Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Biosignal comparison for autism assessment using machine learning models and virtual reality

Maria Eleonora Minissi [a,*], Alberto Altozano [a], Javier Marín-Morales [a], Irene Alice Chicchi Giglioli [a], Fabrizia Mantovani [b], Mariano Alcañiz [a]

[a] Instituto Universitario de Investigación en Tecnología Centrada en El Ser Humano (HUMAN-tech), Universitat Politécnica de Valencia, Valencia, Spain
[b] Centre for Studies in Communication Sciences "Luigi Anolli" (CESCOM), Department of Human Sciences for Education "Riccardo Massa'', University of Milano - Bicocca, Building U16, Via Tomas Mann, 20162, Milan, Italy

ARTICLE INFO

Keywords:
Virtual reality
Statistical machine learning
Biosignal
Autism spectrum disorder
Eye movements
Motor skills

ABSTRACT

Clinical assessment procedures encounter challenges in terms of objectivity because they rely on subjective data. Computational psychiatry proposes overcoming this limitation by introducing biosignal-based assessments able to detect clinical biomarkers, while virtual reality (VR) can offer ecological settings for measurement. Autism spectrum disorder (ASD) is a neurodevelopmental disorder where many biosignals have been tested to improve assessment procedures. However, in ASD research there is a lack of studies systematically comparing biosignals for the automatic classification of ASD when recorded simultaneously in ecological settings, and comparisons among previous studies are challenging due to methodological inconsistencies. In this study, we examined a VR screening tool consisting of four virtual scenes, and we compared machine learning models based on implicit (motor skills and eye movements) and explicit (behavioral responses) biosignals. Machine learning models were developed for each biosignal within the virtual scenes and then combined into a final model per biosignal. A linear support vector classifier with recursive feature elimination was used and tested using nested cross-validation. The final model based on motor skills exhibited the highest robustness in identifying ASD, achieving an AUC of 0.89 (*SD* = 0.08). The best behavioral model showed an AUC of 0.80, while further research is needed for the eye-movement models due to limitations with the eye-tracking glasses. These findings highlight the potential of motor skills in enhancing objectivity and reliability in the early assessment of ASD compared to other biosignals.

1. Introduction

1.1. Why use biosignals in clinical assessments

Implicit social cognition theories suggest that humans lack conscious control over many psychological and internal processes [1]. Consequently, patients may find it challenging to objectively analyze and report their behaviors during clinical and psychological assessments. Assessment procedures depend on patients' anamnesis, self-reports, and clinical observations, all of which may be susceptible to subjectivity bias, either from the clinician's side or the patient's [2,3]. To overcome these challenges, an innovative framework from computational psychiatry introduces objective measures based on human neurobiological activity. This approach applies machine learning models to neurobiological data with the goal of detecting dysfunctions underlying clinical symptoms [4]. Biosensors, such as eye tracking, galvanic skin response, electroencephalography, and functional magnetic resonance, record neurobiological signals that reflect internal processes, potentially revealing clinical biomarkers. Biomarkers are disorder biosignatures accurately provokable and measurable [5]. In contrast to explicit symptoms directly observed and reported by patients, biomarkers operate beyond individual awareness, emerging from the analysis of biosignals representing physical states and mental functions in various contexts [6].

Biosignals can be categorized into explicit and implicit types: explicit biosignals derive from explicit measures like semantic utterances and behavioral responses (e.g., scores, reaction times, accuracy), while implicit biosignals tap into human neurobiological underpinnings, such as

[*] Corresponding author. Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano (HUMAN-tech), Universitat Politécnica de Valencia, Polytechnic City of Innovation: Access N, Building 8B3rd Floor Camino de Vera s/n 46022, Valencia, Spain.
E-mail address: meminiss@htech.upv.es (M.E. Minissi).

https://doi.org/10.1016/j.compbiomed.2024.108194
Received 5 September 2023; Received in revised form 8 February 2024; Accepted 18 February 2024
Available online 24 February 2024

eye movements, motor skills, skin conductance, and respiration. Data-driven techniques can process and analyze biosignals to enhance the objective identification of clinical disorders and their treatment [7]. Indeed, machine learning applied to both implicit and explicit biosignals has successfully identified biomarkers related to various disorders, including neurodevelopmental and psychiatric disorders [8,9].

The promising contribution of machine learning applied to biosignals in clinical assessments may significantly improve the healthcare system by reshaping clinical evaluations [10]. Neurodevelopmental research is increasingly focusing on testing machine learning models applied to implicit and explicit biosignals, providing objective measurements in clinical assessment to complement traditional techniques rather than replace them.

1.2. Machine learning applied to biosignals for the assessment of autism spectrum disorder

Autism spectrum disorder (ASD) is a neurodevelopmental disorder affecting 1 in 100 children worldwide, equating to at least 78 million affected children [11]. Traditional ASD assessment procedures face limitations due to the absence of objective and neurobiological measurements [12–15]. Consequently, one-fourth of children aged eight years or younger exhibiting signs and symptoms of ASD remain undiagnosed [16].

In ASD research, successful machine learning models for assessment purposes exist, differing in the type of biosignal tested (i.e., either implicit or explicit; see [8,17,18] for reviews). However, despite numerous studies have explored different biosignals for ASD classification, research on the superiority of certain biosignals over others in detecting this condition is scarce. Most studies have investigated automatic ASD classification using diverse samples and procedures, making effective biosignal comparisons challenging (see Related work section). Consequently, while biosignal-based ASD classifications exist, determining the superiority of specific biosignals to improve ASD diagnosis remains difficult due to methodological inconsistencies across studies. Specifically, within the context of computational psychiatry and current ASD assessment limitations, the open question of whether implicit biosignals are more robust than explicit biosignals in classifying autism remains unaddressed.

This study aims to fill this research gap by investigating whether machine learning models applied to implicit and explicit biosignals differ in their capacity to identify ASD under the same experimental setup and conditions for biosignal collection. Implicit biosignals, including eye movements and motor skills, and explicit biosignals, such as behavioral and verbal-related performance, were measured in children with ASD and typically developing (TD) children. Machine learning models with reliable validation methods were computed for each implicit and explicit biosignal. The study hypothesizes that machine learning models based on implicit biosignals would achieve better classification performance than those based on explicit biosignals due to the more representative neurobiological information inherent in implicit biosignals.

Furthermore, to ensure study replicability and ecological validity in the assessment procedure, biosignal recording was conducted while participants were immersed in a virtual environment (VE). Evidence suggests that virtual reality (VR) elicits realistic reactions in users like those provoked by real settings [19], overcoming the ecological limitations of clinical settings where assessments are typically conducted (i.e., clinics and laboratories). Besides the enhanced ecological validity, objectivity in VR-based assessments is increased due to the automatic recording of biosignals managed directly by the system. Here, implicit and explicit biosignals were recorded while children experienced a VR playful environment consisting of four scenes. To test the hypothesis, machine learning models for each biosignal in each virtual scene were developed and compared to their combination in a unique model for each biosignal. Finally, the best model in each biosignal between the virtual-scene-specific and the combined one was compared across biosignals to determine the most effective approach.

2. Related works

To our knowledge, no studies have systematically compared implicit and explicit biosignals for the automatic classification of ASD when they are recorded in ecological settings. Indeed, most studies focused on ASD detection using a unique biosignal in the form of unimodal assessment (see [8,20] for reviews). For instance, some studies identified ASD using data from implicit biosignals, such as motor skills (e.g., [21,22]) and eye movements (e.g., [23,24]), while other works used explicit biosignals like behavioral scores derived from patients' observations and interviews (e.g., [25,26]).

Notably, comparing biosignal performance of previous works is challenging due to inconsistencies in the biosignal used, as well as in study procedures, such as (1) varying sample sizes and characteristics, (2) diverse data-processing techniques and machine learning models, and (3) different experimental paradigms and setups. For instance, Carette et al. [23] compared machine and deep learning algorithms for ASD identification using visual scanpaths of children watching videos, while Ardalan et al. [21] employed an easily reproducible classifier based on the full-body motor skills of adolescents, and Wall et al. [26] applied alternating decision trees on behavioral scores coming from reports of the Autism Diagnostic Observation Schedule 2 (ADOS-2). As a result, comparisons across studies seem challenging due to the different methodologies.

Furthermore, there are also studies using biosignal combinations rather than unimodal assessment. One instance is Kang et al. [24], in which a simple classifier was developed collecting EEG data during resting state and eye movements during the visual exploration of face images in young children. The non-simultaneous recording of biosignals may have compromised the ecological validity and prolonged the procedure. On the contrary, Vabalas et al. [22] improved data collection by simultaneously recording motor skills and eye movements during an imitation task involving hand movements. Their machine learning technique was robust and reliable, nevertheless, the presented task lacked ecological validity, and the sample consisted of adults, limiting the ability to make early diagnoses.

From this background arises the academic interest in enhancing the objectivity of ASD assessments using biosignals and machine learning models. However, it seems there is a gap in research as no studies have compared explicit and implicit biosignals to determine which is superior in the classification of ASD.

To define a promising procedure for early ASD diagnosis, recommendations suggest applying machine learning algorithms with reliable validation methods to various biosignals collected from young children during ecologically valid tasks [8,27]. As mentioned earlier, the use of VR systems that present realistic situations to users can provide ecologically valid and objective ASD assessments [28,29]. In this regard, a few studies tested this approach by applying machine learning to unimodal ASD classification based on VR procedures (e.g., [12,13]). In particular, machine learning models applied to full body movements in children with ASD and TD immersed in a VE representing a realistic urban street intersection achieved an 89.36% of accuracy [13]. Likewise, the eye movements of children with ASD and TD recorded during the visual exploration of a virtual scenario resembling a city mall led to the automatic identification of ASD with an 86% of accuracy [30]. Nevertheless, comparing biosignal and model efficiency across these studies is challenging due to the aforementioned differences, which highlights the need for specific and consistent procedures to determine the most promising biosignal in ASD assessment.

3. Procedure

3.1. Participants

81 children took part in the study. Participants' age ranged between 3 and 7 years (see Table 1 for sample characteristics). The sex imbalance between group was in line with the prevalence ratio of autism (4 males, every 1 female diagnosed; [31]). The participants of the ASD group had a diagnosis made previously by expert clinicians through the administration of the ADOS-2. The day of the study, caregivers of the ASD children were asked to bring the assessment report of the ADOS-2. The absence of comorbidities, such as cognitive and language impairments, anxiety, personality, and further neurodevelopmental disorders, was checked in the ASD group by expert clinicians, while, in the TD group, the absence of clinical reports of either diagnosis or risk of the above-mentioned disorders was required to be included in the group. Nevertheless, before children's participation in the study, caregivers of the participants with TD answered a short ad hoc developed questionnaire regarding the presence of any potential symptoms in their children.

Participants' recruitment was made by the Development Neurocognitive Centre Red Cenit (Valencia, Spain), which promoted the study on the social media. The study participation was voluntary and free. The Ethical Committee of the Polytechnic University of Valencia approved the study (ID: P_06_04_06_20).

3.2. The VR system and biosensors

The CAVE Automatic Virtual Environment (CAVE) was chosen as VR system due to its suitability for the ASD population and the reduced risk of discomfort [12,13,32–36]. Head-mounted displays (HMDs) could present difficulties for children with ASD due to their pronounced sensory dysfunction. The extended utilization of HMDs may be disconcerting for young autistic children with severe symptoms, as they might feel suddenly immersed in a virtual environment disconnected from the real world outside [37]. Additionally, the extended use of HMDs in young children (more than 5 min) tends to be unrecommended due to the ongoing development of their visual system [38]. The CAVE was also a good option due to its opportunity to track the whole-body movements of participants while they are free to move in the room with no need to using sensors.

The CAVE was set in the Development Neurocognitive Centre Red Cenit. It consisted of a room of the dimensions of 4 m × 4 m x 3 m in which three ultra-short lens projectors positioned in the ceiling projected 100° images at 55 cm of distance in three walls. Specifically, the main parts of the scenes were presented in the central wall, while the projections on the two lateral surfaces fostered the sense of immersion in the VE. Fig. 1 presents a picture of the CAVE. The sound system used was the Logitech Speaker System Z906 500W 5.1 THX Digital. Besides visual and auditory stimulation, olfactory stimuli were provided by the Olorama Technology™ in two specific moments of the virtual experience (see The VR experience section). The Olorama Technology™ had 12 rechargeable channels which can be selected and triggered by means of a UDP packet, and a programmable fan system that dissipated the odor. In this study, two channels were used for two different odors (wet grass and rose). The intensity of the odor, represented by the time the odor valve was open, was set at the maximum (300 ms).

Table 1
Study sample characteristics.

	ASD (N = 39)	TD (N = 42)
Mean age in months	53.14 (*SD* = 12.38)	57.88 (*SD* = 11.62)
Males	32	19
Females	7	23
ASD diagnosis	ADOS-2	N/A
Comorbidities	No	
Hand dominance	right-handed	
Nationality	Spanish	
Medications	drug naïve	
Vision	normal or corrected to normal	

The interaction in the VE was ensured by the Azure Kinect DK. It was set on a 40 cm high tripod in front of the CAVE central surface, not interfering with participants' vision.

Besides providing interaction in the VE, during the virtual experience, the Azure Kinect DK recorded the whole-body movements of the participant by a computer vision algorithm. It used a depth camera in the resolution mode of 640 x 576 at 30 frames per second. The camera's depth of field allowed participant's body tracking in the entire room. However, to avoid the tracking of further people during the experimental procedure (i.e., the experimenter), a user's interaction area of 3 m × 3 m was set in which body tracking was ensured. Participants were invited to stay in the central portion of this area which size was of 1 m × 1 m (see Fig. 2 for a representation of the experimental setting). The central portion of the user's interaction area was indicated by a grey carpet on the floor and participants were instructed to stay and interact within the area. In the VE, when a virtual action was required from the participant, they could see a transparent virtual human shape mirroring the movements of their head, trunk, and limbs.

Finally, Tobii Pro Glasses 2 were used to measure participants' gaze during the first scene of the virtual experience (see the *VE presentation* scene in the section below). This eye-tracking device recorded what the participant was observing in the dynamic VE from their perspective, while they were free to move in the CAVE. It was equipped with a front camera facing the external environment (i.e., the VE in this case), and two micro-cameras recording the eyes. It was also equipped with an accelerometer and a gyroscope that supported the elimination of the impact of head movements on eye movement data. The eye movements were recorded only in the first scene of the VE rather than during the entire virtual experience due to ergonomic limitations related to the use of eye tracking glasses in children. Specifically, Tobii Pro Glasses 2 are designed for adults and may fit large to children, who wore the eye tracking glasses with an eyeglass lanyard to stabilize them as much as possible. The prolonged use of the eye-tracking glasses may have caused discomfort in children, particularly in those of little age and more severe ASD condition. Consequently, wearing the eye tracking throughout the entire experience would have reduced children's motor freedom during the virtual interaction, affecting the reliability of the implicit biosignal of motor skills. For this reason, eye tracking glasses were used only in the first scene of the VE in which participant's motor movement was reduced and social visual stimulation was increased (see Fig. 3 for the workflow diagram of the procedure).

3.3. The VR experience

The VE was developed in Unity® and it represented a park with an urban playground in which there were two virtual humans: the principal avatar (PA), who was a kid, and the virtual therapist (VT), who was a customized adult appearing when participant's interaction was not as expected. Participants of the two groups experienced four virtual scenes (see Fig. 3a and the supplementary material for detailed information on the virtual humans and the virtual scenes). The usability and user experience of the tasks presented in the virtual scenes was previously tested [32].

In the *VE presentation*, participants attended PA while he presents himself and the park. This scene is characterized by a high-level of social content. During his speech, PA initiates joint attention indicating three targets, and he asks the participant three questions. Participants can answer questions at three distinct levels presented in an orderly manner: by an open-ended answer (first level), by a close-ended answer (second level), and by manually selecting among three options (third level). When the participant answers the question, the further levels are skipped. On the contrary, if the participant does not answer the question

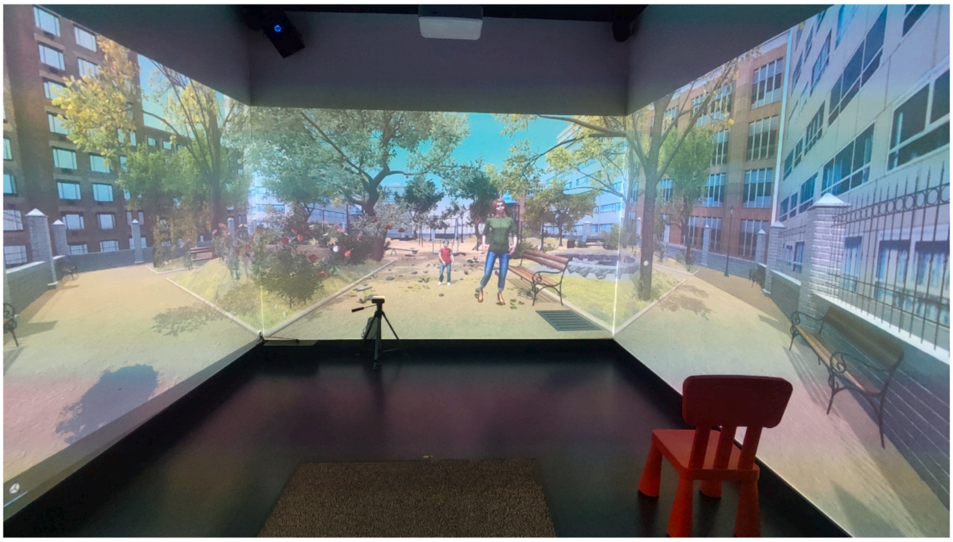

Fig. 1. Experimental setting.

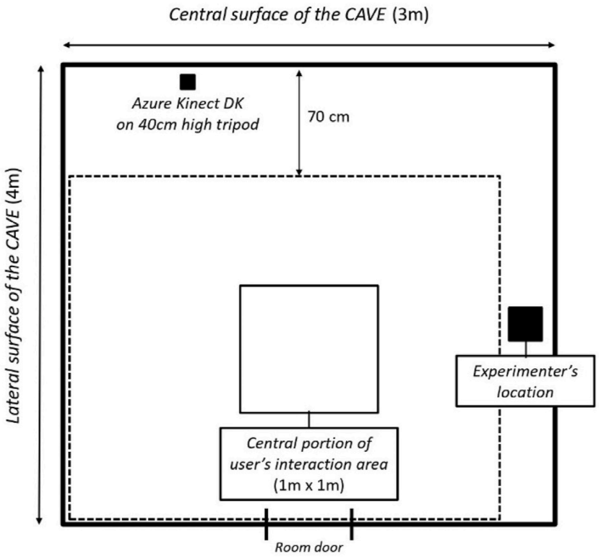

Fig. 2. Experimental setting representation. Please note that figure is not scaled according to real dimensions. The dotted line enclosed the user's interaction area in which participants could interact virtually with the system and in which the body tracking was guaranteed. Within this area, participants were invited to stay in the central portion. The biggest filled-black square represents the experimenter's location during the study.

over the three levels, the lack of the answer is recorded (see Fig. 1 in the supplementary material). During the park presentation by PA, the odor of wet grass is released by the Olorama Technology™. Participant's behavioral responses related to the questions (response time and level of response), eye movements and motor skills were recorded.

In the *kick task (KT),* the participant must kick the ball back to PA for five times. The ball trajectory was predetermined. Participant's behavioral responses (i.e., reaction times, number of kicks, times that VT appeared) and motor skills were recorded.

In the *bubble task (BT),* the participant must blow up thirty soap bubbles made by PA. The participant explodes them by touching. Participant's behavioral responses (i.e., reaction times, number of bubbles exploded, times that VT appeared) and motor skills were recorded.

In the *flower task (FT),* the participant helps PA to pick five flowers. While PA is presenting the task, the odor of rose is released by Olorama Technology™. Then, the participant picks the flowers and take them on the bench. Participant's behavioral responses (i.e., reaction times, number of flowers picked, times that VT appeared) and motor skills were recorded.

3.4. Experimental procedure

The experimental study has been administered between March 2022 and December 2022. Each child underwent to a singular experimental session. Caregivers were informed about the study and gave written consent prior to the children's participation. Sociodemographic data, such as the child's gender, age in months the day of the experimental session, and family's socioeconomic status were recorded before the study.

At the beginning of the study, the participant chose between male and female virtual human shape to facilitate the meta-self-recognition with the avatar projected by the Azure Kinect DK. This choice was made in an anonymous virtual setting differing completely from the one of the virtual experience.

In the same virtual setting, participant had the chance to familiarize themselves with the projection of the virtual human shape and the virtual setting (see experimental procedure section in the supplementary material for more information). When the participant was ready to start the experiment, the experimenter put on the child the eye-tracking glasses and the virtual experience started.

During the virtual experience, the *VE presentation* was always seen as first, while the further tasks were presented in a randomized and counterbalanced order between participants. After the exposition to the *VE presentation*, the eye-tracking glasses were removed. Due to the participants' little age and neurodevelopmental condition, breaks from task performance were taken in case it was needed. Data on eye movements were recorded only in the *VE presentation*, while data regarding the behavioral performance and motor skills were recorded in the four virtual scenes.

3.5. Biosignal data processing

3.5.1. Implicit biosignal processing

3.5.1.1. Motor skill tracking and processing. The device used to track motor skills can obtain the positions of 27 joints of the body in real time (see Fig. 3b). This includes all people present in the user's interaction area, which in the raw data includes not only the participant who was

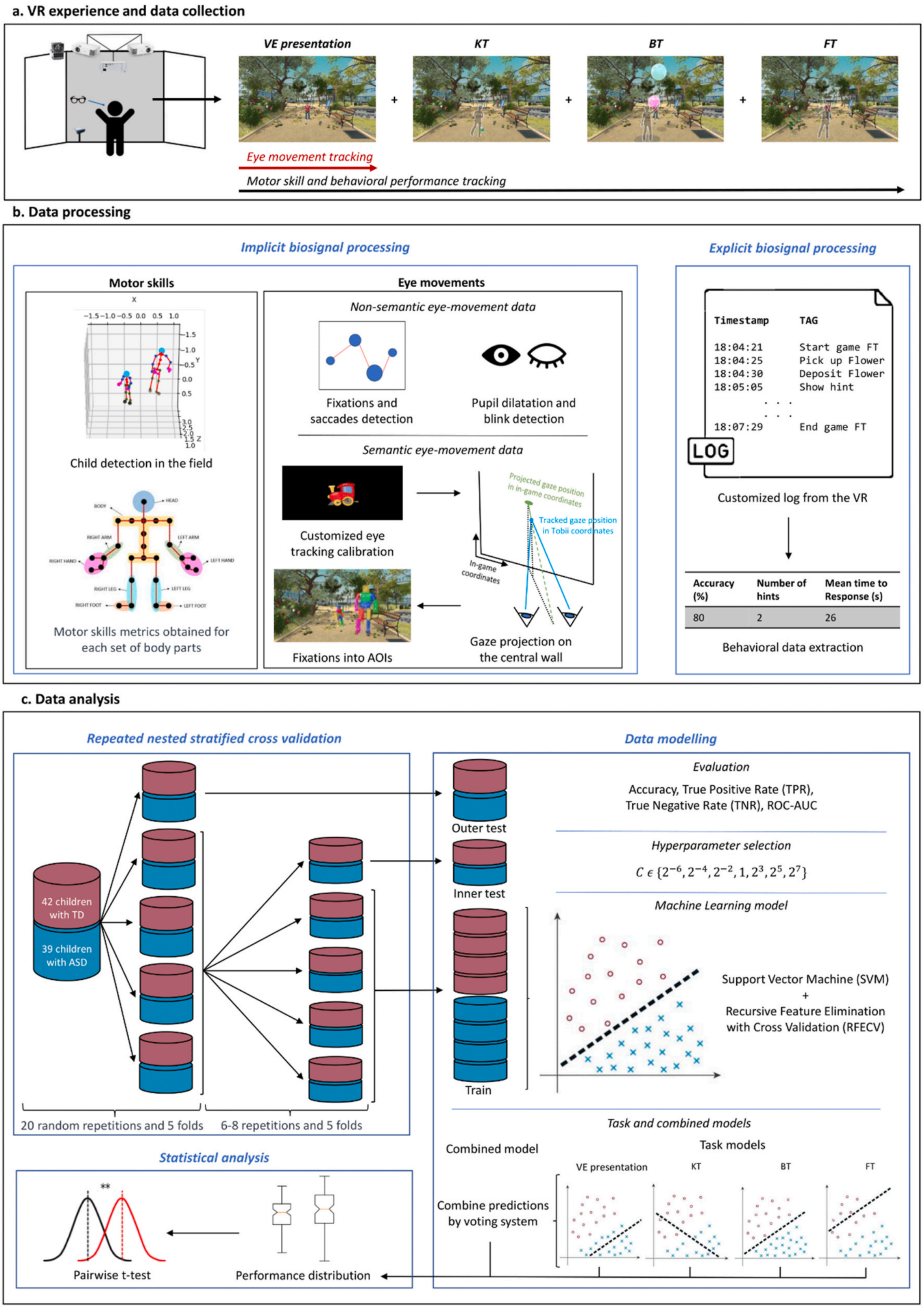
a. VR experience and data collection
VE presentation
KT
BT
FT
Eye movement tracking
Motor skill and behavioral performance tracking
b. Data processing
Implicit biosignal processing
Explicit biosignal processing
Motor skills
Child detection in the field
Motor skills metrics obtained for each set of body parts
HEAD
BODY
RIGHT ARM
LEFT ARM
RIGHT HAND
LEFT HAND
RIGHT LEG
LEFT LEG
RIGHT FOOT
LEFT FOOT
Eye movements
Non-semantic eye-movement data
Fixations and saccades detection
Pupil dilatation and blink detection
Semantic eye-movement data
Customized eye tracking calibration
Projected gaze position in in-game coordinates
Tracked gaze position in Tobii coordinates
In-game coordinates
Gaze projection on the central wall
Fixations into AOIs
Timestamp TAG
18:04:21 Start game FT
18:04:25 Pick up Flower
18:04:30 Deposit Flower
18:05:05 Show hint
18:07:29 End game FT
LOG
Customized log from the VR
Accuracy (%) Number of hints Mean time to Response (s)
80 2 26
Behavioral data extraction
c. Data analysis
Repeated nested stratified cross validation
42 children with TD
39 children with ASD
20 random repetitions and 5 folds
6-8 repetitions and 5 folds
Data modelling
Outer test
Inner test
Train
Evaluation
Accuracy, True Positive Rate (TPR), True Negative Rate (TNR), ROC-AUC
Hyperparameter selection
$C \in \{2^{-6}, 2^{-4}, 2^{-2}, 1, 2^{3}, 2^{5}, 2^{7}\}$
Machine Learning model
Support Vector Machine (SVM) + Recursive Feature Elimination with Cross Validation (RFECV)
Task and combined models
Combined model
Task models
VE presentation
KT
BT
FT
Combine predictions by voting system
Statistical analysis
**
Pairwise t-test
Performance distribution

Fig. 3. Workflow diagram of the procedure.

carrying out the experience, but in some cases also the experimenter when their intervention in the interaction area was needed due to the young age of children. To distinguish between the movements of children and the experimenter, an automatic participant detection method based on continuity of movement and participant's height was implemented following two assumptions: (1) the centre of participant's body followed a smooth movement, that is, the displacement produced between frames was less than 20 cm during the entire virtual experience; (2) considering the height of children, it was checked that their maximum height was less than that of the experimenter. Once the participant was correctly identified throughout the recording, the noise from the low precision of the device was removed using a smoothing technique. A moving average with a uniform window size of 5 was used for this purpose. This process was performed five times to obtain smooth and derivable position and velocity curves. Then, motor skills metrics were obtained for each set of body parts shown in Fig. 3b. These included the mean, maximum, and minimum value of the displacement, velocity, acceleration and tangential acceleration between consecutive frames, as well as position. Additionally, the total number of missing values was also recorded. The considered metrics are summarized in Table 1 in the supplementary material.

3.5.1.2. Eye-movement data processing. We considered two key aspects of eye movements: the semantic and non-semantic eye movements. Non-semantic eye movements refer to the gaze features not directly related to the surrounding context, such as fixations, saccades, blinks, and pupil diameter, while semantic eye movement features are closely related to the elements of the scene and how an individual visually attends them, such as the frequency of gaze towards the virtual humans or the objects of interest. Given their processing and connotation differences, we considered these two aspects as different datasets, and extracted different metrics for each type of data. For a detailed list of metrics, see Table 1 in the supplementary material.

Non-semantic eye-movement data. Data consisted of two main files provided by the Tobii Pro Glasses 2 for each subject: the first file contained information about the tracked position of the subject's gaze in 3D space, and the second file contained the registered pupil diameter (see Fig. 3b for data processing). Both signals were provided as time series with a given timestamp for each value. First, we considered the file containing gaze to study fixations and saccades.

To detect fixations with the given gaze, a velocity threshold algorithm on the raw eye movement data was used following the procedure described in Salvucci and Goldberg [39]. This method considers consecutive eye movements with a velocity of less than 30/s as belonging to the same fixation. A minimum fixation duration threshold of 60 ms was established. To account for missing values in the raw eye movement data, a maximum threshold of 60 ms of missing values within a fixation was defined. If the duration exceeds this threshold, the fixation is automatically terminated. From these, we were able to detect fixations for each participant across the virtual scene. Then, the total number of fixations and the mean and variance of the duration, frequency, and the area covered by fixations were calculated.

Saccades are typically interpreted as eye movements that occur between fixations. However, whether there are many missing values, fixations or eye movements that are not long enough to form a fixation may be mislabelled. To address this issue, another velocity threshold algorithm was used to detect saccades. This method considers consecutive eye movements with a velocity greater than 30/s as belonging to the same saccade. Like the fixation detection algorithm, a maximum missing value threshold was used to separate saccades, although in this case the threshold was reduced to 40 ms due to the shorter duration of saccades. Then, the total number of saccades, the ratio of saccades to fixations, the mean and variance of the duration, and the frequency and amplitude of the saccades were computed.

In addition to eye movements, we studied the recorded pupil diameter. First, the mean and variance of pupil diameter and its velocity of dilation were calculated. Additionally, we studied the presence of missing values which can indicate a blink. Following the algorithm proposed by Hershman et al. [40], the mean and variance of blink duration and blinks per minute were computed, where blinks were considered as a succession of missing values shorter than 100 ms.

Semantic eye-movement data. A relevant aspect of eye movement analysis is determining whether the participant is looking at an Area of Interest (AOI). In the current experimental setup, participant's gaze was determined relative to the video captured by the camera embedded in the eye-tracking glasses and using the participant's point of view. On the contrary, the AOI data was referenced relative to the projection in the central surface of the CAVE. Consequently, a discrepancy emerged between the reference coordinates of the fixations and the AOIs. Therefore, to determine where the participant was looking at any given moment within the context of the experience, it was necessary to project their gaze onto the central wall projection and then check if they were looking at an AOI (see Fig. 3b).

To accomplish this, the location of the central surface of the virtual projection was detected in each frame of the video recorded by the eye-tracking glasses. It should be noted that to obtain a projection of the participant's gaze onto the CAVE central projection surface, a faithful and precise recording of the gaze must be available. For this reason, eye tracking calibration are usually employed before gaze recording. However, conventional automatic calibration procedures involve staring at a fixed point for a certain time interval, and challenges with young children not consistently focusing on the calibration point were encountered. Thus, to accurately track participants' gaze and project it onto the CAVE central surface projection, a customized eye-tracking calibration method was employed. Prior to initiating the virtual experience, a 10 s image of a pulsating train was displayed on the central wall of the CAVE (see Fig. 2 in the supplementary material). Children were instructed to focus their gaze on the train to start the virtual experience. By capturing in the eye-tracking video the position of the train in the central wall and correlating it with participants' fixations at that moment, the eye movements were manually calibrated throughout the entire virtual scene for each participant, and they were projected to the CAVE central projection surface. Then, fixations with a deviation of more than 300 pixels were removed to eliminate those that were too scattered due to head movements. Finally, it was checked if projected fixations belonged to an AOI or not. This was achieved by applying a minimum threshold of 60 ms, requiring participants to gaze at an AOI for at least this duration on each projected fixation. Fixations were attributed to the corresponding AOIs if one or multiple AOIs were fixated upon for more than 60 ms within a fixation. The AOIs studied were the parts of the virtual humans' bodies – head, trunk, legs and arms – and objects pointed out by PA during the virtual experience when he is initiating joint attention. Finally, the mean and variance of the duration of fixations on each AOI were obtained.

3.5.2. Explicit biosignal processing

3.5.2.1. Behavioral data processing. To process behavioral data, a custom log was created in Unity® including a timestamp for every item appearing in the virtual scenes, and every participant interaction with the virtual avatars and the virtual elements. The custom log captured the initiation timestamps for each interaction, as well as the timestamps for the appearance of PA and VT, and the instantiation of each item in the virtual tasks, such as the balls in the KT, the bubbles in the BT, and the flowers in the FT.

Using this custom log, behavioral data were extracted in each virtual scene, including the mean response times, the number of hints needed from the VT, and the accuracy in each task. The task accuracy was in the *VE presentation*, the answer level in which participant gave an answer to PA's questions, in the KT the number of balls to be kicked, in the BT the number of bubbles to be blown up, and finally in the FT the number of flowers participants had to take to the bench (see Fig. 3b for visual representation of data processing, and Table 1 in the supplementary material for a detailed list of metrics).

3.6. Data analysis

First, group differences in control variables such as age were tested using the independent sample *t*-test.

Second, to classify children with ASD, a statistical multivariate machine learning model was developed for each data type to explore the importance of each biosignal. Specifically, thirteen models were developed: eight machine learning models on motor skills and behavioral performance in the four virtual scenes respectively; two machine learning models on semantic and non-semantic eye movements in the *VE presentation*; and three machine learning models based on the combination of models related to the same biosignal. These combined models were computed using a voting system combining all models for each biosignal.

Given the small dataset and the large number of features, a simple linear support vector classifier (LinearSVC) integrated in a recursive feature elimination was used to avoid overfitting. Specifically, LinearSVC model was selected over other models after initial explorative validation, where nonlinear models such as SVC with gaussian kernels yielded worse performance, potentially due to the large number of features and requiring extensive finetuning. Our validation strategy consisted in a nested cross-validation (see Fig. 3c), including an outer cross-validation procedure was applied to generate several test partitions differing from those used to develop the model, which was combined with an inner cross-validation that chooses the optimal hyperparameters of the LinearSVC regularization (C), ranging logarithmically between 2^{-6} and 2^{7}, while also choosing the optimal number of features selected by the recursive feature elimination model. On the one hand, the inner cross-validation consisted of a repeated stratified k-fold which used 5 folds of the train data of each outer loop. This inner cross-validation was repeated 6–8 times (for motor skills data and eye movements data, respectively) to obtain stable and robust results. On the other hand, the outer cross-validation consisted of another repeated stratified k-fold with 20 repetitions and 5 folds, which resulted in 100 test sets per model. This stratified inner and outer strategy was chosen to train and test models with a more representative subsample of the real data. Additionally, due to the imbalanced dataset, a class-balanced regularization approach was used for the LinearSVC classifiers.

To evaluate the models a set of metrics were considered: accuracy (i.e., percentage of subjects correctly recognized), true positive rate (i.e., percentage of ASD subjects correctly labelled), true negative rate (i.e., percentage of control subjects recognized as control), and Receiver Operating Characteristic – Area Under the Curve (AUC), which describes the ability of the model to distinguish between positive and negative classes (0.5 being indistinguishable from a random class assignation and 1 being a perfect discrimination). The models were optimized to achieve the best AUC.

Finally, to assess the top-performing models within each biosignal category, as well as the comparative performance of the machine learning models in relation to the specific biosignal type, a post-hoc pairwise *t*-test was conducted. This statistical analysis involved evaluating the AUC results obtained from the 100 test sets, which represented the performance distribution for each task. By applying *t*-test, the significance of performance differences among the models was determined.

4. Results

First, participants' sociodemographic data did not differ between groups (*ps* > 0.05).

Regarding the thirteen machine learning models developed, findings are shown in Table 2, in which their performance is reported depending on the biosignal. Table 2 includes the accuracy, true positive rate, true negative rate and AUC means of the machine learning models trained for each biosignal in the correspondent virtual scene, as well as in the performance of the combined model of each biosignal in the test outer loop of the nested cross-validation. The best model was based on motor skills and combined all the virtual scenes, with an 81% of accuracy mean (*SD* = 9%) and 0.89 (*SD* = 0.08) of AUC mean. The model also recognized the TD with an 89% of accuracy mean (*SD* = 11%). Motor skills models on the KT, BT and *VE presentation* also achieved promising accuracy means varying from 75% (*SD* = 10%) to 80% (*SD* = 11%). In addition, the combined model on behavioral performance achieved a 75% of accuracy mean (*SD* = 10%) and 0.80 (*SD* = 0.11) of AUC mean, and the rest of the models presented accuracy means lower than 70%.

Table 2
LinearSVM + RFECV test nested cross-validation results by biosignal and virtual scenes. Results include mean and standard deviation.

Biosignal	Virtual scene	Mean accuracy (*SD*) %	Mean TPR (*SD*) %	Mean TNR (*SD*) %	Mean AUC (*SD*)
Motor skills	Combined	81 (9)	72 (17)	89 (11)	0.89 (0.08)
	VE presentation	75 (10)	63 (19)	86 (11)	0.81 (0.10)
	KT	80 (10)	75 (18)	84 (11)	0.86 (0.10)
	BT	80 (11)	80 (17)	80 (13)	0.84 (0.13)
	FT	65 (13)	53 (25)	72 (19)	0.68 (0.16)
Eye movements	Combined (NS + S data)	59 (12)	29 (19)	82 (16)	0.58 (0.17)
	VE presentation (NS data)	69 (13)	57 (25)	75 (16)	0.75 (0.15)
	VE presentation (S data)	54 (15)	59 (32)	51 (21)	0.54 (0.19)
Behavioralperformance	Combined	75 (10)	57 (18)	91 (13)	0.80 (0.11)
	VE presentation	73 (9)	55 (17)	88 (14)	0.72 (0.09)
	KT	71 (11)	51 (20)	87 (12)	0.70 (0.12)
	BT	71 (11)	68 (21)	73 (17)	0.71 (0.11)
	FT	68 (10)	37 (23)	86 (13)	0.62 (0.11)

NS = non-Semantic eye-movement data; S = Semantic eye-movement data.

Regarding model variability over test splits, the combined model on motor skills showed reduced AUC variance (0.08) than its virtual-scene specific counterparts, while also being the least variable across all models, making it the most robust model. Eye tracking, on the contrary, reports the greatest variability over splits, showing a variance of 0.19 and 0.17 for semantic and combined models, respectively.

Tables 3–5 show the model comparisons depending on the biosignal. Findings indicate that combined models were significantly better than virtual-scene specific models on motor skills and behavioral performance, while for the implicit biosignal of eye movements, the non-semantic eye movement model significantly outperformed the semantic eye movement model and the combined counterpart.

Finally, in Fig. 4 are shown the comparisons between the best machine learning models of each biosignal. It can be noted that the combined model based on motor skills with AUC mean of 0.89 (*SD* = 0.08), that yielded the best classification results, is also significantly different from the others, while non-semantic eye tracking has a significantly lower performance in comparison to the other models. To further illustrate model differences, Fig. 5 shows the mean AUC curves of the

Table 3
Statistical differences for the machine learning models based on motor skills in AUC distributions over 100 folds. **p < 0.05, **p < 0.01, ***p < 0.001.*

Motor skills				
VE presentation	BT	KT	FT	Combined
1.000	0.057	>0.001 (***)	>0.001 (***)	>0.001 (***)
	1.000	0.235	>0.001 (***)	0.001 (**)
		1.000	>0.001 (***)	0.011 (*)
			1.000	>0.001 (***)
				1.000

Table 4
Statistical differences for the eye movement models in AUC distributions over 100 folds. **p < 0.05, **p < 0.01, ***p < 0.001.*

Eye movements		
Non semantic	Semantic	Combined
1.000	>0.001 (***)	>0.001 (***)
	1.000	0.168
		1.000

Table 5
Statistical differences for the machine learning models based on behavioral performance in AUC distributions over 100 folds. **p < 0.05, **p < 0.01, ***p < 0.001.*

Behavioral performance				
VE presentation	BT	KT	FT	Combined
1.000	0.580	0.106	>0.001 (***)	>0.001 (***)
	1.000	0.318	>0.001 (***)	>0.001 (***)
		1.000	>0.001 (***)	>0.001 (***)
			1.000	>0.001 (***)
				1.000

combined models. It is observed that mean AUC curves are separated approximately by one standard deviation from each other.

5. Discussion

The primary objective of the present study is to compare implicit and explicit biosignals in the VR-based automatic and early identification of ASD. To this end, our approach involved systematically evaluating multiple biosignals within the same framework and dataset.

This rigorous methodology allowed for a robust statistical comparison of the efficacy of each biosignal in ASD detection and classification, overcoming the challenges posed by prior studies that utilized disparate methodologies, hindering direct comparison. Specifically, a total of eighty-one children between the ages of 3 and 7 were divided into two groups: a group consisting of children previously diagnosed with ASD, and a group of children with TD. The participants were immersed in a realistic VE where they interacted and played with a virtual child in four virtual scenes (*VE presentation*, KT, BT, and FT). The assessment was conducted in VR to ensure ecological validity and objectivity in the measurement. During the experience, both implicit biosignals (motor skills and eye movements) and explicit biosignals (behavioral and verbal-related performance) were recorded.

For each biosignal, virtual scene-specific machine learning models were developed to classify ASD. Additionally, a machine learning model was constructed for each biosignal by combining the virtual scene-specific models using a voting system. The performance of the models (both virtual scene-specific and combined) was compared to determine the top-performing model for each biosignal. Finally, the top-performing models of biosignals were compared to identify the most effective biosignal (and corresponding portion of the virtual experience) for the early and objective identification of ASD. The study hypothesized that implicit biosignals would demonstrate higher performance in ASD classification compared to explicit biosignals.

5.1. Machine learning model evaluation and stability

The machine learning model chosen for the automatic ASD identification was LinearSVC, primarily due to previous evidence supporting its efficiency with small sample sizes and reduced computational cost, particularly when dealing with implicit biosignals [8]. Additionally, a nested cross-validation procedure was implemented, which involved an inner loop for feature selection and hyperparameter tuning, and an outer loop to designate test partitions. Importantly, an outer k-fold cross-validation approach with 20 repetitions and 5 folds was utilized, resulting in a total of 100 distinct test partitions that were not seen during the training or validation of the machine learning model. This methodology provided a substantial collection of unseen test partitions, facilitating the evaluation of the model's variability and stability across various partitions. Such evaluation methodology resulted in greater model stability, particularly when compared with traditional methods used in ASD assessments based on biosignals. Conventional methods

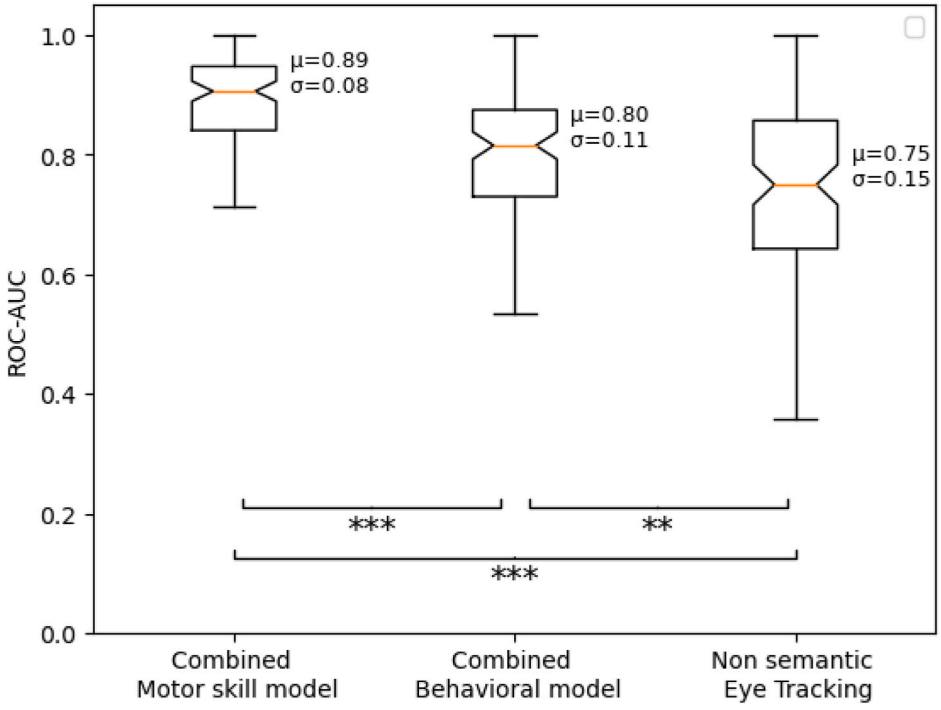

Fig. 4. Statistical differences in AUC distributions of the best model among data types. **p < 0.05, **p < 0.01, ***p < 0.001.*

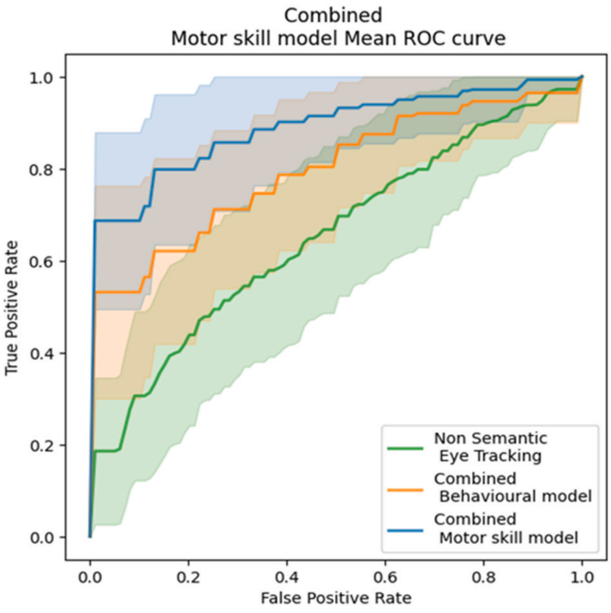

Fig. 5. AUC curve distributions of the best model among data types. Highlighted areas represent the standard deviation of the AUC for each curve value, while the continuous line represents the mean AUC curve across outer folds.

typically rely on basic cross-validations (e.g., [21]) that do not involve an unseen test set or a hold-out procedure producing only a single test partition.

The thirteen LinearSVC models achieved varying performances in ASD classification, with accuracy means ranging from 54% ($SD = 15\%$) to 89% ($SD = 9\%$), and AUC means ranging from 0.54 ($SD = 0.19$) to 0.89 ($SD = 0.08$). Notably, nine out of the thirteen models achieved acceptable to excellent discrimination (AUC means ≥ 0.70; [41]). These included four models based on the implicit biosignal of motor skills (three virtual scene-specific models and the combined model), one model based on the explicit biosignal of behavioral performance (combined model), the model based on non-semantic eye movement data, and three virtual scene-specific models based on behavioral data. These nine models also demonstrated significant true negative rate means, indicating their ability to correctly identify children in the TD group.

Conversely, the four models that demonstrated poor discrimination performance (AUC means < 0.70; [41]) were based on motor skills and behavioral data in the FT, the combined model on eye movements, and the model relying on semantic eye movements.

In the subsequent sections, the machine learning models gathered for each biosignal are discussed.

5.2. Machine learning models based on the implicit biosignal of motor skills

Among the machine learning models on motor skills, the model utilizing motor data in the FT exhibited the lowest performance with an AUC mean of 0.62 ($SD = 0.11$). However, there is evidence indicating that virtual tasks like the FT can provide insights into motor differences between children with ASD and TD at a descriptive level, which suggests the potential for promising classification accuracies based on motor skills in this task. Specifically, children with ASD within the same age range demonstrate whole-body motor abnormalities during task execution compared to their TD peers [33].

Nevertheless, in the present study, the FT did not exhibit the expected sensitivity in objectively classifying children with ASD using motor skills. Likely, the motor performance of children in our study differed from Minissi et al. [33] due to modifications made to the FT, such as including the VT and aesthetic elements (e.g., the background and auditory feedback). It should be noted that the FT used in previous studies was more basic, requiring participants to perform the target action in a minimal virtual environment. Inserting decorative elements and virtual humans yielded to different performance than expected.

On the contrary, the other three virtual-scene specific models based on motor skills demonstrated promising AUC means ranging from 0.81 ($SD = 0.10$) to 0.86 ($SD = 0.10$). This can be attributed to the specific motor requirements of the virtual scenes that asked to move the whole body in different manners.

In addition, the combined model utilizing motor skills achieved the highest performance in ASD classification compared to the other models based on this implicit biosignal. It attained an AUC mean of 0.89 ($SD = 0.08$), indicating excellent discrimination ability of the model [41]. Additionally, it yielded a general accuracy mean of 81%, a true positive rate mean of 72% ($SD = 17\%$) and a true negative rate mean of 89% ($SD = 11\%$). The inclusion of motor data from the entire virtual experience through the combination of virtual-scene-specific models significantly enhanced the classification performance. Consequently, the FT may be considered equally relevant to the other tasks for the automatic identification of ASD, owing to its contribution in boosting the performance of the combined model.

5.3. Machine learning models based on the implicit biosignal of eye movements

Regarding the other implicit biosignal, the machine learning model applied to non-semantic eye movements demonstrated good discrimination performance due to the valuable information carried by non-semantic data. It was the highest-performing model based on eye movements, achieving an AUC mean of 0.75 ($SD = 0.15$), a mean true positive rate of 57% ($SD = 25\%$), and a mean true negative rate of 75% ($SD = 16\%$). This model was among the six models that achieved good discrimination performances while the other two models applied to eye movements were the worst performing, achieving AUC means representing performances close to random class assignation. In particular, the combined model and its semantic counterpart achieved AUC means of 0.54 ($SD = 0.19$) and 0.58 ($SD = 0.17$) respectively.

Despite impaired social visual attention being a common symptom of ASD [42], semantic eye movements, which are directly linked to social visual attention, did not prove to be a sensitive implicit biosignal for the automated classification of ASD. Consequently, the combined model utilizing eye movement data ranked second to last in classification performance due to the influence of the model based on semantic eye movements, which significantly reduced the accuracy achieved by the non-semantic eye movement model. The reduced classification performance of these models can likely be attributed to the labour-intensive post-processing technique employed for semantic information. Specifically, semantic eye movements were extrapolated involving the projection of corrected fixations onto the central surface of the CAVE, where the AOIs were presented. This two-step process may have impacted the quality of the semantic data, thereby reducing the effectiveness of the classification model. Furthermore, the automatic recursive feature elimination process may have also played a role in further reducing the classification performance.

To the best of our knowledge, Alcañiz et al. [30] is the only study that presents a machine learning model for ASD classification based on eye movement data recorded in VR. A similar experimental paradigm was employed in acquiring eye movements, and various machine learning models achieved an accuracy of 86% in discriminating ASD. However, although the VR system and eye-tracking glasses were the same, the virtual experience used in Alcañiz et al. [30], where eye movements were recorded, was longer compared to the current study (24 min in Alcañiz et al. [30] versus 4 min of eye movement recording in our study). This may have introduced greater variability in the data, thereby enhancing the opportunity to develop efficient classification models. Additionally, the AOIs for social stimuli were broader in Alcañiz et al.

[30], encompassing larger virtual areas rather than specific small areas as in our study (e.g., the body parts of AP and TV). This might have increased the detection of semantic information in eye movements. Furthermore, there were differences in the post-processing techniques between the studies. In Alcañiz et al. [30], semantic eye movements were not separated from non-semantic eye movements, and their combination was utilized to develop a classification model without an external replication dataset. In our study, in turn, the two types of eye movements were distinguished, and the virtual content aesthetics differed. Keeping semantic and non-semantic eye movements together during the development of machine learning models may lead to more robust discriminations when working on this type of implicit biosignal.

5.4. Machine learning models based on the explicit biosignal of behavioral performance

Concerning the machine learning models based on the explicit biosignal of behavioral performance, the model utilizing the FT exhibited the lowest AUC mean (0.62 (*SD* = 0.11)). Similar to the motor performance observed in the FT, prior evidence indicates that behavioral performance in this task is impaired in children with ASD [32], who require more time to complete the task compared to their typically developing counterparts. Therefore, the reason behind the reduced AUC mean may be attributed to the same factors that contributed to the poor performance in the model based on motor skills, such as the aesthetic modifications made to the task that were aimed at enhancing the sense of realism and immersion.

On the other hand, the other three virtual scene-specific models achieved AUC means greater than 0.70 (*SD* = 0.12). This indicates that the selected features in these virtual scenes were more effective in identifying ASD compared to the FT. Moreover, the combination of the four virtual scenes yielded the highest discrimination ability among the behavioral models. It achieved an AUC mean of 0.80 (*SD* = 0.11), a mean true positive rate of 57% (*SD* = 18%), and a mean true negative rate of 91% (*SD* = 13%). Similar to the classification models based on motor skills, the diversity in the behavioral demands of the virtual scenes enhanced the classification performance of the combined model incorporating this explicit biosignal.

Overall, the combined models on both the implicit and explicit biosignals recorded throughout the experience (motor skills and behavioral performance) surpassed the classification performances of their virtual-scene specific counterparts. However, it is worth noting that the implicit biosignal of eye movements exhibited the lowest classification performances, except for the model based on non-semantic eye movements.

5.5. Comparison of the top-performing models of biosignals

The statistical comparison among the top-performing models within the biosignals led to the identification of the combined model based on motor skills as the most accurate in ASD classification. It outperformed both the combined model based on behavioral performance and the model based on non-semantic eye movements. The combination of models based on motor skills in the four virtual scenes resulted in the development of a robust algorithm capable of identifying ASD with excellent accuracy. The strong performance of this model can be attributed to the stability in motor recording, as well as the diversity of motor information collected throughout the virtual experience. Each of the four virtual scenes had distinct motor requirements, varied body parts involved in movement, and different degrees of freedom in target movement [33]. This finding confirms the study hypothesis that neurobiological information underpinned by the implicit biosignal of motor skills is more sensitive in identifying ASD compared to the explicit biosignal of behavioral performance. Notably, while impairments in motor skills are not currently included in the ASD diagnostic criteria, their inclusion is highly recommended [43]. On the other hand, behavioral performance related to ASD symptoms is well-known and forms the foundation of traditional ASD assessment procedures.

The ADOS-2 is considered the gold standard among various tests for ASD assessment. The ADOS-2 has four modules used for assessment based on the child's characteristics. These modules exhibit a true positive rate ranging from 83% to 98% and a true negative rate ranging from 50% to 94% in identifying children with ASD [14]. We can posit that the combined model based on motor skills recorded in the current virtual experience performed similarly to traditional assessment procedures in terms of true positive and negative rates. This finding encourages further research on ASD assessment based on motor skills. Furthermore, the absence of subjective observations made by expert clinicians in the virtual experience overcomes the limitation of ASD assessment regarding the lack of objective measurements. Indeed, the proposed ASD classification is based on the implicit biosignal of motor skills, which is objectively recorded by the VR system and automatically analysed by specific software.

To our knowledge, there are previous instances of ASD identification based on motor skills recorded in the whole-body. However, among these studies, only two measured motor skills in standardized and ecological settings (i.e., [13,21]). On one hand, Ardalan et al. [21] achieved 89% accuracy in ASD classification, but the sample size was smaller, and the participants were older compared to the current study (ages 7–17 years). Moreover, whole-body motor skills were measured during the experience of a video game aimed at training balance in children with ASD, requiring multiple playing sessions rather than a specific session as in this study. Reducing the time and cost of assessment is an important consideration that could be addressed by implementing a single diagnostic session guided by the automatic recording of objective data [15]. On the other hand, in Alcañiz Raya et al. [13], the experimental session was unique, and children experienced an immersive and realistic virtual environment characterized by high ecological validity and objectivity in biosignal recording. An 82.98% accuracy in ASD classification was achieved, and despite whole-body movements being recorded, the feature extraction process led to the development of a machine learning model based on the movement of specific body parts (head, trunk, and feet). In summary, the present findings on the combined model of motor skills go beyond previous studies by enhancing the level of realism in the virtual experience and sample size, reducing participants' age, and improving the design of the machine learning model.

Finally, regarding the comparison between eye movement and behavioral models, the combined behavioral model significantly outperformed the model based on non-semantic eye movements. We can conclude that study hypothesis of better classification performance using implicit rather than explicit biosignals was confirmed in one case out of two. However, non-semantic eye movements provided promising results, and their model outperformed the virtual-scene specific behavioral models, suggesting that the neurobiological information carried by non-semantic eye movements is more precise for identifying ASD than virtual scene-specific behavioral performance.

As mentioned, in both motor skills and behavioral performance, the model combination outperformed the virtual-scene specific models. Therefore, if semantic eye movements exhibited a classification performance similar to non-semantic eye movements, the combined model based on their integration could have surpassed the combined behavioral model due to the better AUC means of the semantic and non-semantic models. We believe that the machine learning models based on eye movements did not achieve the expected performance due to limitations regarding the short recording duration and the processing of semantic information. If eye movements had been recorded using a different device more suitable for children, with an extended recording duration during the experience, they might have outperformed the behavioral models.

To our knowledge, at the time of implementing the study, there were no eye tracking glasses suitable for children available on the market. However, this changed in 2023 when the company Pupil Labs released the Neon Product, which is suitable for children between 2 and 8 years old. Future studies should utilize this type of product to assess eye movements in children immersed in realistic VR experiences, as it overcomes the ergonomic limitations of the Tobi Pro Glasses 2 and employs a more efficient post-processing technique for semantic information.

Finally, overcoming the current limitations of clinical assessment is an ongoing debate, and the combination of VR, biosignals, and machine learning may fulfill the need for ecological settings and objective measurements in assessment. The current study proposed a comparison of biosignals based on machine learning models for the identification of ASD in realistic virtual settings. It led to the determination of motor skills as a robust implicit biosignal that may facilitate the objective and ecological assessment of ASD. In particular, its combined model identified ASD in test set with an AUC of 0.89, while eye movements and behavioral responses best models achieved an AUC of 0.75 and 0.80 respectively.

5.6. *Limitations and future directions*

The present study encompasses certain limitations that warrant acknowledgment. Apart from the ergonomic and data processing limitations associated with the eye tracking glasses, there are limitations concerning the sample composition. Specifically, the two groups were not matched in terms of IQ, and only participants who accomplished the familiarization with the system were included. There was also a different sex ratio between groups which may have affected data results, even though was in accordance with the sex ratio of ASD diagnosis. In addition, the assessment of symptom severity was not conducted in children with ASD. This stems from the study's aim to assess ASD irrespective of symptom severity rather than focusing on specific ASD subgroups, which is line with conventional assessment procedures. Nonetheless, it is advisable for future studies to include control for IQ and symptom severity, thereby developing machine learning models that are sensitive to these factors and potentially capable of stratifying the disorder. Furthermore, it is recommended for subsequent studies to employ eye tracking tools suitable for young children (as described earlier) and explore the potential of the implicit biosignal of motor skills recorded in VR in detecting other neurodevelopmental disorders, such as attention deficit hyperactivity disorder. Finally, it may be interesting to compare more implicit biosignals for the automatic and ecological detection of ASD, such as heart rate variability, respiration, and body temperature, in order to determine whether motor skills remain the superior biosignal. Additionally, it could be of scientific interest to combine biosignals and compare the resulting combinations in the framework of multimodal ecological assessments (studies are underway).

CRediT authorship contribution statement

Maria Eleonora Minissi: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Alberto Altozano:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Formal analysis, Data curation. **Javier Marín-Morales:** Writing – review & editing, Validation, Supervision, Software, Formal analysis. **Irene Alice Chicchi Giglioli:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Fabrizia Mantovani:** Writing – review & editing, Supervision. **Mariano Alcañiz:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT by OpenAI in order to improve readability and language of the paper. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Spanish Ministry of Economy, Industry, and Competitiveness-funded project IDI-20201146. It was also co-founded by the European Union through the Operational Program of the European Regional development Fund (FEDER) of the Valencian Community 2014–2020 (IDIFEDER/2018/029 and IDIFEDER/2021/038). Furthermore, MEM received partial funding by the Valencian Community (GRISOLIAP/2019/137; CIBEFP/2021/38), and by the Polytechnic University of Valencia (Predoctoral Mobility Grant of 2021). JMM received partial funding from the Polytechnic University of Valencia (PAID-10-20). Funding for open access charge: CRUE-Universitat Politècnica de València.

The authors acknowledge the valuable contribution of Manuel Martínez González, 3D artist, and Jose Roda Belles, virtual reality and information technology programmer, for the development of the virtual experience. Authors are also thankful to all the families who participated in the study for their interest and dedication to research in the autism field. Authors also acknowledge the contribution of Neus Centelles, who coordinated the administration of the study, as well as the aid of Luis Abad and Marian Sirera of the Neurocognitive Centre Red Cenit. Finally, authors are thankful to Martina Mari and Giampaolo Abate Romero Landini for their dedication in data post processing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.108194.

References

[1] B.A. Nosek, C.B. Hawkins, R.S. Frazier, Implicit social cognition: from measures to mechanisms, Trends Cognit. Sci. 15 (4) (2011) 152–159.

[2] I.H. Bell, J. Nicholas, M. Alvarez-Jimenez, A. Thompson, L. Valmaggia, Virtual reality as a clinical tool in mental health research and practice, Dialogues Clin. Neurosci. 22(2) (2022) 169-177.

[3] E. Sajno, S. Bartolotta, C. Tuena, P. Cipresso, E. Pedroli, G. Riva, Machine learning in biosignals processing for mental health: a narrative review, Front. Psychol. 13 (2022).

[4] X.J. Wang, J.H. Krystal, Computational psychiatry, Neuron 84 (3) (2014) 638–654.

[5] K. Strimbu, J.A. Tavel, What are biomarkers? Curr. Opin. HIV AIDS 5 (6) (2010) 463.

[6] B. Definitions, Group W. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework, Clin. Pharmacol. Ther 69 (3) (2001) 89–95.

[7] Q.J. Huys, T.V. Maia, M.J. Frank, Computational psychiatry as a bridge from neuroscience to clinical applications, Nat. Neurosci. 19 (3) (2016) 404–413.

[8] M.E. Minissi, I.A. Chicchi Giglioli, F. Mantovani, M. Alcaniz Raya, Assessment of the autism spectrum disorder based on machine learning and social visual attention: a systematic review, J. Autism Dev. Disord. 52 (5) (2022) 2187–2202.

[9] G. Orru, W. Pettersson-Yeo, A.F. Marquand, G. Sartori, A. Mechelli, Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review, Neurosci. Biobehav. Rev. 36 (4) (2012) 1140–1152.

[10] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, A. T. Chronopoulos, Computational intelligence approaches for classification of medical data: state-of-the-art, future challenges and research directions, Neurocomputing 276 (2018) 2–22.

[11] A.J. Baxter, T.S. Brugha, H.E. Erskine, R.W. Scheurer, T. Vos, J.G. Scott, The epidemiology and global burden of autism spectrum disorders, Psychological medicine 45 (3) (2015) 601–613.

[12] M. Alcañiz Raya, I.A. Chicchi Giglioli, J. Marín-Morales, J.L. Higuera-Trujillo, E. Olmos, M.E. Minissi, L. Abad, Application of supervised machine learning for behavioral biomarkers of autism spectrum disorder based on electrodermal activity and virtual reality, Frontiers in human neuroscience 90 (2020).

[13] M. Alcaniz Raya, J. Marín-Morales, M.E. Minissi, G. Teruel Garcia, L. Abad, I. A. Chicchi Giglioli, Machine learning and virtual reality on body movements'

behaviors to classify children with autism spectrum disorder, J. Clin. Med. 9 (5) (2020) 1260.

[14] S. Goldstein, S. Ozonoff (Eds.), Assessment of Autism Spectrum Disorder, Guilford Publications, 2018.

[15] F. Thabtah, F. Kamalov, K. Rajab, A new computational intelligence approach to detect autistic features for autism screening, Int. J. Med. Inf. 117 (2018) 112–124.

[16] L.D. Wiggins, M. Durkin, A. Esler, L.C. Lee, W. Zahorodny, C. Rice, J. Baio, Disparities in documented diagnoses of autism spectrum disorder based on demographic, individual, and service factors, Autism Res. 13 (3) (2020) 464–473.

[17] N. Cavus, A.A. Lawan, Z. Ibrahim, A. Dahiru, S. Tahir, U.I. Abdulrazak, A. Hussaini, A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder, J. Personalized Med. 11 (4) (2021) 299.

[18] H.S. Nogay, H. Adeli, Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging, Rev. Neurosci. 31 (8) (2020) 825–841.

[19] M. Clemente, B. Rey, A. Rodríguez-Pujadas, A. Barros-Loscertales, R.M. Baños, C. Botella, C. Ávila, An fMRI study to analyze neural correlates of presence during virtual reality experiences, Interact. Comput. 26 (3) (2014) 269–284.

[20] K.K. Hyde, M.N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D.R. Dixon, E. Linstead, Applications of supervised machine learning in autism spectrum disorder research: a review, Review Journal of Autism and Developmental Disorders 6 (2019) 128–146.

[21] A. Ardalan, A.H. Assadi, O.J. Surgent, B.G. Travers, Whole-body movement during videogame play distinguishes youth with autism from youth with typical development, Sci. Rep. 9 (1) (2019) 1–11.

[22] A. Vabalas, E. Gowen, E. Poliakoff, A.J. Casson, Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis, Sci. Rep. 10 (1) (2020) 8346.

[23] R. Carette, M. Elbattah, F. Cilia, G. Dequen, J.L. Guerin, J. Bosche, Learning to predict autism spectrum disorder based on the visual patterns of eye-tracking scanpaths, in: HEALTHINF, 2019, February, pp. 103–112.

[24] J. Kang, X. Han, J. Song, Z. Niu, X. Li, The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data, Comput. Biol. Med. 120 (2020) 103722.

[25] D. Bone, M.S. Goodwin, M.P. Black, C.C. Lee, K. Audhkhasi, S. Narayanan, Applying machine learning to facilitate autism diagnostics: pitfalls and promises, J. Autism Dev. Disord. 45 (2015) 1121–1136.

[26] D.P. Wall, J. Kosmicki, T.F. Deluca, E. Harstad, V.A. Fusaro, Use of machine learning to shorten observation-based screening and diagnosis of autism, Transl. Psychiatry 2 (4) (2012) e100, e100.

[27] M. Alcañiz, I.A. Chicchi Giglioli, M. Sirera, E. Minissi, L. Abad, Biomarcadores del trastorno del especto autista basados en bioseñales, realidad virtual e inteligencia artificial, Medicina 80 (2020) 31–36.

[28] S. Parsons, Authenticity in Virtual Reality for assessment and intervention in autism: a conceptual review, Educ. Res. Rev. 19 (2016) 138–157.

[29] T.D. Parsons, Parsons, Clinical Neuropsychology and Technology, vol. 2016, Springer, 2016.

[30] M. Alcañiz, I.A. Chicchi-Giglioli, L.A. Carrasco-Ribelles, J. Marín-Morales, M. E. Minissi, G. Teruel-García, L. Abad, Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: a proof of concept for diagnosis, Autism Res. 15 (1) (2022) 131–145.

[31] E. Fombonne, Epidemiology of pervasive developmental disorders, Pediatr. Res. 65 (6) (2009) 591–598.

[32] M.E. Minissi, I.A.C. Giglioli, F. Mantovani, M. Sirera, L. Abad, M. Alcañiz, A qualitative and quantitative virtual reality usability study for the early assessment of ASD children, ANNUAL REVIEW OF CYBERTHERAPY AND TELEMEDICINE 2021 (2021) 47.

[33] M.E. Minissi, L. Gómez-Zaragozá, J. Marín-Morales, F. Mantovani, M. Sirera, L. Abad, M. Alcañiz, The whole-body motor skills of children with autism spectrum disorder taking goal-directed actions in virtual reality, Front. Psychol. 14 (2023) 1140731.

[34] M.E. Minissi, G.A.R. Landini, L. Maddalon, S.C. Torres, I.A.C. Giglioli, M. Sirera, M. Alcañiz, Virtual reality-based serious games to improve motor learning in children with autism spectrum disorder: an exploratory study, in: 2023 IEEE 11th International Conference on Serious Games and Applications for Health (SeGAH), IEEE, 2023, August, pp. 1–6.

[35] E. Pastorelli, H. Herrmann, A small-scale, low-budget semi-immersive virtual environment for scientific visualization and research, Procedia Comput. Sci. 25 (2013) 14–22, https://doi.org/10.1016/j.procs.2013.11.003.

[36] S. Wallace, S. Parsons, A. Westbury, K. White, K. White, A. Bailey, Sense of presence and atypical social judgments in immersive virtual environments: responses of adolescents with autism Spectrum disorders, Autism 14 (2010) 199–213, https://doi.org/10.1177/1362361310363283.

[37] R. Bradley, N. Newbutt, Autism and virtual reality head-mounted displays: a state of the art systematic review, Journal of Enabling Technologies 12 (3) (2018) 101–113.

[38] N. Newbutt, R. Bradley, I. Conley, Using virtual reality head-mounted displays in schools with autistic children: views, experiences, and future directions, Cyberpsychol., Behav. Soc. Netw. 23 (1) (2020) 23–33.

[39] D.D. Salvucci, J.H. Goldberg, Identifying fixations and saccades in eye-tracking protocols, in: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, 2000, November, pp. 71–78.

[40] R. Hershman, A. Henik, N. Cohen, A novel blink detection method based on pupillometry noise, Behav. Res. Methods 50 (2018) 107–114.

[41] D.W. Hosmer Jr., S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, vol. 398, John Wiley & Sons, 2013.

[42] M. Chita-Tegmark, Social attention in ASD: a review and meta-analysis of eye-tracking studies, Res. Dev. Disabil. 48 (2016) 79–93.

[43] A.N. Bhat, Motor impairment increases in children with autism spectrum disorder as a function of social communication, cognitive and functional impairment, repetitive behavior severity, and comorbid diagnoses: a SPARK study report, Autism Res. 14 (2021) 202–219, https://doi.org/10.1002/aur.2453.

Maria Eleonora Minissi is a cognitive psychologist and PhD candidate in New Technologies for Health and Well-Being at Polytechnic University of Valencia (Spain). Her PhD program is in cotutelle regime with the PhD program Education in Contemporary Society of Università degli Studi di Milano-Bicocca. She works at the University Research Institute of Human-Centered Technologies where she does research on the use of new technologies, particularly virtual and extended reality, to improve mental health and the assessment of clinical and neuropsychological disorders.

Alberto Altozano received his B.S in Physics and M.Sc in Data Science from the University of Valencia and the Autonomous University of Madrid respectively. He is currently a novel researcher at the University Research Institute of Human-Centered Technologies in the Polytechnic University of Valencia. His research interests include machine learning, deep learning, intelligent agents, reinforcement learning and cognitive science.

Javier Marín-Morales received his M.Eng. and Ph.D. degrees from the Polytechnic University of Valencia (Spain). He is currently a postdoctoral researcher at the University Research Institute of Human-Centered Technologies of UPV. His research interests include machine learning, affective computing, statistical biomedical signal processing, virtual reality, and psychological assessment. He has authored more than 50 international scientific contributions in these fields, which have been published in peer-reviewed international journals and conference proceedings. He is involved in several international research projects.

Irene Alice Chicchi Giglioli actually is the Chief Scientific Officer at Sincrolab. She got a Ph.D. in Technologies for Health and Well-being from the UPV. Her research activity focuses on evaluation, training and neurocognitive interventions through the combined use of Extended Reality technologies (VR/AR) and neuroscientific methods. To this end, she coordinates and investigates in different developments related to neurodevelopmental disorders. Dr. Chicchi Giglioli has more than 45 publications and she is an editor and reviewer of scientific journals and supervisor of Bachelor, Master and Doctoral Theses.

Fabrizia Mantovani, PhD is a full professor at Università degli Studi di Milano Bicocca (Italy) in the Department of Human Sciences for Education ''Riccardo Massa''. She is also the leading director of the Centre for Studies in Communication Sciences "Luigi Anolli" (CESCOM). Professor Mantovani does research on new technologies, emotion, interpersonal communication, and computing in social science, arts and humanities. She has been a coordinator of several European projects and national research projects related to the area, and she has more than 100 scientific publications.

Mariano Alcañiz, PhD is full professor of Computer Graphics at Polytechnic University of Valencia (Spain), where he founded the University Research Institute of Human-Centered Technologies. His research interest has always been focused on how virtual reality-related technologies can augment human abilities and performance in fields like medicine, health, education, and marketing. He has more than 100 scientific publications on these topics, and he has been a coordinator of various European projects and national research projects related to the area. He is also the National Program Coordinator of the Information Society Technology of the Ministry of Science of Spain.