

can contribute to such classification. In turn, this can pave the path for gaze-related *PI* and *BI*-based affect recognizers. Please note that we considered the change in the value of *PIs* with respect to Baseline (Sect. 3.3) in order to offset the variability due to different Baseline values of the participants.

4.3.1 Intra-group classification with regard to varying emotions

In this section, we present our observations of intra-group differences by using the different *GCs* (Sect. 3.6) trained with participants' *PIs* and *BIs* corresponding to various emotions for the ASD and TD groups. The *PIs* and *BIs* computed for the ASD and TD groups were differentiated across the groups and across emotional expressions. Specifically, for the TD group, the % change in the *PIs* on an average were 36.75 ± 6.4, 31.16 ± 6.78 and 39.11 ± 5.85 for Happy, Angry and Neutral expressions, respectively. Again, that for the *BIs*, the values represented in percentage were 42.11 ± 3.83, 32.47 ± 3.15 and 46.61 ± 4.72 for Happy, Angry and Neutral expressions, respectively. In contrast, for the ASD group, the % change in the *PIs* on an average were 0.17 ± 6.43, 18.77 ± 7.17 and 9.15 ± 6.53 for Happy, Angry and Neutral expressions, respectively. Again, that for the *BIs*, the values were 18.08 ± 4.54, 15.72 ± 3.66 and 19.61 ± 5.20 for Happy, Angry and Neutral expressions, respectively.

The classification was done using leave-one-out approach for all the classifiers such as SVM, CT, RT and BNT. The results (Figs. 5a, b) indicate that our σ-tuning approach for SVM while using *PIandBI*s, out-performed all the other classifiers with the highest classification accuracy (with a minimum of 97% for ASD and 97.6% for TD group) for Happy, Angry and Neutral emotions. With CT, we achieved second highest accuracy (with a minimum of 85.3% for ASD and 87.7% for TD). For the RT and BNT, the classification accuracy was less (< 60%). Also, we trained the classifiers with *PIs* and *BIs* in isolation that resulted in reduced %accuracy for various emotions irrespective of the type of classifiers (Fig. 5). Having observed improved classification accuracy using *PIandBIs*, we wanted to verify our observation by computing the Precision-Recall Curve (PRC) for the SVM Classifier. The PRC evaluates the fraction of true positives among the positive predictions. For example, in our study, if the *PIandBIs* of an individual corresponding to avatars' Happy emotion was classified as Happy, then this was referred to as True Positive, else, it was False Positive. Figure 6 shows PRC for varying σ values of the SVM while using the *PIs and BIs* in isolation and together (*PIandBIs*). To estimate the performance of SVM, we estimated the Area Under Curve (AUC) of the PRC, since increased area infers greater confidence in future classification performance even with gaze data having variations (Saito and Rehmsmeier 2015). From Fig. 6, we find that the composite effect of

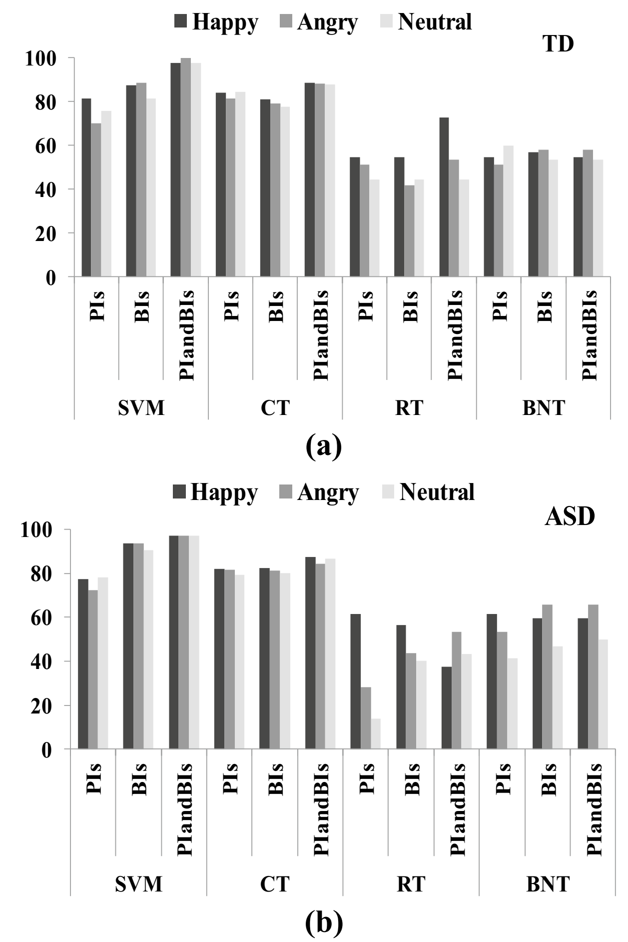

Fig. 5 %Accuracy of the classifiers for avatars' various emotional expression for **a** TD group and **b** ASD group

*PIandBI*s out-performed the cases of the *PIs* and *BIs* used in isolation for both the groups.

4.3.2 Inter-group classification with regard to varying emotions

Having seen that the classification accuracy improved with the composite effect of *PIandBIs* within each group (ASD and TD), we wanted to understand whether the same holds good while classifying the gaze pattern of individuals with ASD from that of their TD counterpart. We trained the *GCs* with gaze-related indices with the two classes as those for ASD group and those for the TD group to carry out inter-group classification for the three emotions. The confusion matrices (Fig. 7a) indicates that %accuracy was lesser while the *GCs* used isolated *PIs* that that with isolated *BI*s or *PIandBIs* having same AUC in the PRC of the SVM classifier (Fig. 7b) for all the three emotions. Also, the *GCs* with RT and BNT classifiers showed better %accuracy (Fig. 7a)

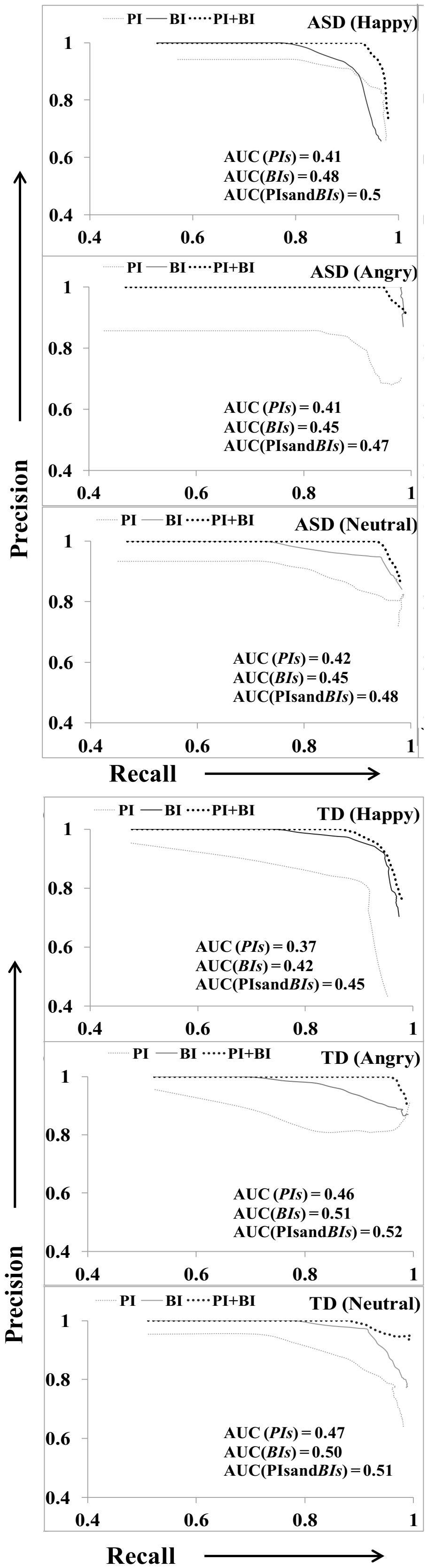

Fig. 6 PRC for Intra groups of ASD and TD

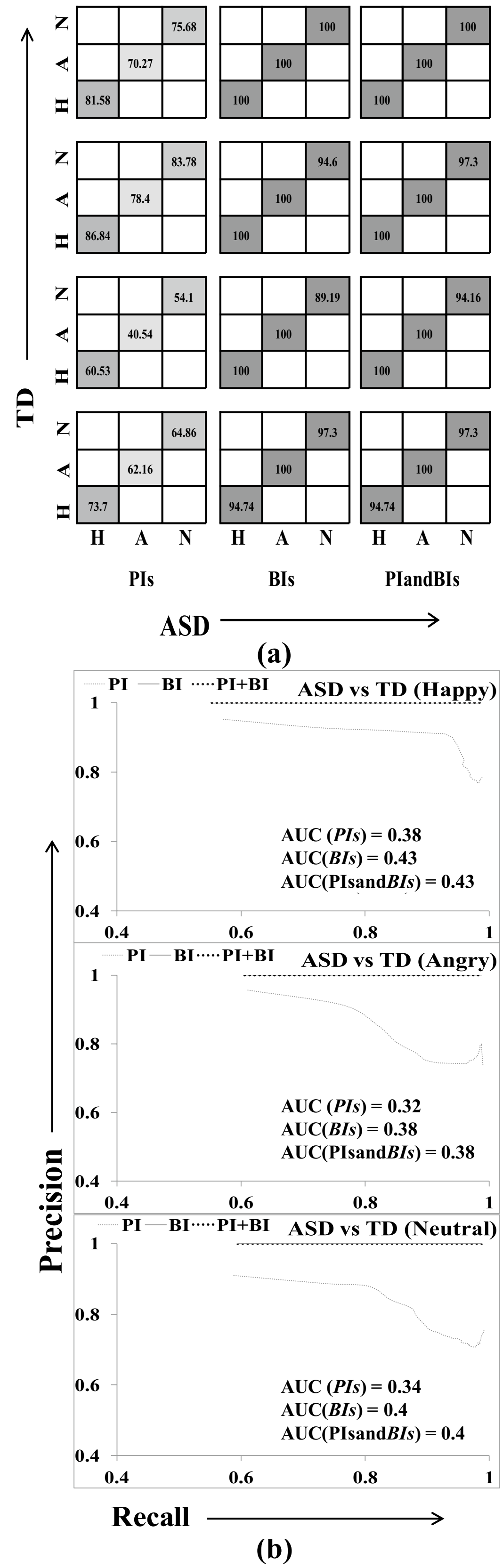

Fig. 7 Inter-group classification **a** %Accuracy and **b** PRC of the classifier. (Note: In Fig. (**a**) H = Happy, A = Angry, N = Neutral and in Fig. (**b**) The PRC curves of the '*BIs*' and '*PIandBIs*' are overlapped)

in the case of inter-group classification while using *BI*s and *PIandBIs* compared to that for intra-group classification.

4.4 Statistical analysis and discussion

We carried out 2x3 Mixed ANOVA (Sect. 3.5) considering all the dependent variables (*PIs* and *BIs)*, such as PD, BR, FD_{FACE}, AFD_{FACE}, FD_{OTHERS} and AFD_{OTHERS} corresponding to all the emotional expressions. As far as the *PIs* were concerned, for the PD there was no significant difference with a small effect ($F(1,16)=0.85$ $r=0.22$) due to emotions within participant groups (intra-group). However, there was marginal statistical difference (p value $=0.08$) on PD with large effect ($F(1,16)=3.5$, $r=0.42$) due to emotions for inter-group analysis. Again, change in *BR* due to emotions did not reflect significant difference with no effect for intra-group analysis. But, there was moderate effect ($F(1,16)=2.35$, $r=0.36$) with no statistical significance for inter-group data analysis as far as BR was concerned.

In contrast, for the *BIs*, such as FD, namely, $\%FD_{FACE}$, we found a significant difference with high effect for intra-group (p value <0.05, $F(1,16)=4.6$, $r=0.47$) and inter-group (p value <0.05, $F(1,16)=182.9$, $r=0.96$) analyses in response to avatars' emotions. No statistically significant distinction was observed for $\%FD_{OTHERS}$ both for intra-group and inter-group data. But, there was moderate effect of emotions on $\%FD_{OTHERS}$ ($F(1,16)=2.8$, $r=0.4$) only for inter-group data. Again, there was moderate effect of emotions for intra-group computation on AFD_{FACE} ($F(1,16)=2.04$, $r=0.33$) without statistical significance. But, for AFD_{FACE} computation for inter-group analysis, there was a significant difference (p value <0.05) with high effect ($F(1,16)=11.54$, $r=0.65$). For AFD_{OTHERS}, both intra- and inter-groups had no statistical difference along with small effect for intra-group ($F(1,16)=1.01$, $r=0.24$) and inter-group ($F(1,16)=0.55$, $r=0.18$) data.

As regards the intra-group analysis, we can infer that the emotional expressions had significant contribution with high effect on at least some of the *BIs* and a small effect on at least some of the *PIs*. Since, there was no significant contribution on either all of the *PIs* or the *BIs*, the composite effect of combined *PIs* and *BIs* helped to achieve higher classification accuracy. Thus, during intra-group classification, the *PIandBIs* was crucial to distinguish the implicit manifestations due to exposure to various emotions, as is evident from classification accuracies (Sect. 4.3.1).

However, we see a different picture for the inter-group classification. In this, we found that the emotional expressions had significant contribution with high effect on all of the *BIs* associated with the FACE_ROI and moderate effects on the *PIs*. Thus, this might have resulted in the *BIs* to contribute to high classification accuracy that was also shown by the *PIandBIs* (Sect. 4.3.2), thereby making the inclusion of *PIs* in inter-group classification not mandatory. Also, the moderate effect of the *PIs* might have contributed to reduced classification accuracy compared with *BIs* in isolation (Sect. 4.3.2). This fact might infer that the *BIs* can have a stronger contribution to the classification accuracy than the *PIs* as far as inter-group analysis was concerned. From this, one can possibly say that the physiological manifestations in response to emotional expressions are not distinctly different between the ASD and TD groups. Thus, though individuals with ASD have restrictions in making explicit expressions of their affective state (Begeer et al. 2008), one's physiological indices can be used as indicators of affective state, similar to that in the literature (Tharp et al. 2015). Here this has been shown to hold good even for VR-based realistic social situations having social communicators who can move about dynamically in the visual scene and narrate social experiences accompanied with context-relevant emotional expressions.

5 Conclusion

In our current research work, we have designed and developed a Gaze-Sensitive VIrtual Reality-based Social-skill Platform (G-VISP). It featured a gaze-based data acquisition (GDA) module that acquired the participants' gaze data while they were exposed to VR-based (ViRSC) tasks with virtual peers (avatars) demonstrating context-relevant facial emotional expressions. Additionally, we have designed *Gaze Classifiers* (*GCs*) to classify the participants' gaze patterns corresponding to three different emotional expressions (Happy, Angry and Neutral). Two groups of participants (ASD and TD) took part in a usability study. The *GCs* for the intra-group classification using Support Vector Machine (SVM) and Classification Tree (CT) showed higher classification accuracy while using composite effect on eye physiological (*PI*s) and behavioral (*BI*s) indices. Again, in the case of inter-group classification, we found that isolated contribution of *BIs* and composite effect of *PI*s and *BI*s have the potential to classify the gaze patterns of individuals with ASD from that of their TD counterpart.

Though the results were promising, our study had a few limitations. Currently, the G-VISP had an inventory of only nine ViRSC tasks that exposed the participants for a limited duration. In future, we are planning to increase the inventory of ViRSC tasks that can expose the participants to more number of social scenarios for a prolonged duration. Another limitation of our present study is limited sample size. However, since this study was not designed as an intervention platform at this stage of research, but instead to show the feasibility of our technology-assisted system, we did not aim for a larger sample size. In future, we plan to enroll a larger sample size for a full-fledged intervention study. Also, our *GCs,* though developed for offline analysis, can be used for online classification purposes, since we are acquiring real-time gaze data. However, in our present study, our aim was to understand the feasibility of developing *GCs*. Thus, we have not integrated the *GCs* with the G-VISP for online classification. In future, we plan to integrate the *GCs* with G-VISP, so that online classification of gaze data is possible. This can help the therapist to make faster decisions regarding modulation of intervention paradigm in an individualized manner with regard to emotion recognition tasks.

In future, a more in-depth longitudinal study incorporating larger participant pool is required before such a platform can be deployed in intervention settings. Also, questions on the optimum duration of exposure of the participants to G-VISP remain. However, we believe that such an integrated system can be potent to pave a way towards designing a powerful complementary tool in the hands of the therapists/interventionists where one therapist can facilitate multiple individuals at the same time. Thus, such a system can hold promise in improving the quality of life for individuals with ASD.

Acknowledgements The authors wish to thank the Pearl Special Needs Foundation, Ahmedabad and KGP Hospital, Baroda for helping us in enrolling the participants in our study. Also we thank Cognitive Science Research Initiative under DST, India for partially funding this research. We would also like to express our gratitude to Ministry of Electronics and Information Technology, India for offering fellowship support to the researcher under Visvesvaraya PhD Scheme.

References

Access Policy Practice Advice: Autism Spectrum Disorder (2011) http://www.dhs.vic.gov.au/__data/assets/pdf_file/0010/681247/dsd_cis_access_policy_practice_advice_utism_spectrum_disorder_1011.pdf. Retrieved 1 May 2016

American Psychiatric Association (APA) (2013) Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub

Aracena C, Basterrech S, Snáel V, Velásquez J (2015) Neural networks for emotion recognition based on eye tracking data. In: 2015 IEEE international conference on systems, man, and cybernetics, IEEE, pp 2632-2637

Autism cases rise in last two decades (2017) The Times of India. https://timesofindia.indiatimes.com/city/visakhapatnam/autism-cases-rise-in-last-two-decades/articleshow/57967289.cms. Retrieved 4 July 2018

Babu PRK, Oza P, Lahiri U (2018) Gaze-sensitive virtual reality based social communication platform for individuals with autism. IEEE Trans Affect Comput 9(4):450–462

Begeer S, Koot HM, Rieffe C, Terwogt MM, Stegge H (2008) Emotional competence in children with autism: diagnostic criteria and empirical evidence. Dev Rev 28(3):342–369

Behoora I, Tucker CS (2015) Machine learning classification of design team members' body language patterns for real time emotional state detection. Des Stud 39:100–127

Bekele E, Crittendon J, Zheng Z, Swanson A, Weitlauf A, Warren Z, Sarkar N (2014) Assessing the utility of a virtual environment for enhancing facial affect recognition in adolescents with autism. J Autism Dev Disord 44(7):1641–1650

Bours CCAH, Bakker-Huvenaars MJ, Tramper J, Bielczyk N, Scheepers F, Nijhof KS, Buitelaar JK (2018) Emotional face recognition in male adolescents with autism spectrum disorder or disruptive behavior disorder: an eye-tracking study. Eur Child Adolesc Psychiatry 27(9):1143–1157

Breiman L (2017) Classification and regression trees. Routledge

Brock SE, Change C, Catewood V (2004) The identification of autism spectrum disorders. http://www.csus.edu/indiv/b/brocks/Workshops/NASP/Autism.pdf. Retrieved 22 June 2018

Chandler S, Charman T, Baird G, Simonoff E, Loucas TOM, Meldrum D, Pickles A (2007) Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. J Am Acad Child Adolesc Psychiatry 46(10):1324–1332

Cheong S, Oh SH, Lee SY (2004) Support vector machines with binary tree architecture for multi-class classification. Neural Inform Process Lett Rev 2(3):47–51

Coon H, Villalobos ME, Robison RJ, Camp NJ, Cannon DS, Allen-Brady K, McMahon WM (2010) Genome-wide linkage using the Social responsiveness scale in utah autism pedigrees. Molecular Autism 1(1):8

Cortes C, Vapnik V (1995) Support-vector networks. Machine learning 20(3):273–297

Data and Statistics (2018) Centers for disease control and prevention. www.cdc.gov/ncbddd/autism/data. Retrieved 10 Apr 2018

Elisseeff A, Pontil M (2003) Leave-one-out error and stability of learning algorithms with applications. NATO Sci Ser Sub Ser iii Comput Syst Sci 190:111–130

Frazier TW, Klingemier EW, Parikh S, Speer L, Strauss MS, Eng C, Youngstrom EA (2018) Development and validation of objective and quantitative eye tracking—based measures of autism risk and symptom levels. J Am Acad Child Adolesc Psychiatry 57(11):858–866

Han L, Embrechts M, Szymanski B, Sternickel K, Ross A (2013) Sigma tuning of gaussian kernels: detection of ischemia fro-m magnetocardiograms. U.S. Patent No. 8,527,435. Washington, DC: U.S. Patent and Trademark Office

Hofmann SG, Smits JA (2008) Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. J Clin Psychiatry 69(4):621

Kliemann D, Dziobek I, Hatri A, Steimke R, Heekeren HR (2010) Atypical reflexive gaze patterns on emotional faces in autism spectrum disorders. J Neurosci 30(37):12281–12287

Kuriakose S, Lahiri U (2015) Understanding the psycho-physiological implications of interaction with a virtual reality-based system in adolescents with autism: a feasibility study. IEEE Trans Neural Syst Rehabil Eng 23(4):665–675

Matsuda S, Minagawa Y, Yamamoto J (2015) Gaze behavior of children with ASD toward pictures of facial expressions. Autism Res Treat. https://doi.org/10.1155/2015/617190

Nesterova A, Aysina R, Suslova T (2015) Recent technologies to improving social and communication skills in children with asd: systematization of approaches and methods. Mod Appl Sci 9(11):38

Nugrahaningsih N, Porta M (2014) Pupil size as a biometric trait. In: Cantoni V, Dimov D, Tistarelli M (eds) Biometric authentication. BIOMET 2014. Lecture notes in computer science, vol 8897. Springer, Cham, pp 222–233

Russell SJ, Norvig P (2016) Artificial intelligence: a modern approach. Pearson Education Limited, Malaysia

Saito T, Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10(3):e0118432

Schurgin MW, Nelson J, Iida S, Ohira H, Chiao JY, Franconeri SL (2014) Eye movements during emotion recognition in faces. J Vision 14(13):14–14

Silberberg D, Arora N, Bhutani V, Durkin M, Gulati S (2013) Neurodevelopmental disorders in India–an INCLEN Study (IN6-2.001). Neurology, 80(7 Supplement), IN6-2

Spence SH, Barrett PM, Turner CM (2003) Psychometric properties of the spence children's anxiety scale with young adolescents. J Anxiety Disord 17(6):605–625

Teaching Aspergers Students Using Visual Imagery. My Aspergers Child (2007) http://www.myasperger-schild.com/2007/08/using-visual-thinking-skills-to-support.html. Retrieved 10 Aug 2016

Tharp JA, Wendelken C, Mathews CA, Marco EJ, Schreier H, Bunge SA (2015) Tourette syndrome: complementary insights from measures of cognitive control, eyeblink rate, and pupil diameter. Front Psychiatry 6:95

Tsang V (2018) Eye-tracking study on facial emotion recognition tasks in individuals with high-functioning autism spectrum disorders. Autism 22(2):161–170

Uljarevic M, Hamilton A (2013) Recognition of emotions in autism: a formal meta-analysis. J Autism Dev Disord 43(7):1517–1526

Vabalas A, Freeth M (2016) Brief report: patterns of eye movements in face to face conversation are associated with autistic traits: Evidence from a student sample. J Autism Dev Disord 46(1):305–314

Wang Q, Lu L, Zhang Q, Fang F, Zou X, Yi L (2018) Eye avoidance in young children with autism spectrum disorder is modulated by emotional facial expressions. J Abnorm Psychol 127(7):722

Watling R, Deitz J, Kanny EM, McLaughlin JF (1999) Current practice of occupational therapy for children with autism. Am J Occup Ther 53(5):498–505

Yoon HJ, Carmichael TR, Tourassi G (2014) Gaze as a biometric. In: Medical Imaging 2014: image perception, observer performance, and technology assessment, vol 9037. International Society for Optics and Photonics. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9037/903707/Gaze-as-a-biometric/10.1117/12.2044303.short?SSO=1