Hindawi
International Journal of Computer Games Technology
Volume 2022, Article ID 6738068, 21 pages
https://doi.org/10.1155/2022/6738068

*Research Article*

# Recognition of Emotions for People with Autism: An Approach to Improve Skills

**Adilmar Coelho Dantas and Marcelo Zanchetta do Nascimento**

*Faculty of Computer Science, Federal University of Uberlândia, Uberlândia, Brazil*

Correspondence should be addressed to Adilmar Coelho Dantas; akanehar@gmail.com

Received 8 October 2021; Revised 8 December 2021; Accepted 18 December 2021; Published 15 January 2022

Academic Editor: Cristian A. Rusu



Autism spectrum disorder refers to a neurodevelopmental disorders characterized by repetitive behavior patterns, impaired social interaction, and impaired verbal and nonverbal communication. The ability to recognize mental states from facial expressions plays an important role in both social interaction and interpersonal communication. Thus, in recent years, several proposals have been presented, aiming to contribute to the improvement of emotional skills in order to improve social interaction. In this paper, a game is presented to support the development of emotional skills in people with autism spectrum disorder. The software used helps to develop the ability to recognize and express six basic emotions: joy, sadness, anger, disgust, surprise, and fear. Based on the theory of facial action coding systems and digital image processing techniques, it is possible to detect facial expressions and classify them into one of the six basic emotions. Experiments were performed using four public domain image databases (CK+, FER2013, RAF-DB, and MMI) and a group of children with autism spectrum disorder for evaluating the existing emotional skills. The results showed that the proposed software contributed to improvement of the skills of detection and recognition of the basic emotions in individuals with autism spectrum disorder.

## 1. Introduction

The representations obtained through the facial expressions of a person are important for social interactions in the society [1]. Emotional expression and eye gaze direction play crucial roles in nonverbal communication [2]. According to the authors in [3], facial emotion recognition is a fundamental concept within the field of social cognition. Individuals with autism spectrum disorder (ASD) may have characteristics related to repetitive behavior patterns, impaired social interaction, impaired verbal and nonverbal communication, and limited social-emotional reciprocity present from childhood [4]. People with ASD can present difficulty improving and developing important social-emotional skills throughout life [5]. Facial emotion recognition is a skill important for adult life but traditional methods and group interactions may not be promising for these tasks, as described in [6].

Several types of research have been developed to present tools based on the use of information and communication technology (ICT) for the improvement of skills related to emotions in an individualized way [6–9]. These solutions enable the user to have more comfort, entertainment, and insertion in the real context safely and playfully, as described in [10]. Moreover, these tools can be used on several devices and platforms, such as tablets, desktops, and smartphones. The development of techniques with ICT in digital game environments is being adopted in our society according to data published in the study presented in [4]. One of the approaches employed in this context is the use of serious games (SGs). According to the authors in [11], SGs provide entertainment and the means to improve skills related to facial emotion recognition. Moreover, computational techniques employed for facial emotion recognition can contribute to tools such as SGs.

There are challenges in the area for new tools that can encompass methodologies for improving emotion-related skills with SGs and facial emotion recognition techniques in a way that quantifies individual's data during the treatment sessions.

## 2. Contributions of This Work

In this paper, we present an SG to improve emotional skills through facial expression recognition methods for individuals with ASD. Firstly, a camera captures user's face. Then, the Haar-like feature algorithm is applied to detect regions of interest from user's face. The dlib library is employed to detect the face keypoints in the image. With the detected keypoints, the optical flow algorithm is employed to keep the locations independent of user's face movement. The descriptors obtained with the histogram of oriented gradients are extracted from the regions of interest. A convolutional neural network model is also employed over the regions of interest. The resulting data in the flattening layer are associated with the handcraft information and keypoints for classification in the softmax layer. This game includes characters to assist in the practice activities. Our SG explores multimodal aspects, such as facial expressions, vocal prosody, and body language. The user receives positive feedback after proper representation of the emotion, as well as the status and faults. A group of public domain datasets was employed for evaluation of the facial emotion recognition methods. In a second step, a dataset of volunteers with ASD was selected to evaluate the game and investigate the contribution of the tool to the improvement of emotional skills.

The contributions of this paper can be summarized as follows:

(i) We present an approach for feature extraction and emotion classification based in a hybrid model (handcraft and learning features)

(ii) We developed an SG involving characters that express emotions and can assist in improve facial expression skills to individuals with ASD

(iii) We design a tool that allows for the capture of information from intervention sessions to help specialists in the improvement of individuals with ASD

The article is organized in sections with information and details: Section 3 describes about the theoretical background. Section 4 presented the related work about strategies to improve facial expression skills. Section 5 shows the methodologies used in the various stages of the proposed approach. Section 6 presents the results and discussions regarding the performance of the various steps of the work. Finally, the conclusions are described in Section 7.

## 3. Theoretical Background

*3.1. Facial Expressions of People with Autism.* The term autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects social communication and behavior in children and adults. Lorna Wing and Judith Gould were the first researchers to observe and define the term ASD in 1979 [12]. According to data published by the World Health Organization, one in 160 children have characteristics that may be related to ASD [13]. In 2020 in the United States, the Centers for Disease Control and Prevention reported an increase of 178% in the number of children diagnosed with ASD compared to data published in 2000 [14]. In Brazil, with a population of over 200 million, there are estimated 2 million individuals with ASD [15].

The perception of the face is one of the most important skills developed in humans [1]. Factors, such as behavioral difficulties that occur due to lack of visual attention, compromise the socialization of children with ASD [16]. Depending on the level, individuals with ASD can also have difficulty interacting in collaborative environments and performing actions such as teamwork or public speaking due to the difficulty of muscle analysis and face expression recognition [17, 18]. The development of software with the aim of contributing to this task for ASD individuals has been increasingly employed in our society according to data from an investigation into the global panorama of solutions published in the work presented by [4]. Software for this task that employs SG for the process of improving these skills are being explored in this area.

*3.2. Serious Games to Improve Skill of Facial Expression.* Serious games are primarily aimed at learning or developing skill [19]. This term was proposed and defined by Clark [19] in 1970 the name "Serious Games" for games developed with the educational objective and not only for fun and entertainment, but that does not mean that serious games (SGs) should not be fun for their users, but rather provide a association of these essential elements in this type of application. According to the authors in [11], SGs provide the user with entertainment and exploring improving skills related to emotional expressions. Software for this task is investigated in several studies in the literature as highlighted in these studies [20–22].

*3.3. Facial Emotion Recognition.* Computational techniques for facial emotion recognition (FER) based in computer vision and image processing can contribute to tools such as SGs. These techniques involve face information acquisition, feature extraction, and classification steps. The information acquisition step involves the detection of areas of the face from an image or video. After this step, regions must be selected to determine the area of interest and eliminate the background of the image. From these regions, features related to morphological or nonmorphological information are extracted. According to [23], the performance of a system for human FER depends on the feature extraction methods used. The information extracted must be highly discriminative among the variations of the classes and weakly discriminative within the same class variation. Moreover, the procedure for calculating the features should be described in low-dimensional space and be robust against variation in illumination and noise [24].

Tools for FER can be categorized into methods that explore geometric information, brightness, or texture [25]. Recently, the strategies based on deep learning have been employed in problems for FER and facial emotion expression in computer vision [26–28]. Methods based on deep learning have achieved remarkable success rates in FER

studies [27, 29]. The high accuracy achieved in these recognition tasks can be attributed to large-scale labeled datasets such as AffectNet [26] and EmotioNet [30], which enable convolutional neural networks (CNNs) to explore generalizable representations [31]. These methods with approaches based on CNNs have contributed to the challenges with regard to new approaches for FER.

## 4. Related Works

The use of computer technologies to improve skills has increased in recent years. The work of [32] shows through a systematic review the use of technology in improving the conceptual skills, practical skills, social skills, and general skills to people with ASD. This study shows that people with ASD tend to be motivated by new technologies, and these solutions can be fun in the process of improving skills. The use of technological advancements such as artificial intelligence and augmented reality undoubtedly provides a comfortable environment that promotes constant learning for people with ASD. Moreover, the study in [33] shows a systematic mapping study of the use of technologies in emotion recognition in children with ASD. The study analyzed the main techniques employed and recommendations for user interface and equipment. These strategies can provide flexibility, accessibility, and easy adaptation to real-world scenarios. This section presents the main contributions that have investigated SG strategies for individuals with ASD.

The authors in [34] proposed a system to improve facial expression skills for individuals with ASD using 3D animations overlaid on the face. In this tool, images obtained from regions (frontal and lateral) of user's face were used to elaborate a 3D face mask. With the use of the facial action coding system (FACS) theory, representations of facial micromovements in the 3D animations were captured from the participant. The overlay of the 3D face representation allowed to evaluate the representation of emotions by the participants during the representation. This tool was evaluated with three participants during seven sessions. These sessions are aimed at collecting skills related to emotion recognition based on facial expressions. The results showed an accuracy of 89.94% for recognition of basic emotions. According to the authors, the use of the 3D representation contributed to the evaluation by allowing the participants to maintain their attention and focus. The authors highlight that the system still has restrictions due to the small number of participants as well as the small number of quantitative metrics, which makes further analysis necessary to evaluate the effectiveness of the system.

In 2016, a digital storybook with characters, video element modeling, and augmented reality (AR) was proposed for emotion recognition in the study of [35]. Computational technologies were adopted in order to attract children's attention and focus during sessions. The features available in AR allowed to extend the social features of the story and to direct participants' attention to the important information present in the digital book. The tracking of individual's eye movement was analyzed by the expert. In the sessions, initially, a printed book with visual images was presented. Then, characters modeled in AR of the scenes in the book were presented with facial expressions of emotions to assist the individual in representing the emotion. The evaluation was conducted in three stages: baseline, intervention, and maintenance. The authors observed that the system helped the children to recognize and understand emotions from facial expressions. The experts reported that the tool raised the attention of users with AR use. The authors reported that quantitative measures were not obtained for the performance improvement of emotion.

An SG game for smartphone devices capable of assisting in the development of emotional skills for individuals with ASD was presented by [36]. The tool was composed of a communication interface for the smartphone (app) and a server with services for requests, emotion processing, and emotion classification. The captured images were sent for analysis and classification of emotions on the server. The server had a convolutional neural network trained with 15,000 images of the basic emotions. The tool was evaluated with nine users, ages 18 to 25, during four intervention periods and two-week intervals. According to specialists, the results showed that the participants improved their ability to recognize and express emotions. However, the authors did not employ metrics for quantitative analysis of the improvement in the skill. They also did not analyze the performance of the tool in relation to image processing and classification algorithms.

In [20], the authors developed a game called Emotiplay to assist in improving the emotional skills of children with ASD. The game was designed to run in a desktop environment with access to a browser using HTML5, CSS, and JavaScript technologies. This game was built with animated characters that represented everyday scenes from social life and questionnaires to be answered about the emotions observed in the scenes. Among the proposed activities, the user could recognize emotions through facial expressions, gestures, speech, and other sociobehavioral characteristics. The evaluation with volunteers from different countries analyzes the performance of the tool in people of different cultures and sociobehavioral contexts. This investigation provided quantitative results, and the data showed an evolution of the emotional vocabulary of participants with ASD.

In the work of [21], the authors developed an application using machine learning techniques to help individuals with ASD develop the only four basic emotions: neutral, joy, sadness, and anger, The representations were captured on video and analyzed by Viola and Jones technique to detect the FKs of users' faces. These were used for classification with the random forest algorithm. To make the game attractive, the authors developed scenarios involving social situations that demanded the representation of emotions. A qualitative evaluation was not performed; only questionnaires were used to analyze the level of satisfaction regarding the use of the tool. The authors reinforce that new evaluations should be carried out in future work.

In [37], the authors developed an application that simulates a mirror with a webcam in which convolutional neural networks are employed to analyze the images that are captured by a camera and compare it with the one that the

patient should perform to detect five basic emotions. The tool has been evaluated with people with autism and monitored by experts. In the experiments, the professionals were able to evaluate the acceptance rate and various usability information of the tool. The tool made it possible to carry out the treatment sessions during the period of social isolation. As restrictions, the proposed method presents the need for a specific hardware and proper configurations, which for certain people can be an obstacle for use during daily routine. Thus, it is reported by the authors that approach should be improved in relation to the quality of the interfaces.

In the literature, different works have demonstrated the importance of adapting and evaluating tools in environments more suitable for individuals with ASD. Developing tools that can capture information from user's face in a process of emotion detection and recognition can generate important data for the specialist. These specialists can track the progress of this skill at each session and further explore customization to each user's limitation during the skill development process.

## 5. Methodology

The proposed tool is a free software developed to improve emotional skills from the recognition of facial expressions in people with ASD. Figure 1 presents the modules of this tool: SG module, detection and classification module, and data module. Firstly, the initial instructions of the SG are presented to the user in the SG module. Then, the knowledge evaluation step is presented to the user to evaluate the emotional skills. After this step, the user receives new information to express emotions following the instruction of the virtual character.

The facial expressions of the user are captured by the webcam, which can be obtained from a computer or smartphone. In this work, a 720p Microsoft webcam H5D00013 was used to capture the expressions which were sent to the detection and classification module. In this step, the information from ROIs on user's face is extracted and classified. The data obtained are stored for evaluation and analysis by the specialist. Information on the attention level, the number of hits, the number of errors, and the time required to express the emotion is stored (see extra parameters, Figure 1). An interface allows the specialist to analyze the data in each session in order to evaluate the improvement of the skills in individuals with ASD.

The proposed game was developed for children aged 6 to 13 years old, but it can be applied to adult people. According to the authors in [32, 33], when a game is developed for children with ASD, some characteristics need to be adopted: the user interface and elements such as color tones and sounds. The narrative during matches and characters with customization, such as choices of clothing and accessories chosen by the user, are important in the game. These features allow the SG to be more attractive and consequently improve the process of skill improvement.

Our scheme was developed using HTML version 5.2 [38], JavaScript, and Python programming language version 3. In addition, the web application was developed in a responsive manner allowing it to run on Desktop or Smartphone devices. In this SG, no predefined game engine from the literature was used, in order to allow for customization based on the particular characteristics of the user. The algorithms were implemented using a cloud computing architecture, with 4 GB of RAM, a 1 TB HD, and a 5 GB GPU. The videos were captured with 30 frames per second in the RGB color model.

*5.1. SG Module.* Before the SG is launched, the user chooses one of the available characters, as shown in Figure 2. For this software, the characters Ana and Juninho were provided. Figure 2 presents the character Ana that helps the user during the stages of the game.

In this SG, animated characters are used to express the six basic emotions. These animations were developed based on an analysis of the videos in the dataset [39], which contains 2,900 videos and a set of images of 75 people expressing each of the six emotions, according to FACS theory. The characters take between 40 and 45 seconds, on average, to express these emotions. The animations and the quality of the representations of the emotions were evaluated by specialists who work with ASD individuals. Figure 3 presents one of the characters used to express emotions.

After this character has been defined, an interface allows the accessories to be customized (see Figure 4). This customization step allows interaction so that the user can have motivation and more attention according to [40].

The user is then directed to the first stage of the game. At this stage, the user must identify the emotions shown in a sequence of three images of people. The individual watches six videos with a person representing each of the basic emotions during one minute. The videos with the basic emotions are presented randomly. At the end of each video, the images are shown for the user to select the emotion as shown in Figure 5. During five minutes, the user can choose the image that represents the emotion. In this phase, no feedback is presented to the individual with ASD. The data captured during this stage is shown only at the end of this phase. This strategy was adopted in order not to demotivate the user during this assessment. Participant's score is displayed in the data module so that specialists can analyze and track participant's progress.

*5.2. Evaluating Serious Games with People with ASD.* After the evaluation step, the user begins the SG, following the path in the world of emotions as shown in Figure 6. In this step, the basic emotions will be presented, and the character will assist the user.

A camera will capture the video of the user expressing the emotion, and the detection and recognition algorithms analyze this representation. User's performance during the emotion expression process can be accompanied by a status bar. Each facial muscle expressed adequately allows for the filling of the status bar. After the status bar has been filled in completely, the system will allow the user to proceed to the next phase, and a new emotion will be presented. In Figure 7, the SG interface is illustrated with the target emotion (upper right, yellow borders), and in the upper left

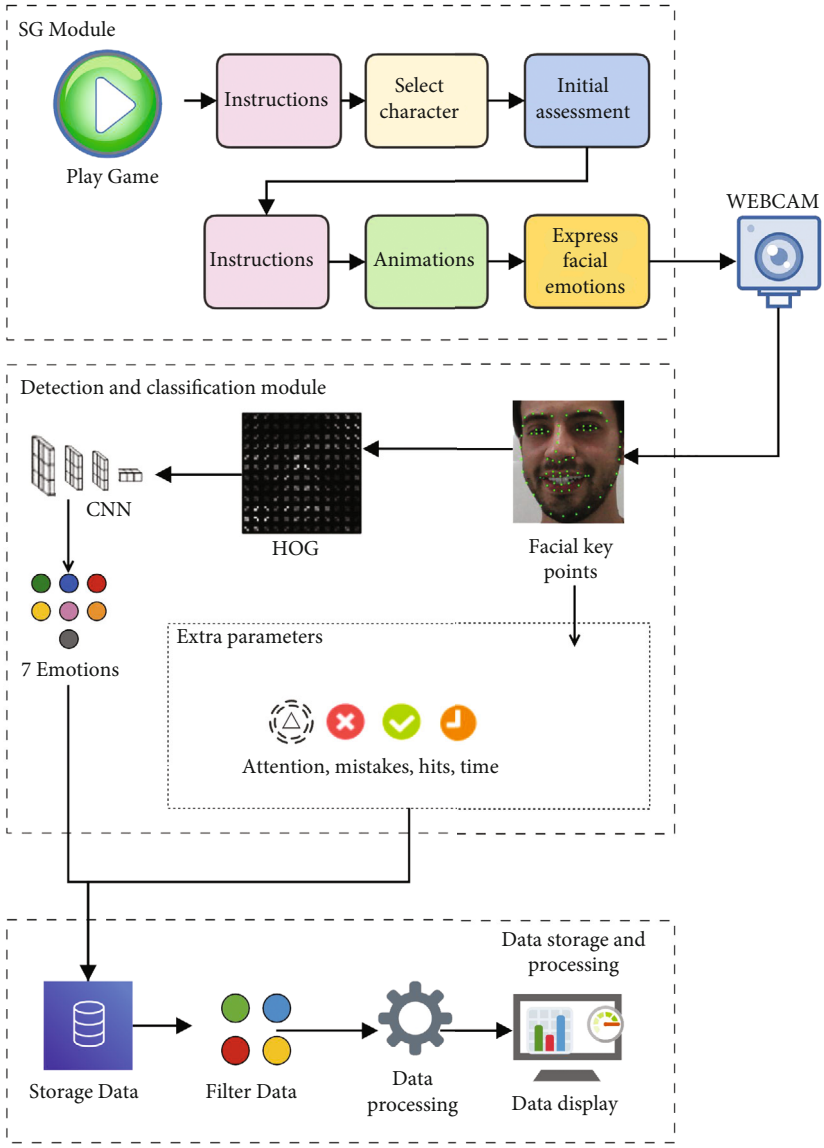


Figure 1: Flowchart of the steps of the proposed tool for improving emotion skills.

(purple borders), the character expresses the emotion based on information from the user (center of the screen). The points marked on the face allow the user to monitor the expression of emotion.

When a user is playing a match of this game, this participant expresses the target emotion as represented by the avatar defined in Figure 7 (see right side of the interface). In Figure 7, the tutor (see left side of the interface) shows the movements needed to represent the facial expression of the required emotion. These movements are shown as represented in Figure 3. The movements of user's face and the keypoints (marked in white) are presented in the central region of the interface. Information such as time, eye-tracking of the regions of the screen viewed by the user, the muscles activated in the face, and the hit and miss numbers is stored. This information can be analyzed by the healthcare professional during the follow-up of a match.

When the user presents a sequence of errors in the expression of an emotion, hints are presented with the objective of helping the user correctly represent that emotion during a match of this game. However, when the participant does not manage to represent the emotion, the tool does not allow the user to advance to the next phase. The participant is guided to the training phase with images and animations. In the training phase, the tutor expresses the emotion, and a voice speaks the name of the emotion presented. There is also an image of a person showing that emotion (see Figure 8). The six emotions are presented in this phase, and this process is repeated twice for each emotion. After this step, a message is displayed asking to continue training or to return to the game phase. In this process, the training information is stored for the specialists for the purpose of assisting them in making decisions and helping the participants.


Figure 2: The interface presents the character Ana that was chosen to aid in the process of improving emotions.

Figure 3: A character expressing an emotion that was employed in the game.

*5.3. Detection and Classification Module.* At this module, the methods responsible for processing and classifying emotions are presented in Figure 9.

Firstly, the video is captured in the RGB color model and converted to grayscale. This process allows decreasing the color scale present in the frames as reported in the studies proposed by [41, 42]. The frames are analyzed by the algorithm proposed by Viola-Jones for feature detection of regions of user's face [43]. This method analyzes the appearance of the objects and searches for descriptors [43, 44]. Then, initially, it is necessary to perform a calculation on all of the pixels of the face. For this, an integral image was computed in this stage where $I(x, y)$ represents the value in the integral image and $i(x, y)$ represents the input image which is given by:

$$I(x, y) = i(x, y) + I(x - 1, y) + I(x, y - 1) - I(x - 1, y - 1). \quad (1)$$

After obtaining this information, the method is employed to obtain edge features, line features, and center-surround features (see Figure 10). In this method, we also classify the features using the Adaboost algorithm (Figure 11) and a tree structure defined by a cascade of classifiers [45]. If all of the classifiers accept the image of user's face, then it is displayed with the detected regions (see Figure 12). In this stage, the ROIs detected are the eyebrows, eyes, lip, nose, and jaw.

Facial keypoint (FK) detection model based on the Dlib library was employed in these regions of face [48–50]. In this stage, the dlib function called *shape_predictor* is employed. This function can take a ROI as input and then output a set of locations with 68 FKs, which are shown in Figure 13:

(i) Right eyebrow—points 18, 19, 20, 21, and 22

(ii) Left eyebrow—points 23, 24, 25, 26, and 27

(iii) Right eye—points 37, 38, 39, 40, 41, and 42

(iv) Left eye—points 43, 44, 45,46, 47, and 48

(v) Nose—points 28, 29, 30, 31, 32, 33, 34, 35, and 36


Figure 4: Interface for character customization with options for clothing and accessories (cap, hat, and glasses).

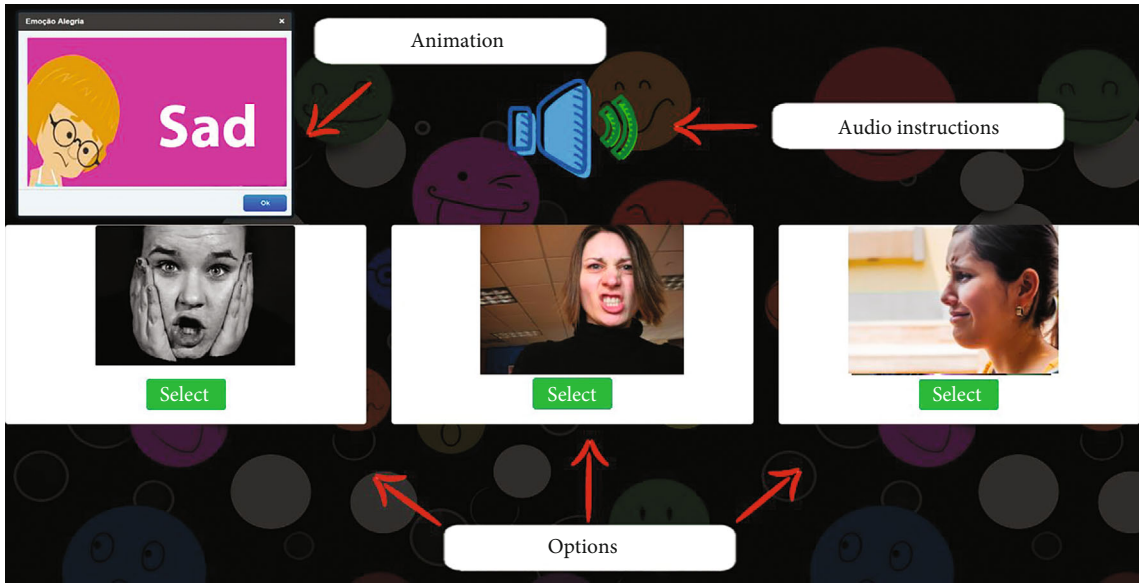


Figure 5: Game interface for evaluating skills related to basic emotions.

Mouth:

(i) Upper outer lip—points 49, 50, 51, 52, 53, 54, and 55

(ii) Upper inner lip—points 61, 62, 63, 64, and 65

(iii) Lower inner lip—points 61, 65, 66, 67, and 68

(iv) Lower outer lip—points 49, 55, 56, 57, 58, 59, and 60

The Euclidean distance is computed among the detected FKs for detected face regions. Due to the movement of muscles in the face, we apply the algorithm proposed in [51] to compute the optical flow and to estimate the movement of the face in the images making up the video sequence. If the user moves during the emotion investigation process, the algorithm can describe these movements. Finally, the information from the FKs is stored in the flatting vector before the fully connected layers of CNN.

The histogram of oriented gradient (HOG) method was computed to the detected ROIs. The HOG method is a feature descriptor that allows determining the occurrences of gradient orientation in localized portions of an image for extracting object features [52]. Gradients are computed for each image per block where the block is obtained of a grid

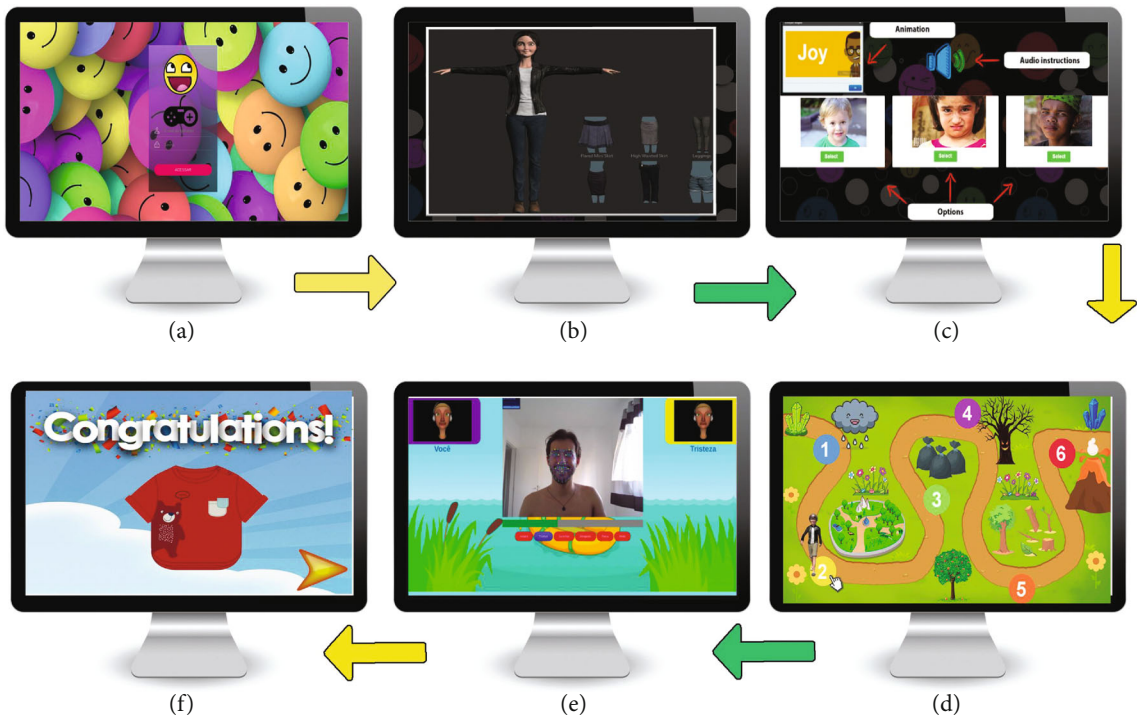


Figure 6: Gameplay of the proposed SG: (a) user identification (ID); choice of one of the characters; (c) evaluation of skills about the basic emotions; (d) scenario with 6 different interactions about the emotions; (e) user can improve the emotional skills; and (f) each emotion expressed correctly, a surprise item is present to the character that can advance game map.


Figure 7: Interface for training the emotional competencies of facial expressions and functionality.

of pixels composed from the magnitude $g$ and direction $\Theta$ of the change in the pixel intensities:

$$\Theta = \arctan \frac{g_y}{g_x}, \qquad g = \sqrt{g_x^2 + g_y^2}. \tag{2}$$

The terms $g_x$ and $g_y$ are the horizontal and vertical components of the pixel intensity change, respectively, in the equation. The feature was calculated in blocks of size $8 \times 8$ pixels for the ROI of $48 \times 48$ pixels. The values for each block are quantized into 9 bins based on the gradient and the magnitudes of the pixels. These values are stored into the feature vector before the fully connected layer.

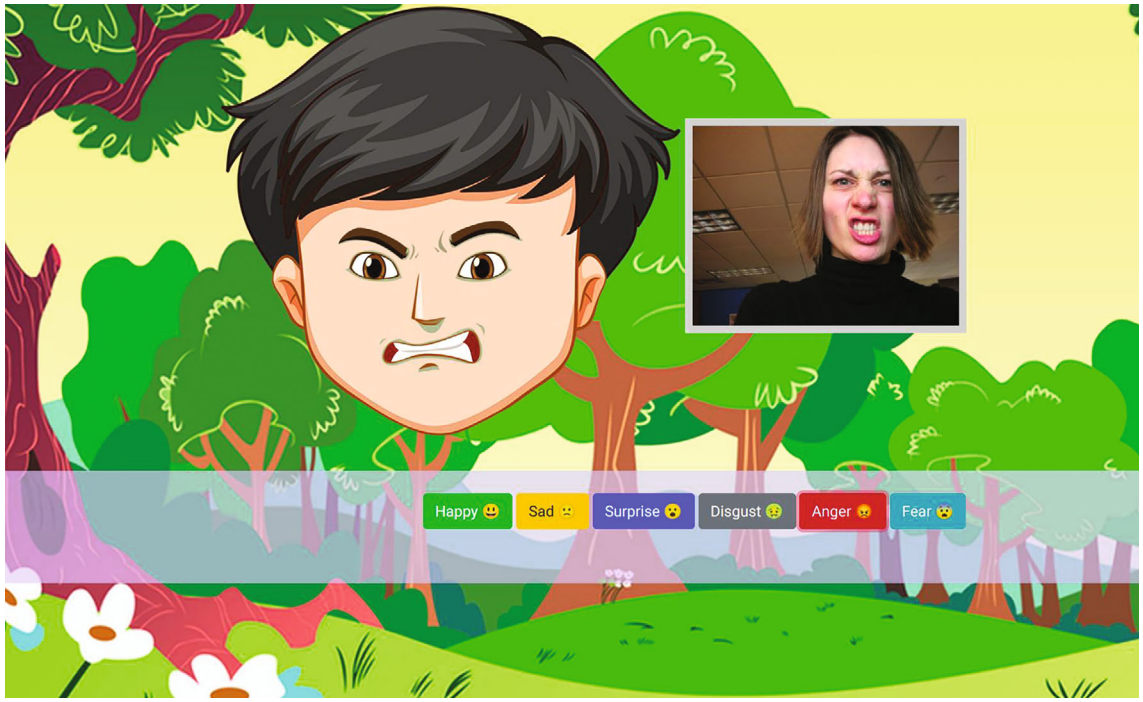


Figure 8: Training interface for emotion recognition and expression skills.

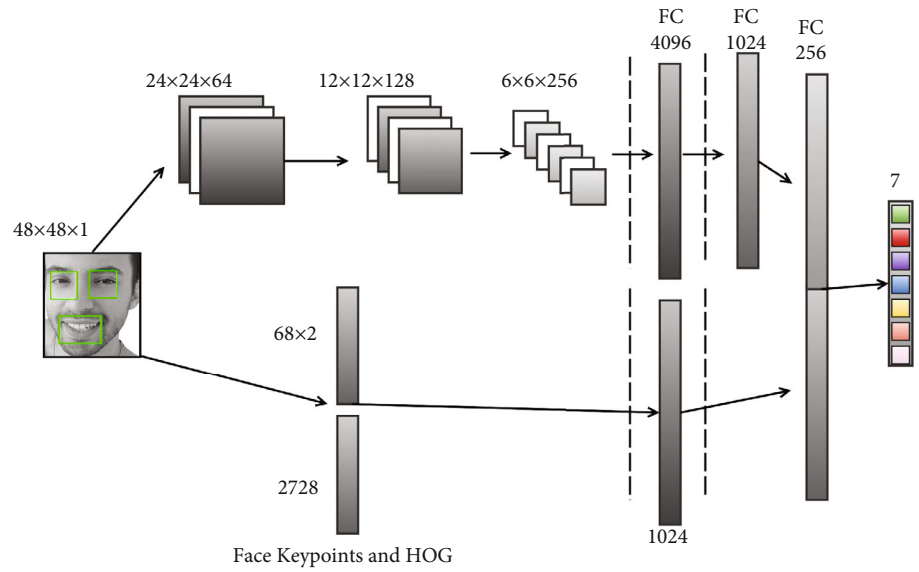


Figure 9: Proposed model for obtaining face information for emotion classification.

The ROIs are also given as input to a CNN approach based on the MobileNet architecture [53, 54]. This model employed depth-separable convolutional (DSC) layers and hyperparameters called width and resolution multipliers that address the computational resource limitations (latency and size) of the applications [53, 54]. These features aim to split the convolution step into two operations—depth-wise convolution and point-wise convolution ($1 \times 1$ size)—thus reducing the set of parameters in the convolutional layers. Figure 14 shows the difference between a traditional convolution filter and a DSC filter, where $D_k$ is the kernel size, $M$ is the number of input channels, and $N$ is the number of output channels.

The batch normalization and activation operations are employed after the convolutional layers. Batch normalization is a technique used to standardize the inputs and contributes to the training stage of the model. For the activation stage, the rectified linear unit (ReLU) operation [53] was used. The pooling layer was employed with the function of simplifying the information in the output of the convolutional layers. In this step, the max-pooling operation was used, where only the largest value is extracted for

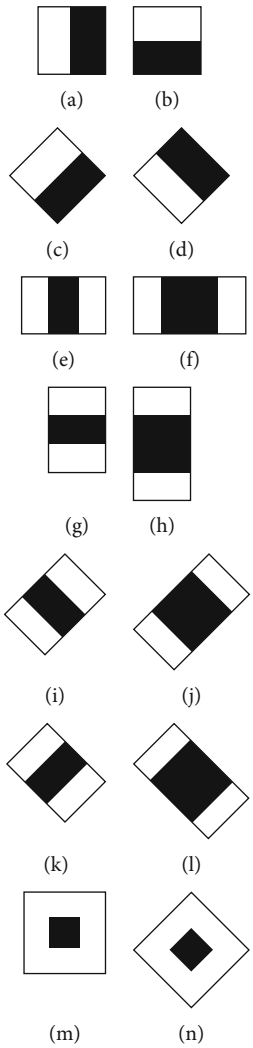


Figure 10: Features extracted with the Haar-like: (a–d) edge features; (e–l) line features; and (m, n) center surround features, adapted from [46].

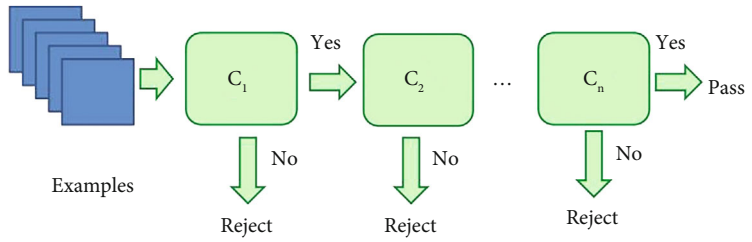


Figure 11: Representation of structure classification based on the cascade model, adapted from [47].

the output. This data summarization serves to reduce the number of weights to be learned and avoid overfitting. In the final stage, the flattening and softmax layers were employed [53]. In the proposed approach, an adaptation was performed where three fully connected (FC) layers were inserted (as shown in Figure 9). The features obtained by the convolutional layers and the handcraft features (FK and HOG) were employed in the classification. According to the authors at [55], the strategy of using handcraft features can help the CNN architecture with the ability to learn problem-specific features and consequently improve results. The model was trained using the 70% ROIs from each dataset during 1000 epochs using an Adam optimizer and a batch size of one image patch. The configuration for the Adam optimizer used the value provided by the Keras framework (default learning rate of 0.001).

*5.4. Data Module.* This tool also has a module for storing game match data. This module has an interface that allows analyzing the data that was captured in the emotion

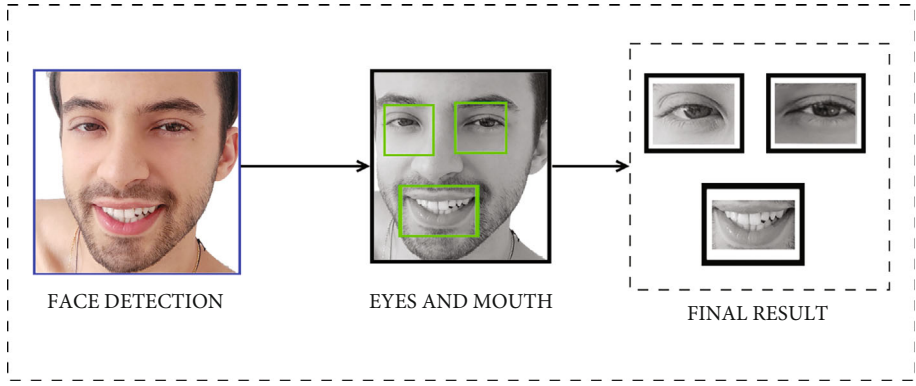


Figure 12: Stages of the detected regions of the face (face, eyes, and mouth).

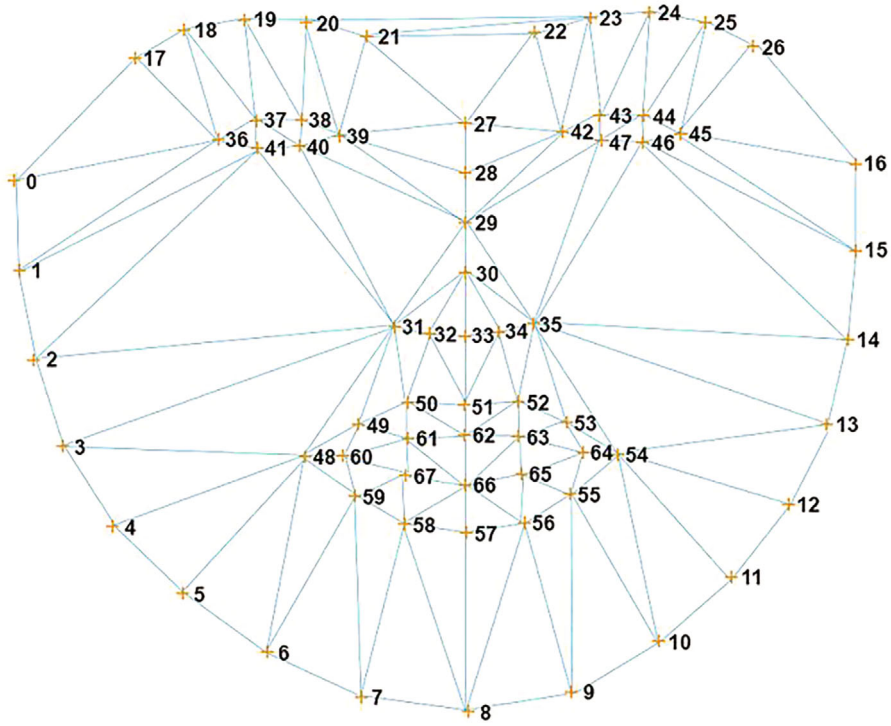

Figure 13: Markers inserted in the image of user's face.

improvement process (see Figure 15). The specialist can use this data to evaluate the improvement of skill of the treatment sessions of each individual. Moreover, the specialist can evaluate the number of sessions, the emotions in which the individual has difficulty, the time spent expressing the emotions, and the muscles activated for each emotion.

*5.5. Evaluation Strategies.* The hold-out method was employed to evaluate the proposed methods at the training and test stages. The image dataset was split into subsets: the training with 70% images and the test with 30% images. The evaluation of the classification performance was performed using the metrics accuracy, recall, precision, and *F* 1-score. The accuracy metric indicates the performance of the model in reaction to the instances that the model classified correctly. Recall evaluates the number of positive samples that have been correctly classified. Accuracy of a model can be defined as the ability to predict positive values. The metric *F*1-score is defined by:

$$F_1 = \frac{\text{Precission} \times \text{Recall}}{\text{Precission} + \text{Recall}}. \quad (3)$$

*F*1-score is used when it is desired to find the balance between the metrics precision and recall, which adapts to the solution as can be observed in the studies of [56–58].

To evaluate the proposed SG, we employed a multiple baseline design across participants to demonstrate emotional skill. This strategy is divided into baseline, intervention, and maintenance phases. This assessment model is widely used in medical and psychological research and for individuals with ASD, as reported in [34, 35, 59–61].

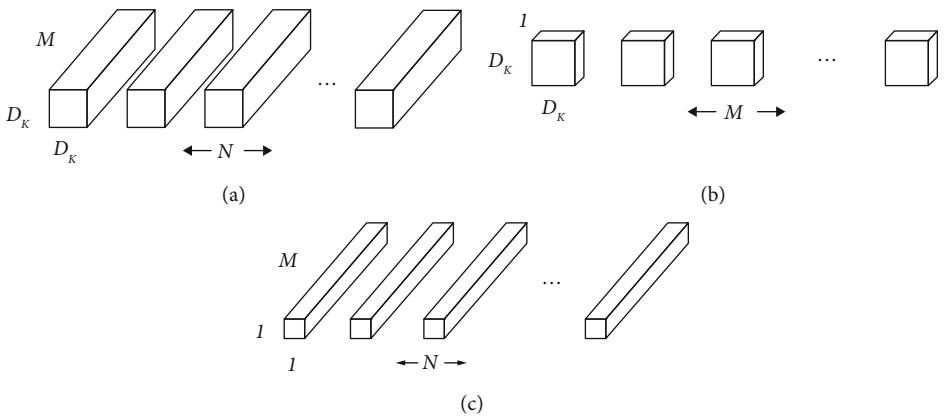


FIGURE 14: Representation of convolutional filters by a traditional method (a). The two-stage operations: in-depth (b) and point-to-point (c). Fonte: [53].

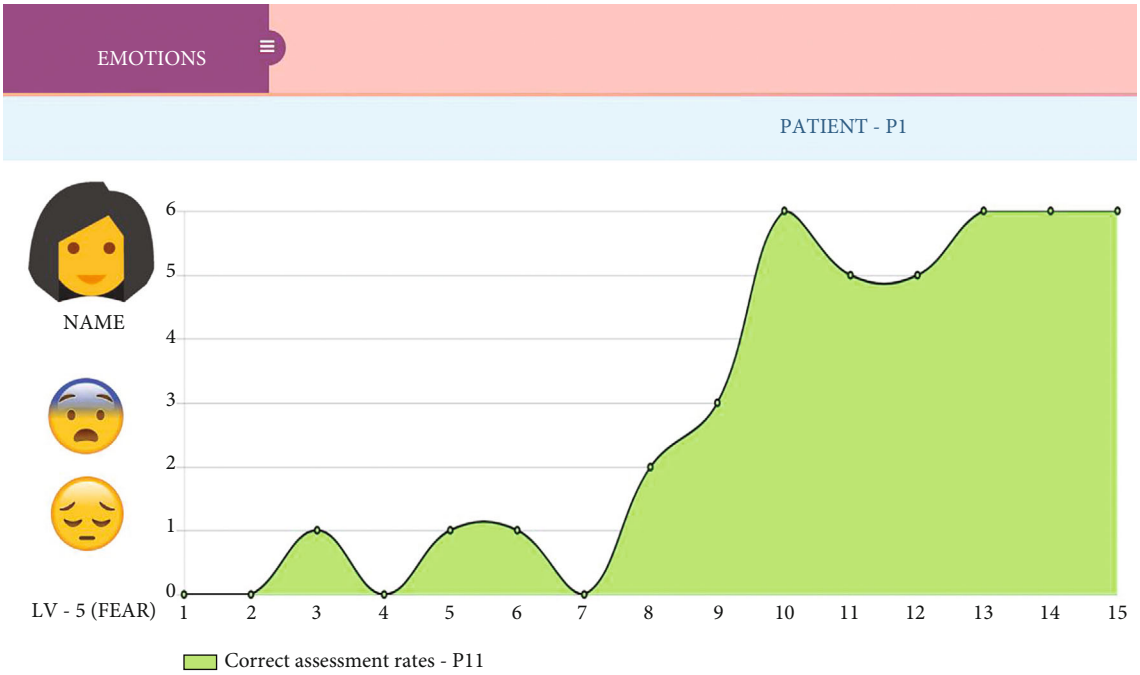


FIGURE 15: Dashboard with user performance information provided to the specialist.

The information about the skill of each individual is collected in the baseline stage. In each section, six videos are shown to each participant with a person representing the basic emotions during one minute. The videos are presented in order randomly to avoid that the participant memorizes the sequence. At the end of each video, three images with different emotions were presented to choose the emotion that was presented in the video. During five minutes, participants chose the image that corresponded to the emotion shown in the video. The baseline sessions occurred once a weekday up to five times a week and for approximately 30 minutes.

In the first session of the intervention phase, participants were instructed by the researchers that the character will express the target emotion. Participants were also oriented about the scoring and functionality of the interface. After the instructions, participants were to express the emotion presented in the video. The facial mimics of the participant were captured with a camera and represented by the character. The participant has five minutes to represent each emotion. The individual observed the representation of the emotion through a progress bar. When the emotion was correctly expressed, a reward is provided to the player. This phase was also employed once a weekday up to five times a week in sessions of approximately 30 minutes. The videos with the basic emotions were presented randomly. The characters used in the intervention session were different from those used in the baseline and maintenance phases to avoid memorization by the participants.

Finally, maintenance sessions were conducted four weeks after the intervention phase to demonstrate the

emotional skills of the individuals. The time interval between intervention and maintenance is aimed at checking whether the improved or developed skills remain in the individuals. During this step, all individuals watched the videos with the emotions presented randomly. The participants chose the image that corresponds to the emotion shown in the video. These sessions occurred once per weekday up to five times a week and for approximately 30 minutes.

## 6. Experiments and Results

This section presents the public datasets and evaluation of the computational algorithms for emotion recognition (see Section 6.1) section as well as the experiments performed to evaluate the improvement of emotional skills with the use of the game with individuals with ASD (see Section 6.2).

### *6.1. Evaluation of the Emotion Recognition and Detection Methods*

*6.1.1. Public Datasets.* For evaluation of the detection and recognition, algorithms were employed the public domain datasets: CK+ [62], FER2013 [63], Real-world Affective Faces Database (RAF-DB) [64], and MMI Facial Expression Database [65].

The CK+ database [62] has images of people expressing basic emotions, which are divided into joy (324), sadness (253), anger (183), surprise (328), disgust (182), and fear (182). This basis has been employed in work in the literature for evaluating the emotion detection and recognition algorithms [56–58]. This database is composed of videos/images from over 200 adults between the ages of 18 and 50 of Euro-American and African-American individuals. The images were taken in periods, where the capture begins with the face in neutral and captures the transition to the required emotion. These images were captured in grayscale or RGB with a resolution of 640 Ã— 490 or 640 Ã— 480 pixels. The second database investigated was FER2013 [63], which has a total of 35,887 grayscale $48 \times 48$ pixel images of faces, divided into the following categories: anger (4593), (547) disgust, (5121) fear, (8989) joy, (6077) sadness, (4002) surprise, and (6198) neutral. The faces were captured so that they were centered and occupied the same proportion of space in each image. The images are not distributed in temporal states, being considered the maximum state of each one. The third database investigated was RAF-DB [64]. This is a large-scale facial expression image base with approximately 30,000 facial images obtained over the Internet in RGB color standard and a size of $640 \times 480$ pixels. From this dataset of images, a total of 12,271 images are divided into surprise (1290), fear (281), disgust (717), joy (4772), sadness (1982), anger (705), and neutral (2524). An important feature of this dataset is that the images have great variability in the age of the participants and variations in ethnicity, head poses, lighting conditions, occlusions (e.g., glasses, facial hair, or self-occlusion), postprocessing operations (e.g., various filters and special effects), etc. The last dataset evaluated was the MMI Facial Expression Database [65]. This database consists of over 2,900 high resolution $640 \times 480$ RGB samples. In this database only 174 samples contained the necessary annotations for emotion classification. The images used are distributed as follows: joy (35), disgust (25), fear (24), anger (27), surprise (35), and sadness (28).

Figure 16 shows images of individuals expressing emotions from each of the databases used: (a) CK+, (b) FER2013, (c) RAF, and in (d) MMI. In the images presented in Figure 16, it is possible to observe a good distribution of characteristics related to gender, age, and ethnicity of the individuals. These features contribute to the evaluation of the robustness and generalization of the algorithms over the various contexts.

*6.1.2. Analysis of the Face Emotional Recognition Algorithms.* In this experiment, an investigation was performed with the association of the extracted features employed for emotion classification on the images. Table 1 shows the results of the average accuracy for the image datasets. The values shown in Table 1 indicated that the association between the CNN learned features and the handcraft descriptors contributed to increasing the values of the accuracy. The best performance with the accuracy metric was obtained on the CK+ dataset (accuracy = 98.8%). Table 2 shows the results for the proposed approach and other CNN architectures. As better results for the proposed approach were obtained with the CK+ dataset, this dataset was chosen for comparison with the other models: the proposed model was 2.03% and 5.03% better than the inception architecture and MobileNet model, respectively.

The quantitative metrics for evaluating the detection and recognition steps for each of the emotions are presented in Table 3 with the CK+ dataset. These results show that the worst measurements occurred with the sad, fear, and disgust emotions.

In general, the results presented by the proposed algorithm were relevant, with an average accuracy rate of 0.98. The values obtained through the recall metric were satisfactory, with an average value of 0.99, showing that the model was able to detect a relevant number of cases from the CK+ database. The $F1$ score metric considers false positives and false negatives and is generally more important than other measures such as precision, especially if you have an uneven class sample distribution, as in our experiments.

Figure 17 presents the confusion matrix of the results obtained in the experiments for the dataset. In the matrix, it is possible to observe that the emotions with the highest number of errors are sadness, disgust, and fear. These are emotions in which some of the systems present in the literature show difficulties in classification similar to the proposed method. This was observed in the studies proposed by Jain et al. [66] and Wu and Lin [67]. In general, the proposed method shows promising results compared with studies in the literature. For this, these methods were employed as part of the SG for detection and recognition of emotions during the process of improving emotional skills.

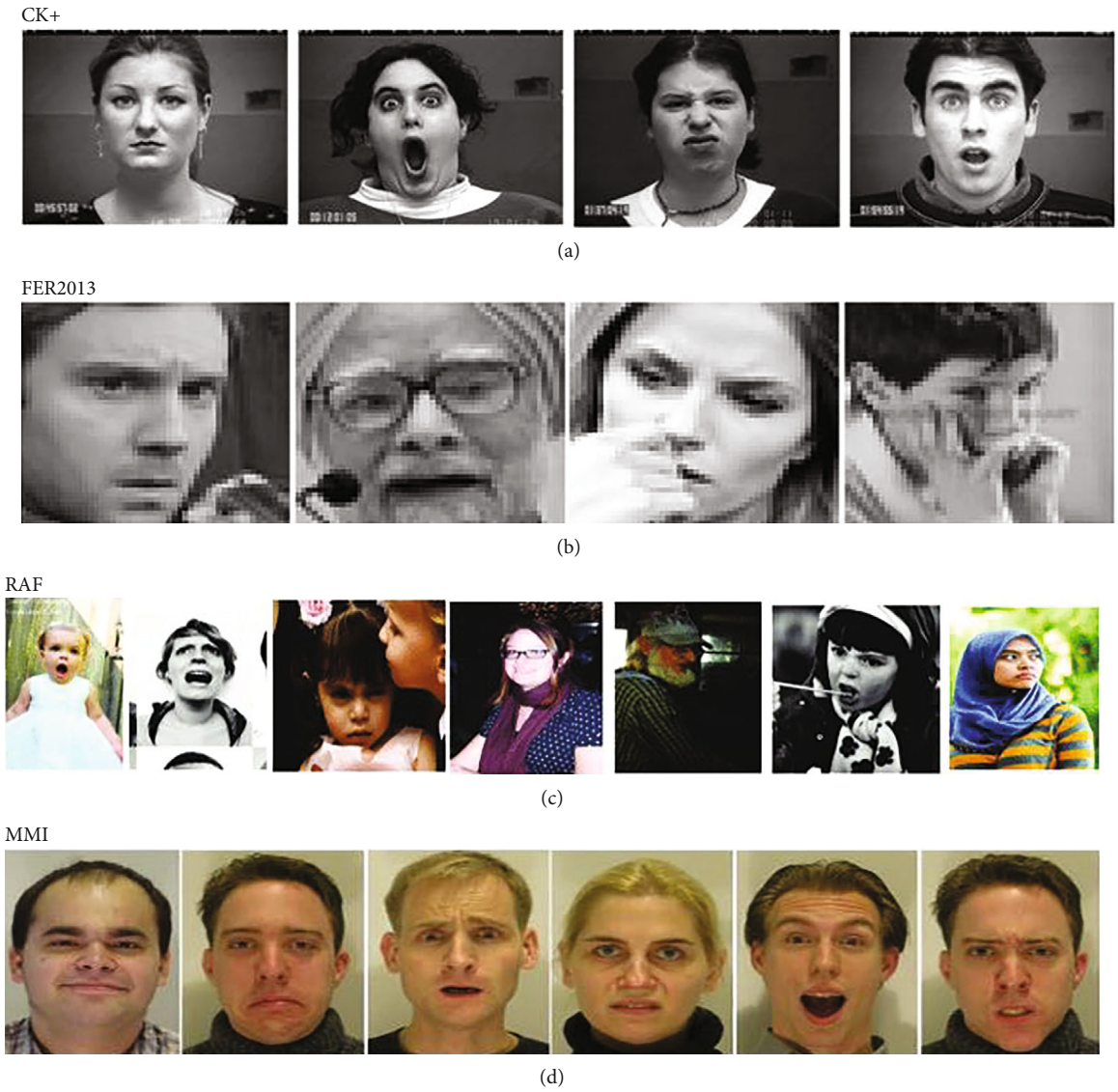


Figure 16: Images with representation of emotion expressions in the datasets: (a) CK+, (b) FER2013, (c) RAF, and (d) MMI.

Table 1: Results with the accuracy metric (%) obtained for the various bases: CK+, FER2013, RAF-DB, and MMI.

| Proposed model | Datasets | | | |
|---|---|---|---|---|
| | CK+ | FER2013 | RAF-DB | MMI |
| CNN (on raw pixels) | 83.30 | 71.30 | 82.60 | 74.50 |
| CNN + face keypoints | 91.20 | 73.10 | 85.65 | 77.20 |
| CNN + face keypoints + HOG | **98.80** | 78.80 | 88.47 | 79.87 |

Table 2: Evaluation of the model with different CNN architectures: proposed model, MobileNet, and inception v3 on Ck+ base.

| Model | Accuracy | Recall | Specificity |
|---|---|---|---|
| Proposed model | 98.80 | 98.70 | 99.75 |
| MobileNet | 81.82 | 77.81 | 96.39 |
| Inception v3 | 91.30 | 90.81 | 98.38 |

Table 3: The model with different CNN architecturecision, recall, and $F1$-score metrics for each of the basic emotions with the model evaluating the images from the CK+ image dataset.

| Emotion | Accuracy | Precision | Recall | $F1$-score |
|---|---|---|---|---|
| Happy | 100.0 | 100.0 | 99.0 | 1.00 |
| Sad | 98.0 | 98.0 | 98.0 | 0.98 |
| Anger | 100.0 | 99.0 | 100.0 | 1.00 |
| Disgust | 98.0 | 99.0 | 97.0 | 0.98 |
| Surprise | 99.0 | 99.0 | 100.0 | 0.99 |
| Fear | 98.0 | 97.0 | 100.0 | 0.98 |

Several methods have been developed to study for DRE of image from the CK+ dataset. But none of them has used our proposed association for DRE. An illustrative overview is now propitious to show the good quality of our method. Table 4 lists several values of accuracy in the literature, including those presented in this work.

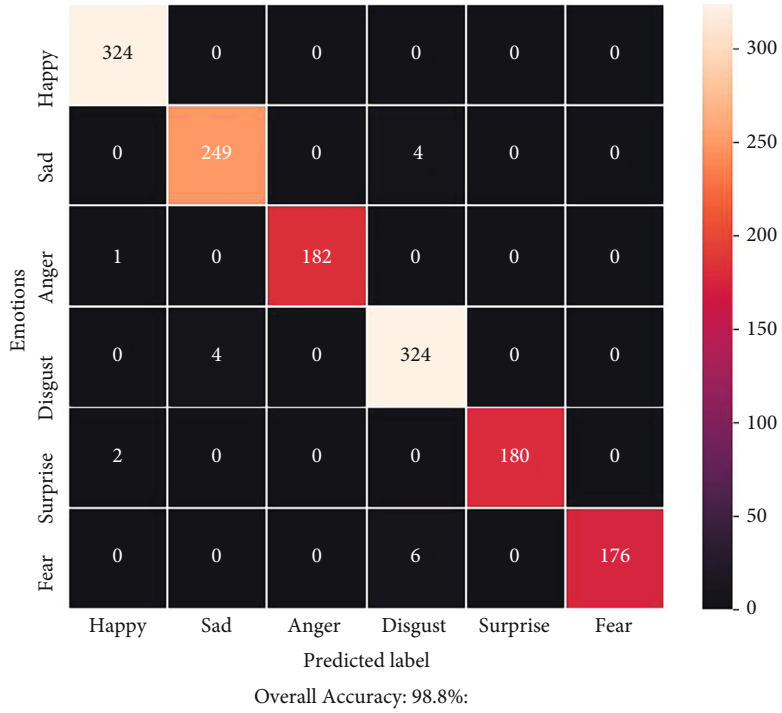


Figure 17: Confusion matrix of the proposed model in sample obtained of the CK+ dataset.

Table 4: Evaluation of the methods used for DRE.

| Author | Feature extraction | Classifier | Accuracy (%) |
|---|---|---|---|
| Vinicius et al. [68] | Learning features | Softmax | 88.58 |
| Wu and Lin [67] | AlexNet features | SVM | 86.83 |
| Wu and Lin [67] | Adaptive feature mapping | Softmax | 89.84 |
| Bilkhu et al. [69] | Cascade regression | SVM | 89.00 |
| Noor et al. [70] | Histogram of oriented gradients | SVM | 90.79 |
| Jain et al. [66] | Learning features | Softmax | 93.24 |
| Cossetin et al. [71] | LBP and Weber local descriptor | Pairwise classifier | 98.91 |
| Salman et al. [72] | Geometric features | MPL | 99.00 |
| Hernandez et al. [73] | Gabor functions | SVM | 99.00 |
| Proposed method | **Learning and handcraft features** | **Softmax** | **98.84** |

### 6.2. Evaluating Serious Games with People with ASD

*6.2.1. Participants.* A group of eight individuals diagnosed with ASD was selected to evaluate the proposed SG. The individuals diagnosed with ASD were selected at the university hospital of the Federal University of UberlÃ¢ndia (UFU), located in UberlÃ¢ndia, Minas Gerais, Brazil. Data collection was carried out after approval by the Research Ethics Committee of UFU (Opinion No. 82555417 0 0000 5152), and during data collection, the subjects were asked to sign the informed consent form, and this work was following Resolution 196/96.

To select the group was considered the following information: (1) being diagnosed with ASD, (2) having no motor limitations, and (3) having undergone clinical evaluation. This clinical evaluation adopted the parameters of the Diagnostic and Statistical Manual of Mental Disorders [74]. The participants were literate and without cognitive delays, which did not compromise the application of the game during the experiments. Although there was no severe impairment in cognitive skills, all participants had social deficits and difficulty recognizing emotions as reported by specialists.

A visit was made to the clinic, and the tool was presented to the specialists. Then, the specialists were trained to use the tool and doubts about its usability were answered. A meeting with the parents presented the tool, and the doubts were clarified. To assess participants' cognitive, social, and communication skills, interviews were conducted with the parents and the specialists. The selected participants with ASD were between 6 and 12 years old in which four boys and four girls. The SG was evaluated with children with different intensity levels (from level 1–level 3) of ASD. During the

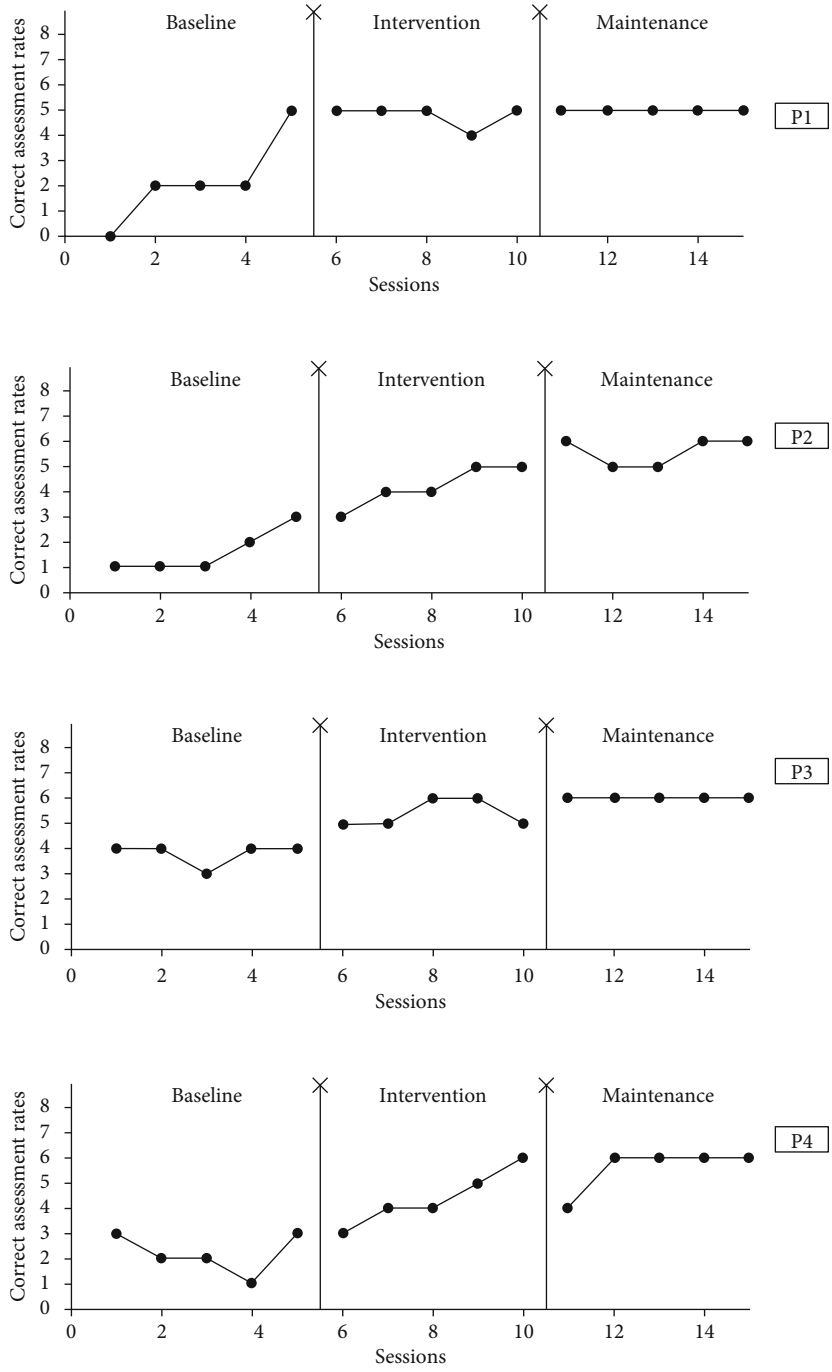


Figure 18: Correct assessment rates of the participants during the three phases of the participants with ASD: P1, P2, P3, and P4.

application of multiple baseline designs (baseline, intervention, and maintenance phases), a psychologist accompanied the participants in order to support the use of the tool. For this group, the application of the tool occurred on an individualized basis with psychologists in therapy sessions. For the application of the SG, a desktop computer with a webcam was used.

*6.2.2. Analysis of the Effectiveness of the SG.* The second stage of investigation is aimed at evaluating the performance of individuals with ASD using the SG. The group of participants was analyzed in the three phases of investigation. This stage can contribute to show whether the intervention was effective and how each individual with ASD had improved. Figures 18 and 19 show the results for the eight participants with ASD. Table 5 shows the number of times that the SG alerted the user due to loss of focus in each of the evaluation steps (eye-tracking).

In the baseline phase, the specialists observed that participants with ASD had their focus not directed to the main part of the tool (see Table 5). Usually, their focus was directed to other parts of the SG interface. The specialists

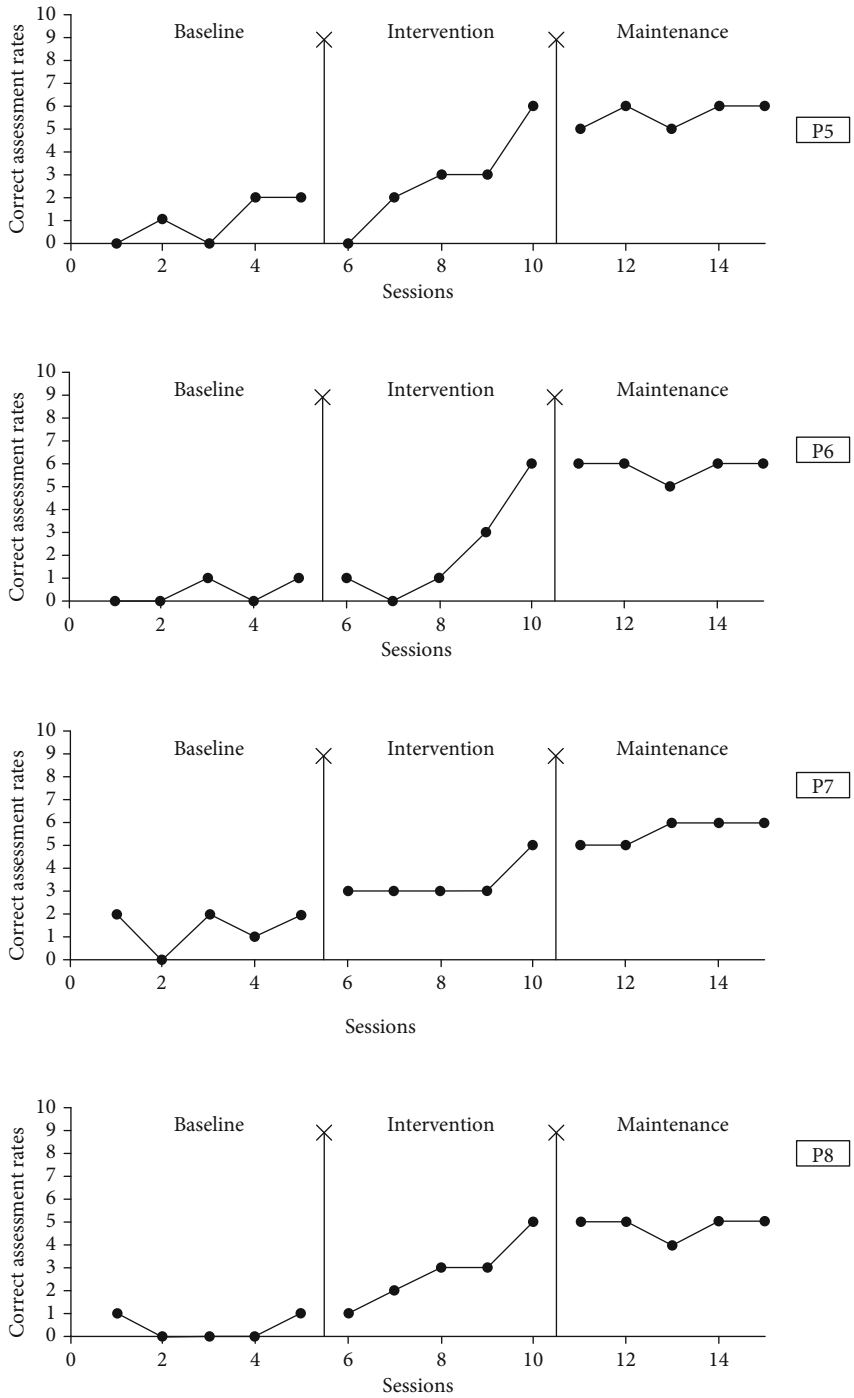


Figure 19: Correct assessment rates of the participants during the three phases of 4 participants with ASD (P5, P6, P7, and P8).

reported that some participants were able to correctly speak the word that represented the emotion but were not able to express the facial muscles to represent it. These sessions contributed to improving the representations and also to usability training of the tool. In the intervention phase (see Figures 18 and 19), the specialists guided participants to focus on the animations while keeping attention on the facial expressions for representing emotions. Participants showed an improvement in emotion identification scores when compared to the baseline stage. It is possible to observe from the results that some participants had an improvement in the hit rates at each session (P1–P8). This information is mainly observed in the intervention phase. In the maintenance phase, all participants managed to increase their performance scores over previous phases. Table 5 shows that all participants had higher indices of loss of focus in the baseline phase and that these values decreased during the sessions. It can also be noted that P6 presented higher rates of inattention, which may be related to the low performance rate in the initial phase (baseline).

Table 5: Occurrence of the alert number that the application generated to the user due the lack of attention.

| Users | Baseline | Attention alert Intervention | Maintenance |
|---|---|---|---|
| P1 | 3 | 2 | 1 |
| P2 | 2 | 1 | 0 |
| P3 | 1 | 1 | 0 |
| P4 | 2 | 0 | 0 |
| P5 | 5 | 3 | 0 |
| **P6** | **7** | **5** | **1** |
| P7 | 2 | 1 | 1 |
| P8 | 4 | 1 | 0 |

Table 6: Summarized results (%) for the individuals with ASD.

| Users | Baseline | Correct rate Intervention | Maintenance |
|---|---|---|---|
| P1 | 63.33 | 80.00 | 83.33 |
| P2 | 30.00 | 70.00 | 93.33 |
| P3 | 63.33 | 90.00 | 100.0 |
| P4 | 60.00 | 73.33 | 93.33 |
| P5 | 20.00 | 36.66 | 93.33 |
| **P6** | **6.66** | **36.66** | **96.66** |
| P7 | 23.33 | 46.66 | 93.33 |
| P8 | 6.66 | 46.66 | 80.00 |

Table 6 presents the correct rates for each stage of the multiple baseline design. The participants presented an advance in the results (score) for the identification of emotions compared to the baseline stage and had an advance in relation to the scores obtained at the baseline stage. The most significant increase was seen for individual P6.

## 7. Conclusions

The computational techniques employed in the SG allowed us to build a tool aimed at individuals with ASD to improve emotion detection and recognition skills without the use of the popular game engines and hardware. The concepts employed in the SG allowed the development of an interactive and dynamic application capable of awakening the interest of the users involved in the interventions in an interactive and relaxed manner during game's evaluation phases. For the person with ASD and with problems related to emotional skills (recognition and expression of emotions), this tool is an aid system that can contribute to the specialists in the area of treatment in a way that can be employed in combination with other techniques.

This SG has a module that employs emotion recognition and detection algorithms for individuals with ASD. This characteristic helps participants to develop or improve their emotional skills: joy, sadness, anger, fear, grief, and surprise. The Viola and Jones technique, the dlib library, and the features learned by the convolutional layers of CNN contributed to the detection of the features captured on the face. The softmax layer enabled the classification of the features for detection of emotion. The dashboard interface enabled the specialist to analyze the data to assist in making decisions to improve the strategies of the emotional skills of the individuals with ASD during the sessions.

The experiments showed that in the baseline stage, the participants with ASD were not yet familiar with the tool and did not make significant progress during the sessions. Participants were not focused on the game scenes as occurs in everyday situations as reported by parents and specialists. In the intervention and maintenance phases, the participants were motivated by the technological resources of the tool. This allowed for greater concentration on the SG and improved the emotional skills.

This study provided important results for the detection and recognition of emotions in public domain databases. The results showed that the system can contribute to the treatment of individuals with ASD to improve emotional skills. One of the limitations presented in this work is the need for a hardware device with a webcam to capture images and an internet connection for evaluation and storage of the data obtained during sessions with individuals with ASD. It is also noted that evaluations of microexpressions and investigation of balanced datasets may contribute to the model. In further work, we intend to investigate the data collected in the baseline and intervention phases and employ CNN architectures to provide recommendations to participants in real-time, enabling feedback capable of helping in the improvement of emotional skills.

## Data Availability

Application results and face data used to support the conclusions of this study are available upon request to the author for correspondence. Participant data may not be released due to the research ethics committee.

## Conflicts of Interest

The authors declare that they have no conflict of interest.

## Acknowledgments

The authors thank everyone who took part in this survey, autism or nonautism people, their parents who allowed their participation, the multidisciplinary team from the schools where these children study, and all the specialists who gave their opinions about this work. This study was financed in part by the Coordination of Improvement of Higher-Level Personnel, CAPES—finance code 88882.429122/2019-01. The authors gratefully acknowledge the financial support of National Council for Scientific and Technological Development (CNPq) (grant # 304848/2018-2).

## References

[1] W. C. D. Souza, M. A. G. Feitosa, S. Eifuku, R. Tamura, and T. Ono, "Face perception in its neurobiological and social context," *Psychology & Neuroscience*, vol. 1, no. 1, pp. 15–20, 2008.

[2] C. C. de Santana, W. C. D. Souza, and M. A. G. Feitosa, "Recognition of facial emotional expressions and its correlation with cognitive abilities in children with down syndrome," *Psychology & Neuroscience*, vol. 7, no. 2, pp. 73–81, 2014.

[3] N. J. Sasson, A. E. Pinkham, K. L. Carpenter, and A. Belger, "The benefit of directly comparing autism and schizophrenia for revealing mechanisms of social cognitive impairment," *Journal of Neurodevelopmental Disorders*, vol. 3, no. 2, pp. 87–100, 2011.

[4] A. M. O. D. Lima, M. R. D. A. Medeiros, P. D. P. Costa, and C. A. S. Azoni, "Analysis of softwares for emotion recognition in children and teenagers with autism spectrum disorder," *Revista CEFAC*, vol. 21, no. 1, 2019.

[5] E. J. Jones, T. Gliga, R. Bedford, T. Charman, and M. H. Johnson, "Developmental pathways to autism: a review of prospective studies of infants at risk," *Neuroscience & Biobehavioral Reviews*, vol. 39, pp. 1–33, 2014.

[6] R. C. Pennington and M. Carpenter, "Teaching written expression to students with autism spectrum disorder and complex communication needs," *Topics in Language Disorders*, vol. 39, no. 2, pp. 191–207, 2019.

[7] K. Sagayaraj, C. R. Gopal, and S. Karthikeyan, "The efficacy of technology and non-technology based intervention for children with autism spectrum disorder: a meta-analysis," *International journal of Innovative Science and Research Technology*, vol. 5, pp. 863–868, 2020.

[8] O. Manta, T. Androutsou, A. Anastasiou, Y. Koumpouros, G. Matsopoulos, and D. Koutsouris, "A three-module proposed solution to improve cognitive and social skills of students with attention deficit disorder (ADD) and high functioning autism (HFA) innovative technological advancements for students with neurodevelopmental disorders," in *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 1–7, 2020.

[9] S. Strickroth, D. Zoerner, T. Moebert, A. Morgiel, and U. Lucke, "Game-based promotion of motivation and attention for socio-emotional training in Autism," *i-com*, vol. 19, no. 1, pp. 17–30, 2020.

[10] J. W. Tan and N. Zary, "Diagnostic markers of user experience, play, and learning for digital serious games: a conceptual framework study," *JMIR serious games*, vol. 7, no. 3, article e14620, 2019.

[11] A. Solinska-Nowak, P. Magnuszewski, M. Curl et al., "An overview of serious games for disaster risk management - prospects and limitations for informing actions to arrest increasing risk," *International Journal of Disaster Risk Reduction*, vol. 31, pp. 1013–1029, 2018.

[12] C. Bosa, "Autismo: atuais interpretações para antigas observações," *Autismo e educação: reflexões e propostas de intervenção*, vol. 1, pp. 21–39, 2002.

[13] OMS, *Autism spectrum disorders*, Key Facts, 2017.

[14] M. J. Maenner, K. A. Shaw, J. Baio et al., "Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016," *MMWR Surveillance Summaries*, vol. 69, no. 4, pp. 1–12, 2020.

[15] C. D., *Control, prevention, autism spectrum disorder (asd)@-misc*, Centers for Disease Control and Prevention, 2018.

[16] R. H. Zaja and J. Rojahn, "Facial emotion recognition in intellectual disabilities," *Current Opinion in Psychiatry*, vol. 21, no. 5, pp. 441–444, 2008.

[17] M. Elshahawy, K. Aboelnaga, and N. Sharaf, "Codaroutine: a serious game for introducing sequential programming concepts to children with autism," in *2020 IEEE Global Engineering Education Conference (EDUCON)*, pp. 1862–1867, IEEE, 2020.

[18] K. Ribu, *Teaching computer science to students with Asperger's syndrome*, Tapir Akademisk Forlag, 2010.

[19] C. A. Clark, *Serious Games*, Viking, New York, 1970.

[20] S. Fridenson-Hayo, S. Berggren, A. Lassalle et al., "'Emotiplay': a serious game for learning about emotions in children with autism: results of a cross-cultural evaluation," *European Child & Adolescent Psychiatry*, vol. 26, no. 8, pp. 979–992, 2017.

[21] A. Dapogny, C. Grossard, S. Hun et al., "Jemime: a serious game to teach children with ASD how to adequately produce facial expressions," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 723–730, IEEE, 2018.

[22] C. Grossard, L. Chaby, S. Hun et al., "Children facial expression production: influence of age, gender, emotion subtype, elicitation condition and culture," *Frontiers in Psychology*, vol. 9, p. 446, 2018.

[23] E. Salahat and M. Qasaimeh, "Recent advances in features extraction and description algorithms: a comprehensive survey," in *2017 IEEE international conference on industrial technology (ICIT)*, pp. 1059–1063, IEEE, 2017.

[24] D. G. R. Kola and S. K. Samayamantula, "A novel approach for facial expression recognition using local binary pattern with adaptive window," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2243–2262, 2021.

[25] G. Pons and D. Masip, "Supervised committee of convolutional neural networks in automated facial expression analysis," *IEEE Transactions on Affective Computing*, vol. 9, pp. 343–350, 2018.

[26] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: a database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, pp. 18–31, 2019.

[27] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[28] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 11, 2020.

[29] X. Sun, P. Wu, and S. C. Hoi, "Face detection using deep learning: an improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, 2018.

[30] C. Fabian Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Emotionet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5562–5570, 2016.

[31] B. Li, S. Mehta, D. Aneja et al., "A facial affect analysis system for autism spectrum disorder," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4549–4553, IEEE, 2019.

[32] K. Valencia, C. Rusu, D. Quiñones, and E. Jamet, "The impact of technology on people with autism spectrum disorder: a systematic literature review," *Sensors*, vol. 19, no. 20, p. 4485, 2019.

[33] R. Pavez, J. Díaz, and D. Vega, "Emotion recognition in children with ASD using technologies: a systematic mapping

study," in *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, pp. 1–8, IEEE, 2019.

[34] C.-H. Chen, I.-J. Lee, and L.-Y. Lin, "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders," *Research in Developmental Disabilities*, vol. 36, pp. 396–403, 2015.

[35] C.-H. Chen, I.-J. Lee, and L.-Y. Lin, "Augmented reality-based video-modeling storybook of nonverbal facial cues for children with autism spectrum disorder to improve their perceptions and judgments of facial expressions and emotions," *Computers in Human Behavior*, vol. 55, pp. 477–485, 2016.

[36] C. Tsangouri, W. Li, Z. Zhu, F. Abtahi, and T. Ro, "An interactive facial expression training platform for individuals with autism spectrum disorder," in *MIT Undergraduate Research Technology Conference (URTC)*, pp. 1–3, IEEE, 2016.

[37] R. Pavez, J. Diaz, J. Arango-Lopez, D. Ahumada, C. Mendez-Sandoval, and F. Moreira, "Emo-mirror: a proposal to support emotion recognition in children with autism spectrum disorders," *Neural Computing and Applications*, vol. 33, pp. 1–12, 2021.

[38] Mozilla, *Html5*, MDN contributors, 2018.

[39] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, IEEE, p. 5, 2005.

[40] A. Yusoff, *A Conceptual Framework for Serious Games and Its Validation*, Ph.D. thesis, University of Southampton, 2010.

[41] A. Sánchez, J. V. Ruiz, A. B. Moreno, A. S. Montemayor, J. Hernández, and J. J. Pantrigo, "Differential optical flow applied to automatic facial expression recognition," *Neurocomputing*, vol. 74, no. 8, pp. 1272–1282, 2011.

[42] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[43] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[44] A. C. N. da Rocha Gracioso, C. C. B. Suárez, C. Bachini, and F. J. R. Fernández, "Emotion recognition system using open web platform," in *Security Technology (ICCST), 2013 47th International Carnahan Conference on*, pp. 1–5, IEEE, 2013.

[45] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[46] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proceedings. international conference on image processing*, vol. 1IEEE.

[47] A. Azcarate, F. Hageloh, K. V. Sande, and R. Valenti, *Automatic facial emotion recognition*, Universiteit van Amsterdam, 2005.

[48] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1867–1874, 2014.

[49] M. Munasinghe, "Facial expression recognition using facial landmarks and random forest classifier," in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, pp. 423–427, IEEE, 2018.

[50] T. Caramihale, D. Popescu, and L. Ichim, "Emotion classification using a tensorflow generative adversarial network implementation," *Symmetry*, vol. 10, no. 9, p. 414, 2018.

[51] D. Fleet and Y. Weiss, "Optical flow estimation," in *Handbook of Mathematical Models in Computer Vision*, pp. 237–257, Springer, 2006.

[52] R. C. Gonzalez and R. E. Woods, *Image processing*, vol. 2, Pearson, 2007.

[53] A. G. Howard, M. Zhu, B. Chen et al., "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017, http://arxiv.org/abs/1704.04861.

[54] R. Sadik, S. Anwar, and M. L. Reza, "Autismnet: recognition of autism spectrum disorder from facial expressions using MobileNet architecture," *International Journal*, vol. 10, no. 1, pp. 327–334, 2021.

[55] S. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, pp. 112–119, 2020.

[56] M. F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 1, pp. 28–43, 2012.

[57] M. F. Valstar, E. Sánchez-Lozano, J. F. Cohn et al., "Fera 2017-addressing head pose in the third facial expression recognition and analysis challenge," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 839–847, IEEE, 2017.

[58] D. Acevedo, P. Negri, M. E. Buemi, F. G. Fernández, and M. Mejail, "A simple geometric-based descriptor for facial expression recognition," in *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pp. 802–808, IEEE, 2017.

[59] A. L. Apple, F. Billingsley, I. S. Schwartz, and E. G. Carr, "Effects of video modeling alone and with self-management on compliment-giving behaviors of children with high-functioning ASD," *Journal of Positive Behavior Interventions*, vol. 7, no. 1, pp. 33–46, 2005.

[60] B. Ingersoll, A. Dvortcsak, C. Whalen, and D. Sikora, "The effects of a developmental, social—pragmatic language intervention on rate of expressive language production in young children with autistic spectrum disorders," *Focus on Autism and Other Developmental Disabilities*, vol. 20, no. 4, pp. 213–222, 2005.

[61] F. Castelli, "Understanding emotions from standardized facial expressions in autism and normal development," *Autism*, vol. 9, no. 4, pp. 428–449, 2005.

[62] T. Kanade, J. Cohn, and Y.-L. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG '00)*, pp. 46–53, 2000.

[63] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," in *International conference on neural information processing*, pp. 117–124, Springer, 2013.

[64] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.

[65] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database,"

in *Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect*, p. 65, Paris, France, 2010.

[66] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, vol. 120, pp. 69–74, 2019.

[67] B.-F. Wu and C.-H. Lin, "Adaptive feature mapping for customizing deep learning based facial expression recognition model," *IEEE access*, vol. 6, pp. 12451–12461, 2018.

[68] M. V. Zavarez, R. F. Berriel, and T. Oliveira-Santos, "Cross-database facial expression recognition based on fine-tuned deep convolutional network," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 405–412, IEEE, 2017.

[69] M. S. Bilkhu, S. Gupta, and V. K. Srivastava, "Emotion classification from facial expressions using cascaded regression trees and SVM," in *Computational Intelligence: Theories, Applications and Future Directions-Volume II*, pp. 585–594, Springer, 2019.

[70] J. Noor, M. Daud, R. Rashid, H. Mir, S. Nazir, and S. A. Velastin, "Facial expression recognition using hand-crafted features and supervised feature encoding," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1–5, IEEE, 2020.

[71] M. J. Cossetin, J. C. Nievola, and A. L. Koerich, "Facial expression recognition using a pairwise feature selection and classification approach," in *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 5149–5155, IEEE, 2016.

[72] F. Z. Salmam, A. Madani, and M. Kissi, "Emotion recognition from facial expression based on fiducial points detection and using neural network," *International Journal of Electrical and Computer Engineering*, vol. 8, p. 52, 2018.

[73] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "A facial expression recognition with automatic segmentation of face regions," in *International Conference on Intelligent Software Methodologies, Tools, and Techniques*, pp. 529–540, Springer, 2015.

[74] A. Association and A. P. A. Staff, *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR®)*, American Psychiatric Association Publishing, 2010.