



# **Improving Learners' Assessment and Evaluation in Crisis Management Serious Games: An Emotion-based Educational Data Mining Approach**

Ibtissem Daoudi, Raoudha Chebil, Erwan Tranvouez, Wided Lejouad Chaari,  
Bernard Espinasse

## **► To cite this version:**

Ibtissem Daoudi, Raoudha Chebil, Erwan Tranvouez, Wided Lejouad Chaari, Bernard Espinasse. Improving Learners' Assessment and Evaluation in Crisis Management Serious Games: An Emotion-based Educational Data Mining Approach. Entertainment Computing, 2021, 38, pp.100428. <10.1016/j.entcom.2021.100428>. <hal-03203938>

**HAL Id: hal-03203938**

**<https://hal.science/hal-03203938v1>**

Submitted on 9 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

# **Improving Learners' Assessment and Evaluation in Crisis Management Serious Games: an Emotion-based Educational Data Mining Approach**

Ibtissem Daoudi<sup>\*, 1, 2</sup>, Raoudha Chebil<sup>1</sup>, Erwan Tranvouez<sup>2</sup>,  
Wided Lejouad Chaari<sup>1</sup> and Bernard Espinasse<sup>2</sup>

1-National School of Computer Science (ENSI), University of Manouba, Tunisia

2- Computer Science and Systems Laboratory (LIS) UMR CNRS 7020, Aix Marseille  
University & Toulon University, France

\*[ibtissem.daoudi@ensi-uma.tn](mailto:ibtissem.daoudi@ensi-uma.tn)

## 1. Introduction

Crises and risks such as natural disasters and industrial accidents represent a real and major threat for the life and well-being of people causing injuries, diseases and deaths [1]. In this context, *Crisis Management* (CM) aims to reduce injuries and to increase survival under stressful conditions [2]. *Crisis Management* is the application of strategies designed to help stakeholders with various skills or roles, either from different organizations (medical unit, police etc.) or the general public, deal with unexpected negative events in order to limit the severity of its consequences [3]. As a way to define and to test those strategies, training (based on "battlefield"-born experience) and simulation are useful but also highly complex, costly, and time-consuming tasks [4]. Indeed, one of the major characteristics of the field exercises, is the variability of unpredictable events stemming from human actions and interactions as well as technology's shortcomings. Such uncertainty can result from scarce or false information, events judged too improbable to really prepare for, or taking into account people with disabilities [5]. In the former case, each *disability* situation may require specific procedures.

As a computer-based solution, *Serious Games* (SG) in CM can be programmed with various scenarios allowing learners to encounter such situations before they really occur [1]. In particular, *collaborative Crisis Management Serious Games* (CMSG) represent a complementary training tool that can improve individuals' quality of life and behavior by focusing on operational aspects of CM processes. CMSGs have proved their potential for teaching learners both technical and soft skills related to solve different types of crisis problems such as natural disasters (earthquakes, floods), man-made disasters (terrorist attacks), and technological crises (cyber attacks) in an engaging way while reducing training cost and saving time [6]. Soft skills, including emotional and social skills like communication, sense of responsibility and stress management, are necessary in order to make a rapid and efficient response to the crisis.

Despite its obvious interest, the exploitation of the CMSG concept in training processes is not always a guarantee of its usability. As with any learning systems, CMSGs rely on the implicit alignment of the *learning outcomes* (knowledge or skill) and the *game experience* (motivation, engagement). In particular, the usability of collaborative CMSGs depends on different learners' characteristics including *cognitive*, *emotional*, and *social* aspects [1]. Consequently, several works propose to evaluate learners' behavior during their interaction with a CMSG. However by studying the state of the art, we have noticed that researchers often neglect a crucial characteristic of CMSGs namely "*emotions*" [1]. In fact, CM with new training technologies involves not only teaching cognitive procedures related to managing crisis situations but also emotional processes which can significantly impact actors' learning behavior. Indeed, positive affective states such as *engagement/Flow* can contribute to increase learning gains, while negative ones like *frustration* and *boredom* can produce misconceptions and difficulties [7].

According to our findings, we focus on the detection and analysis of learners' emotions expressed in *textual* and *visual* data to infer *Flow game-play experience (engagement)* indicator in CMSGs [8]. These two types of modalities are selected because they allow us to detect emotions in an objective and implicit manner without interrupting players from game-play [9]. To the best of our knowledge, players' engagement measurement and its impact on learning using *Artificial Intelligence* (AI) techniques have not been investigated in such context. Hence, our objective is to propose an emotion-based *Educational Data Mining* (EDM) approach to assess learners' affective dynamics during a collaborative game session for CM training, and to evaluate their final mental states (at an *individual* and *collective* level) based on two different interaction modes: facial expressions and text chat data. Such a proposal would allow a more comprehensive evaluation in and of CMSGs.

The remaining parts of this article are organized as follows. In Section 2, we present works and concepts related to our proposition. In Section 3, we present our main contribution consisting in an EDM-based method for evaluating learn-

ers' affective states in CMSGs. In Section 4, we present the game-based evacuation scenario. In section 5, we show the obtained results from the application of our method on this game scenario, and we discuss our major findings. Finally,  
65 in section 6 we share our conclusion and our future works.

## 2. Related Works

This section presents the most important works on evaluation of CMSGs, and gives an outline of the key concepts used in this article.

### 2.1. Literature Review

70 As the number of CMSGs increases, varying in terms of crisis situation, number of players or learners' characterization, numerous researchers have worked on evaluation of and in game-based learning environments [1]. Many of these works focus on a *summative evaluation*, conducted at the end of the learning process to validate the overall learners' achievements [10] [11] [12], whereas oth-  
75 ers consider a *formative evaluation* (often known as assessment [1]) realized during the learning process by observing learners' behaviors [4].

For instance, the work referenced by [10] and [12] presented respectively an evaluation of the "Dread-Ed" and "Stop Disasters" games using *questionnaires*. More precisely, "Dread-Ed" training objectives involve soft skills in a crisis sit-  
80 uation such as communication and group decision making [10]. Evaluation was then performed via a *questionnaire using a 5-point Likert scale* to conclude whether the game contributes to the learning of social skills, completed with a *debriefing session* at the end of the training process. Dealing with a more general learning outcome, "Stop Disasters" embrace building a safety culture for  
85 emergencies [12]. To verify if the game really improves the awareness of risky situations, the participants answered *questionnaires*, *before* and *after* playing the game, about three aspects : game-play, missions, and game scenarios.

All these works use explicit evaluation techniques like *questionnaires*, *interviews* and *debriefing sessions* proposed before, during or after the game session.

90 Despite their frequent use, these techniques represent a subjective evaluation  
and rely on non-exhaustive players' opinions. For example, as with all ques-  
tionnaires, some learners may understand or answer incorrectly the questions,  
impacting thus negatively the reliability of the game data and its analysis [1]. In  
addition, using questionnaires during the game session is too intrusive and thus  
95 impacts learners' engagement by disrupting the game dynamics. Players' en-  
gagement and accuracy of evaluation results can be improved with non-intrusive  
AI techniques known as "*stealth assessment*" [13]. For example, the work ref-  
erenced by [4] presented a multi-criteria and distributed "stealth" assessment  
approach of learners in a collaborative CMSG called "*SimFor*". To implement  
100 this approach, they have developed a multi-agents system which produces both  
individual and collective learners' evaluation (social and cognitive states).

Although CM training is an emotionally rich experience (stress, surprise,  
fear), a literature study shows a lack of works focusing on evaluating affective  
105 states like frustration and engagement in CMSGs. Moreover, most of these  
works produce a summative evaluation at the end of the game session, which  
can not be very useful for adapting the game to their current emotional states  
compared to formative evaluation [1]. In this article, our objective is to explore  
the potential of an affective computing approach by producing both formative  
110 and summative evaluation of learners' emotions within a collaborative CMSG.  
The following subsection explores what such objective entails.

## 2.2. Background

Having identified the role of emotions in improving CMSG's usability, this  
subsection presents how *affective computing* provides sound conceptual back-  
115 ground allowing to improve players' *engagement* as defined by the *Flow theory*.

### 2.2.1. Affective Computing and Emotions

One of the earliest mentions of the term "*Affective Computing*" (AC) was  
introduced by *Picard* in 1997. According to *Picard*, AC is "*the ability of a com-*

puting device to recognize, interpret and simulate human affective state” [14].

120 This affective state is a complex phenomenon including emotions, sentiments, personality traits, needs and subjective well-being [15]. However, *emotion* remains the most relevant affective state since it plays a crucial role in our daily lives especially in decision making, learning, and situation awareness [16]. In the literature, there are three main categories of models aiming to represent  
125 emotions namely the *discrete model* [17], the *dimensional model* [18], and the *cognitive appraisal theory* [19].

AC aims to gather information about users’ affective states from a variety of data sources including *verbal sources* like textual data and speech patterns as well as *non-verbal sources* such as behavioral features and physiological features  
130 [20]. Generally, the tools that support emotion detection are based on *sensors devices*, *cameras*, and *microphones*. Then, emotional data are analyzed by applying various AI methods such as machine/deep learning algorithms and data mining techniques.

A series of recent studies have proved that *basic emotions* [17] namely anger, joy, fear, surprise, disgust, and sadness do not play a significant role in computer-  
135 based learning [15] [21]. In e-learning processes, we study a subset of *cognitive-affective states* that are the most relevant to learning. These affective states are a combination of basic emotions and include *Flow/engagement*, *frustration*, *confusion*, and *boredom* [22]. For this reason we present, in what follows, the *Flow theory* [8] as well as the *cognitive disequilibrium model* [23] that investigates the  
140 dynamics of affective states over time during learning activities.

### 2.2.2. Flow Theory and Cognitive Disequilibrium Model

Introduced by Csikszentmihályi, the *Flow Theory* is defined as: “rewarding, subjective, emotional state of optimal pleasure that arises when an individual  
145 is absorbed in either work or leisure activities that are perceived as valuable” [8]. Learners can experience *boredom* when the material does not attract them. This state occurs when their skills level is high and the game challenge is low. Conversely, learners can experience *anxiety* (or *frustration*) when they make

mistakes and get stuck. The state of anxiety occurs when their skills level is  
 150 low and the game challenge is high as shown in Figure 1. *Flow*, also known  
 as *concentrated engagement*, is a mental state in which the involvement and  
 enjoyment in the learning process are so strong that time perception and fa-  
 tigue disappear. This state occurs when both the levels of learners' skills and  
 game challenge are high. However, *confusion* state is a state of uncertainty ex-  
 155 perience when learners have difficulty of comprehending the material and are  
 unsure on how to proceed. The confusion state (or apathy) occurs when both  
 the levels of learners' skills and game challenge are low [24].

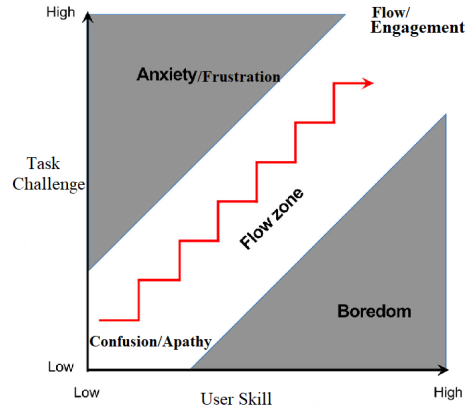


Figure 1: Flow State Diagram (reproduced from [8])

During a computer-based learning session, learners change regularly their  
 affective states. A theoretical model, called the *Cognitive Disequilibrium Model*  
 160 (CDM), addresses transitions between affective states while solving activities  
 in relatively short learning sessions [22]. This model assumes that learners are  
 in a base state of *engagement* until they face errors, uncertainties and others  
 obstacles. *Confusion* is a key signature of the cognitive disequilibrium that  
 occurs when an impasse is detected [23]. Learners revert into the engagement  
 165 state if equilibrium is restored through reflection and problem solving. However,  
 when the source of the confusion is never resolved, the learner gets stuck and  
 important goals are blocked. In this situation, the learner will become *frustrated*,



and then *bored* if the frustration persists. Boredom state represents a critical point at which the learner disengages totally from the learning process [23].

170 Figure 2 illustrates the cognitive disequilibrium model. Transitions (red links) have been confirmed by experimentations described in [23]. However, other authors have proposed that other transitions from frustration to confusion or to engagement can be possible in other scenarios. In addition, the work referenced by [22] observed boredom  $\Rightarrow$  frustration and frustration  $\Rightarrow$  confusion oscillations in a complex learning context (green links).

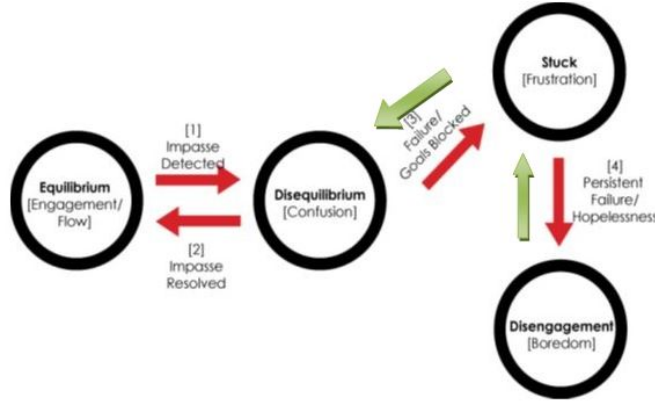


Figure 2: Cognitive Disequilibrium Model [23]

175 The next section presents how the concepts and methods previously described, can be combined to enhance a more comprehensive learners' assessment and evaluation in collaborative CMSGs.

### 3. Contribution : An Emotion-based EDM Method for Learners' Evaluation

180 The aim of the article is to answer our global research question : "*How to evaluate the usability of collaborative CMSGs ?*" To this end, we propose an emotion-based Educational Data Mining (EDM) method for learners' affective

states evaluation. Such investigation in a domain like CM training is a distinctive aspect of our work. This domain seems likely to evoke rich affective experiences that would differ from those typically observed during learning in traditional academic subjects. In this section, we start by presenting a general overview of our method, after that we describe its composing tasks.

### 3.1. Extending Learners' Assessment and Evaluation in Collaborative CMSGs

As discussed in previous sections, the usability of CMSGs can be improved by a more extensive evaluation capacity based on a deep analysis of different learners' characteristics [1]. Extending such capacity does not mean that no SGs fulfill part of the assessment needs, but rather it is possible to complement such games by considering all various criteria needed for a complete evaluation of a particular CMSG. Thus, we consider how each criterion, either qualitative (social interactions, emotions) or quantitative (duration, human and material impacts), and data on which they are based on, can be regrouped and globally analyzed. In this article, we propose a method that can automatically evaluate learners' behavior at two levels : *individual*, which focuses mainly on the emotional aspect, and *collective* dealing with the overall evaluation at a group level using results obtained from the first level. Therefore, the proposed method is composed of three main steps as follows:

1. **Trace recovery:** this step collects players' data produced during a CMSG session. These data cover *visual data* (real-time video recordings from webcams) to capture facial expressions, and *textual data* (exchanged messages during playing) to detect emotions from text.
2. **Basic emotions and stress detection:** the previously collected data are analyzed according to the discrete model [17] to classify learners' basic emotions. Given the sensibility of crisis context to stress factor, we also detect the learners' stress state (which is not considered as a basic emotion) to study its impact on their actions during playing.
3. **Learners' profiles construction:** the detected basic emotions are then analyzed to produce individuals formative and summative evaluations of

215 learners' emotional profiles as well as a collective emotional profile (group emotion) using a decision tree model.

These three steps correspond to specific tasks namely: *data collection and annotation*, *data fusion*, *data analysis*, and *data visualization* as illustrated in Figure 3. These four different tasks are performed using specific AI algorithms presented in the following parts.

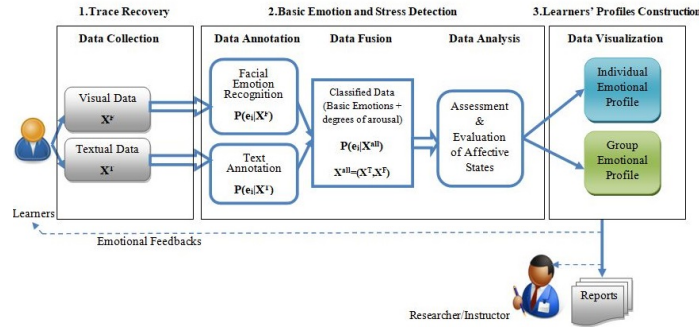


Figure 3: General Overview of the Proposed Method

### 220 3.2. Data Collection and Annotation Task

In order to collect data in a way that is less intrusive compared to physiological measurements [9], we extract *visual data* from facial expressions and *textual data* from exchanged messages during playing. These data sources are annotated as follows:

- 225 • **Text annotation:** the textual content of these messages represents a rich source to detect their emotions that are revealed by deep learning algorithms. In doing so, we utilize an automated predictive analytics tool classified as one of the top AI APIs for emotion detection from raw text strings (shorter instances of text like conversations) with 93.5% of accuracy namely *Indico*<sup>1</sup>. This annotator gives as an output the probability that  
230 the text reflects the Ekman' basic emotions.

<sup>1</sup><https://indico.io/>

- **Video annotation:** the collected visual data are annotated using machine learning and computer vision algorithms to recognize facial expressions. Basing on these methods, we adopt *Openface 2.0*<sup>2</sup>: an automatic facial behavior analysis and understanding toolkit. *Openface* is a powerful toolkit that provides highly accurate facial landmark detection, head pose tracking, eye gaze estimation and *Facial Action Unit recognition* (FAU) [25]. It is possible to save the outputs of the processed data as CSV files containing 18 FAU (codes describing specific facial muscle activations) listed in terms of presence (0 as not present and 1 as present in the face) and intensity (how intense is the FAU on a 5-points scale). Table 1 lists some examples of the facial Action Units (AU) used as features and provides *Intensity information* (I) or *Presence information* (P).

Action Unit Number	Description	Prediction
AU1	Inner brow raiser	I
AU7	Lid tightener	P
AU17	Chin raiser	I
AU45 ...	Blink	P

Table 1: List of some examples of AUs and their Descriptions [25]

In our work, we exploit the output of FAU recognition system since it expresses human emotions according to Ekman [17]. In order to recognize the basic emotions, we need to perform a rules matching between FAU outputs and basic emotions based on the *EMFACS (Emotional Facial Action Coding System)* [26] shown in Table 2. This table represents a common standard for characterizing the six basic emotions from the detected FAU that are well defined and universal.

<sup>2</sup><https://github.com/TadasBaltrusaitis/OpenFace>

Action Units	Basic Emotion
AU6+AU12	Joy
AU1+AU4+AU15	Sadness
AU1+AU2+AU5+AU26 ...	Surprise

Table 2: Examples of Rules Extracted from the EMFACS [26]

### 3.3. Data Fusion Task

In this task, we perform a *bi-modal fusion at the decision-level* which refers to the process of combining data collected from the two modalities after being pre-classified independently to obtain the final classification. Each classified  
255 modality, using the previous annotators, provides one hypothesis on labeled emotion categories; and this integration method gives a global estimate of labeled emotion categories based on partial results [20].

In our case, we consider the face modality as a "*major*" mode from a qualitative point of view but also from a quantitative point of view as, for now,  
260 this interaction mode produces more data than textual mode. So, we rely on the classification provided by the face classifier "*Openface*". For this reason, we adopt the *averaging rule using weights* as a combination rule for the fusion task. Hence, we assign weights as follows:  $\mu^T=0.3$  for the text modality and  $\mu^F=0.7$  for the face modality. We adopt this weighting proposed and validated by works  
265 referenced by [9] and [27]. These works are similar to our study in terms of multimodal emotion detection. In what follows, we describe the adopted approach of decision-level fusion of the individual annotators' outputs:

$X^{all} = (X^T, X^F)$  represents the global feature vector consisting of the text feature vector,  $X^T$  and the face feature vector,  $X^F$ . The two separate classi-  
270 fiers provide the posterior probabilities  $P(ei | X^T)$  and  $P(ei | X^F)$  for text and face, respectively, having to be combined into a single posterior probability  $P(ei | X^{all})$  where  $ei$  represents one of six possible classes of basic emotions ( $e1=joy$ ,  $e2=sadness$ ,  $e3=surprise$ ,  $e4=anger$ ,  $e5=fear$  and  $e6=disgust$ ). Then, we apply the averaging formula using these weights to compute the average probability

275 of the two modalities defined as follows [28]:

$$P(ei | X^{all}=X^T \text{ and } X^F) = (\mu^T * P(ei | X^T) + \mu^F * P(ei | X^F)) / 2$$

### 3.4. Data Analysis Task

In this task, we start by detecting some specific FAU to distinguish stressed from non-stressed participants and to study the impact of stress on affective transitions. Then, we propose a new mapping between basic emotions and  
280 affective states.

- **Stress detection:** stress is one of the most frequently occurring emotions inherent to CM since it affects the actors' way to manage crises [1]. Stress can be defined as: "*the emotional, cognitive, behavioral and physiological reaction regarding to a negative event*" [29]. In general, stress can be  
285 split into the following three categories depending on the time of exposure to stressors: acute stress, episodic acute stress, and chronic stress. The *acute stress* factor is characterized by a short period of time where the body returns to its normal state after the stress factor has passed and it is not considered harmful [29]. It is this category of stress that lies within  
290 the scope of our research. We believe that acute stress has an important impact on learners' engagement and transitions between different affective states. Therefore, we focus in what follows on detecting stress mentalstate. Given stress is related to emotions; facial expressions have been used to detect acute stress by linking some of basic emotions as features. Many  
295 works have proved that, in different contexts like driving and working environments, stress is detected if either anger, fear or a combination of these two negative emotions is detected constantly within a fixed time interval [29] [30]. In particular, they focus on some specific FAU and their activation level extracted in each video frame, described as an indicator  
300 for fear and/or anger including AU2, AU4, AU5, AU7 and AU20.

- **Mapping between affective states and basic emotions:** affective states including engagement, anxiety and boredom are particular combi-

nations of basic emotions as shown in [22]. In our study, we adopt the existing mapping as described in [22] and [9]; and we propose other combinations of basic emotions to deduce learners' affective states based on existing emotions theories [17] [18] [19] and the Flow definition [8].

*Flow or engagement* is defined by a high level of surprise and a low sadness level [9]. Since joy and sadness are opposite emotions as validated by Ekman [17] and Flow is characterized by a full involvement and enjoyment in the activity [8], we can affirm that flow can be defined also by a high level of surprise and a high joy level.

*Anxiety or frustration* is detected at the presence of a high degree of anger and a low degree of joy [9]. Likewise, anxiety can be defined by a high level of anger and a high sadness level. Moreover, basing on the definition of anxiety affective state [8], this state can be mapped to a high level of fear as well as a high level of sadness. In the same way, anxiety can be defined by a high level of fear and a low level of joy.

*Boredom* can be mapped to a high level of disgust as well as a low level of joy. In the same manner, boredom can be defined by a high level of disgust and a high sadness level [9].

In our work, we suppose that the state in which all the levels of six basic emotions are low will represent the *confusion (apathy)* affective state [9]. Table 3 resumes our novel interpretations.

Affective State	Basic Emotions Combination
Engagement/Flow	high level of surprise + high level of joy
Anxiety (Frustration)	high level of anger + high level of sadness or high level of fear + high level of sadness or high level of fear + low level of joy
Boredom	high level of disgust + high level of sadness
Confusion (Apathy)	all the levels of six basic emotions are low

Table 3: Our novel interpretations of basic emotions combinations to deduce affective states

### 3.5. Data Visualization Task

The final task concerns visualization of our analysis results at two levels: *individual and collective*. On the one hand, we visualize the summative individuals emotional profiles by selecting the dominant and the most pronounced emotion [15]. On the other hand, we visualize the aggregation of all individuals emotional profiles based on a decision tree algorithm to decide on the polarity of group emotion (positive or negative) and then to constitute the collective emotional profile [31]. Tree-based methods are easy to explain and to understand since its graphical representation is very intuitive for decision making [32]. In addition, decision trees perform well when they are applied on the educational data for classifying and predicting the learners' behavior [33].

So, we apply *J48 decision tree classifier* [32] to generate a decision on the group emotion based on individual emotions with a *default confidence value = 0.25*. The principle is to decide the class label of group emotion (positive or negative) by learning decision rules inferred from prior training data. For example, the following instance (Engagement=40%, Apathy=50%, Boredom=5%, Anxiety=5%) would be sorted down the leftmost branch of this decision tree and would therefore be classified as a positive instance.

Figure 4 shows the decision tree model of group emotion.

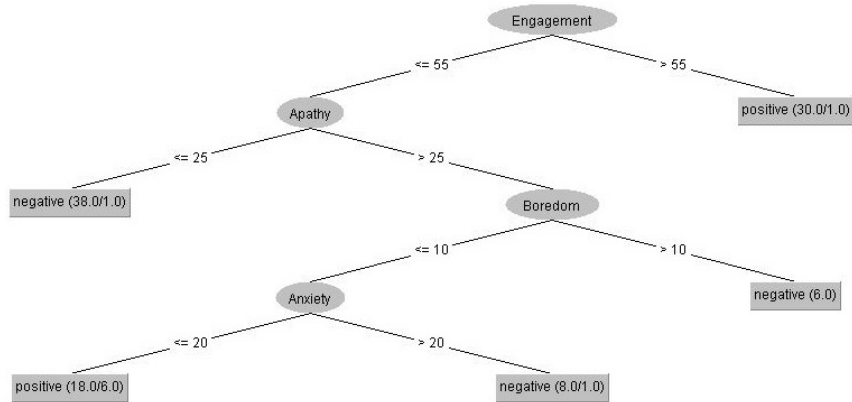


Figure 4: Decision Tree Model of Group Emotion



345       Basing on the following accuracy formula and according to our experimental  
results, this tree-based method reaches an *accuracy of 81% using 5-fold cross-*  
350       *validation.*

$$Accuracy = (correctly\ predicted\ class / total\ testing\ class)$$

The *group emotion* helps instructors to understand the *group performance*.  
Indeed, when the group emotion was positive, it was positively related to the  
350       team’s performance and vice-versa [31]. Moreover, the group emotion helps re-  
searchers to have a better understanding how learners interact with educational  
environments and to make decisions to improve their effectiveness [31].  
In the following section, we report the development of a collaborative evacuation  
scenario based on an existing simulation game platform.

#### 355   4. Experiment: Development of a Game-based Evacuation Scenario

In order to test and to improve the proposed method, we consider an *evacu-*  
*ation scenario*. This scenario is developed in a SG implemented with the com-  
mercialized *iScen* platform. First of all, we introduce this platform, then we  
describe the developed scenario. Finally, we present our experimental protocol.

##### 360   4.1. *iScen* Platform

*Iscen*<sup>3</sup> is a software platform aiming to create interactive scenarios and use  
them in multi-players game sessions on a local network or on the Internet.  
Developed by *Eversim*<sup>4</sup> society, this platform is specifically intended for crisis  
simulation, management, and training. It relies on the *Client-Server* archi-  
365       tecture. The server is used by the *moderator* who creates and administrates  
the scenario and the game session. All the learners (participants) connect to  
the server through their client machines to join the session. *Iscen* has several  
communication systems consisting of *voice chatting, messaging, and a system of*

---

<sup>3</sup><http://www.i-scen.com/home.php?langue=en>

<sup>4</sup><http://www.eversim.com/ns/en/home.php>

*shared notes*. The messaging system let us incorporate different types of content  
370 like text, images, videos, and questionnaires. IScen requires defining *scenarios*  
which the game session instantiates. These scenarios are *script-based* via a  
graphical interface. A script is a combination of several *steps* linked together  
by *paths*, and manipulates a finite set of *game objects*. However, the current  
version of iScen does not allow more than six human players to participate to a  
375 game session simultaneously.

#### 4.2. Game Scenario and Experimental Protocol

In order to test our method, we have built a proof of concept evacuation  
scenario. In what follows, we first describe the scenario, then we detail the  
experimental protocol based on this scenario. Finally, we present the post-  
380 evaluation questionnaire used to prove the obtained results.

##### 4.2.1. Game Scenario Description

We have implemented a collaborative game prototype for building evacuation  
training in case of a fire emergency situation as shown in figure 5.

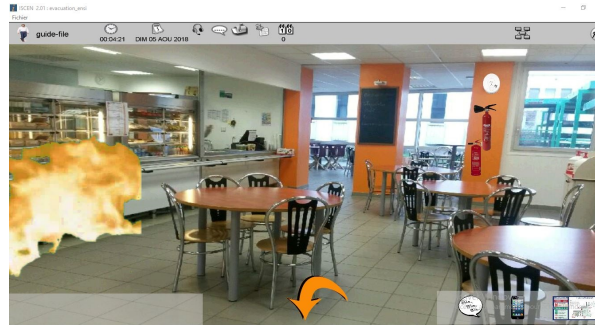


Figure 5: Screen Capture of Crisis Situation Simulated in iScen Platform

Evacuation is defined as the movement of individuals to a safe location. The  
385 scenario aims to train people (professors, students, administrative staff) of a  
public university on evacuating all the present people, especially people with  
special needs, during a fire emergency before the situation becomes dangerous.  
In fact, such scenario aims to respond to a real training need that does not exist

in the cited university. Moreover, the developed evacuation scenario constitutes  
390 a low-cost solution for building evacuation training in a difficult economic context compared to real exercises.

In order to make safe and orderly the evacuation process, all the learners need to perform the appropriate actions that can be taken using the keyboard  
395 and mouse. The evacuation exercise involves a group of participants (including player characters and non-player characters) having different roles namely *coordinator*, *security responsible*, *warden* and *deputy* who must collaborate and coordinate their actions to manage an emergency evacuation procedure.

- *The coordinator*: he/she decides to evacuate or to abandon the floors at  
400 any time, and assures the overall coordination of actions related with emergency evacuation procedure. The coordinator is responsible for ensuring that all stakeholders are aware of their responsibilities and missions, passing on the result of the evacuation to the fire service and other emergency services, as well as reporting back to evacuees.
- *The security responsible*: his/her role consists in assisting warden and  
405 deputy in the emergency procedure by checking that all occupants have proceeded to the designated assembly area. The security responsible should also help people who require special assistance or micro-missions during evacuation in particular vulnerable people like people with mental  
410 and/or physical disabilities as well as panicked and injured people. These micro-missions are for example: providing adequate medical care and special equipment or calling an ambulance if needed, reassuring the crowd, assisting the disabled person to a predetermined safe area, bringing the wheelchair if possible etc. Such micro-missions will stimulate the sense  
415 of responsibility among participants and will help them in dealing with real-life situations.
- *The wardens and deputy wardens*: they are in charge of supervising the evacuation of the building when the fire alarm system sounds, especially

420 guiding rapidly people to a safe area (generally known as assembly point)  
and searching for missing people using some evacuation drills (extinguish-  
ers, fire hose reels etc).

#### 4.2.2. *Experimental Protocol Description*

The experimentation is carried out in the context of a training session for  
university building evacuation where learners are a group of 30 computer science  
425 engineering students. In order to characterize more precisely the task of protocol  
construction in our context, we conducted an approach that consists of the  
following experimental actions :

1. **Information collection on participants:** it is a qualitative pre-evaluation  
via an interview session. It contains questions about personal data, such as  
430 whether the participants have already used some computer-based learn-  
ing software before, especially educational games, as well as their prior  
knowledge on CM. In this way, we separate learners into 3 categories as  
follows: (1) *novice*, (2) *intermediate* and (3) *expert*. This information is  
useful to explain the obtained results during the debriefing session since  
435 learners' evaluation depends strongly on their profile before playing.
2. **Brief presentation of the game:** the purpose of this step is to put  
the participants in the context by explaining the rules/instructions of the  
game, and showing them how to use and interact with the platform.
3. **Launching the game session:** participants were invited to play a net-  
440 work session created by the moderator in iScen. The game session lasted  
on average 20 minutes. Videos of the participants' faces were captured  
while they interacted with the game using webcams equipments.
4. **Collection of interaction traces after the end of the experiment:**  
the collected traces are the video recordings captured during the session  
445 as well as the log files automatically generated by iScen especially the  
exchanged messages using the text chatting system.
5. **Performing a post-evaluation of participants' emotional states:**  
through a statistical analysis of the responses obtained via a post-evaluation

questionnaire to measure the level of engagement for a proof purpose. This  
questionnaire will be described in the following subsection.

6. **A human debriefing session:** this step aims to analyze the players' learning: their performances, and their acquired knowledge but also their missing skills.

#### 4.2.3. Post-Evaluation Questionnaire Description

The *Game Engagement Questionnaire* (GEQ), proposed by [34] and validated by [35], is one of the most commonly used self-report questionnaires in the player experience field for measuring engagement specifically elicited while playing games. The GEQ has a modular structure consisting of: the *core module*, the *social presence module*, and the *post-game module*. All three modules are meant to be administered immediately after the game session has finished. From the perspective of our study, only the *core module* is useful since it allows to measure players' engagement. As shown in the online available link, the core module<sup>5</sup> is a 33-item scale which is designed to measure game players' experience across seven dimensions namely *Immersion*, *Flow*, *Competence*, *Positive Affect*, *Negative Affect*, *Tension*, and *Challenge*. Dimension scores are computed as the average value of their items. Each item consists of a statement on a five-point Likert scale ranging from 1 (not at all) to 5 (extremely) [35]. Examples of statements of the dimensions are: "*I was interested in the game's story*" (*Immersion*), "*I was deeply concentrated in the game*" (*Flow*), "*I felt competent*" (*Competence*), "*I felt happy*" (*Positive Affect*), "*I felt bored*" (*Negative Affect*), "*I felt frustrated*" (*Tension*), and "*I felt pressured*" (*Challenge*). The following section aims to show and to discuss our major findings.

## 5. Results Analysis

In this section, we show the obtained results from the method application on the developed scenario. Then, we discuss our findings by comparing evaluation

---

<sup>5</sup><https://docs.google.com/forms/d/1LoRJwD0KGr6rQTpteHLNYwldTuGFingIpu8XRHbarrY/edit?usp=sharing>

results to learners' responses of the GEQ as well as the limitations of our study.

### 5.1. Obtained Results

As mentioned in the Contribution section, our study aims to answer the following research question: *"How to evaluate the usability of collaborative CMSGs?"*.

480 In doing so, we need to perform an evaluation of learners' affective states at two levels: individual and collective as proposed by our method. Therefore, we divide our results analysis into two main parts as follow:

- **Individual evaluation of learners' affective states:** at this level, we study the impact of stress on affective dynamics to produce both formative and summative evaluation of learners' affective states. In fact, we 485 exploited the proposed mapping to perform a fine-grained analysis (every 20 seconds) of the dynamics of learners' affective states during an approximately 20 minutes training session based on facial features. We evaluated what affective states were present in each 20-second interval. This interval 490 length is chosen after a number of tests showing that shorter clips (or periods) do not provide sufficient and reliable data, and longer clips are hard to evaluate because they often mix different emotional states [22]. Comparing to several predictions proposed by the CDM (shown in figure 2), some of these predictions have been confirmed while others not addressed by the model are identified by our method. The supported predic- 495 tions include the transitions from the state of engagement into confusion, confusion into frustration, and frustration into boredom which naturally occurred. The two predictions that have been identified, but were unexpected in the model, include the transitions from frustration to confusion and boredom to frustration. First, even the transition from frustration 500 into confusion occurred rarely, we believe that some frustrated participants, could view the situation as a challenge and become more energized; and ultimately enter the confusion state while trying to resolve the current misunderstanding. Second, the transition from boredom into frustration 505 occurred significantly when we detect a high activation level of some FAU

characterizing the stress emotion.

To resume, our findings suggest that some aspects of the model might need refinement to be adapted in the context of CMSGs.

Table 4 resumes the possible affective transitions proposed by our method.

510

Several significant transitions from an affective state at time  $T_i$  to another affective state at time  $T_{i+1}$  were identified.

Time $T_i$	Time $T_{i+1}$			
	Boredom	Engagement	Confusion	Frustration
Boredom		-	-	++
Engagement	-		+	-
Confusion	-	+		+
Frustration	+	-	++	

Table 4: Affective Transitions. (+) shows that the transition occurred as predicted by the CDM. (-) shows that the transition is highly unlikely. (++) shows our discovered transitions.

When aggregated across the all participants at the end of training process, our results indicated that 25% of learners felt engagement, 50% expressed boredom, 25% felt frustration, and 0% experienced confusion as shown in figure 6.

515

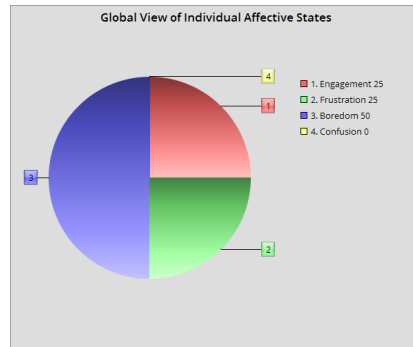


Figure 6: Global View of Individual Affective States of Players

- **Collective evaluation of learners' affective states:** at this level, we exploit all the individuals evaluation of learners' affective states to produce a collective evaluation. In fact, the global view of individual affective states of players (shown in figure 7) allowed us to decide the polarity of group emotion by applying our decision tree model. Hence, we can deduce that the group emotion is negative (25% engagement + 0% apathy + 50% boredom + 25% anxiety  $\Rightarrow$  negative class) and thus the team performance is also negative. In fact, this interpretation can be explained by the fact that all participants are situated, for the first time, in an emergency evacuation procedure based on a virtual training environment. It can also be a consequence of limited learners' guidance and assistance carried out by the instructor during the training process in order to better achieve the game objectives.

The generated results aim to help learners as well as researchers and instructors to understand how participants behave during a session based on a CMSG by analyzing learners' emotions. In addition, such results are of great interest since they can be exploited by researchers to adapt some game aspects (especially the game challenge) according to the detected emotional profiles of learners; and thus to enhance their learning experiences.

## 5.2. Discussion

In order to test our method, the obtained summative evaluation results are compared to self-reported subjective descriptions based on the GEQ. After the collaborative iScen game session, the GEQ core module is presented to learners as an electronic questionnaire to complete. The descriptive statistics obtained from learners responses are reported in Table 5.

As shown in Table 5, positive feelings are much less severe and less frequently experienced compared to negative feelings (lower than the mid-value of the scale). In fact, participants reported the level of positive affect to be low (2.00). More specifically, results analysis shows that immersion (reflecting how players felt



strongly connected with the game) and Flow (indicating whether players lost track of their own effort and/or the passage of time during the game) receive respectively average degrees (2.13 and 2.34).

The dimension negative affect receives the highest value of all (4.28). This  
 550 result indicates that playing the game engendered some negative emotional experiences in particular boredom. In addition, participants experience a certain high degree of tension (3.96) in the form of specific negative emotions like frustration. Moreover, in terms of challenge, participants report that the game environment is difficult and challenging (3.43) according to their level of com-  
 555 petence (2.43). All these results confirm the negative group emotion detected after the application of the proposed method on the same CM scenario.

Dimension	Mean	Standard Deviation	Max	Min
Immersion	2.13	0.60	4.00	1.00
Flow	2.34	0.62	3.55	1.13
Positive affect	2.00	0.50	3.00	1.00
Negative affect	4.28	0.85	5.00	2.00
Tension	3.96	0.77	5.00	1.65
Challenge	3.43	0.68	5.00	2.00
Competence	2.43	0.56	3.00	1.20

Table 5: Descriptive statistics for dimensions of the core module of GEQ

Our findings showed that the final affect annotations obtained via our method correlate well with subjective responses to the GEQ. Indeed, our method focuses on evaluating learners’ affective states using unobtrusive techniques of collect-  
 560 ing and analyzing data without affecting the high level of engagement provided by the game. Compared to the GEQ, our method represents an objective and

rapid manner to analyze learners’ emotions and thus to infer their affective states without the need of physiological sensors or complex algorithms.

However, our research study includes some limitations. In fact, during the use  
565 of the game-based training environment, it appeared that learners experienced more negative affective states, such as boredom and frustration, rather than transited into positive states of engagement, or even confusion. This can be explained by the fact that the change from an emotional state to another one strongly depends on several factors not addressed in this study, including person-  
570 ality traits, social interactions and prior knowledge on the studied subject [15]. These factors should be taken into account in order to more deeply understand the dynamics of affective states over time during a game session. Moreover, our experimental protocol can not validate totally our contribution since the experimentation involves few human participants and generates small data sets. So,  
575 we need to extend this study to a larger sample of participants and to perform many advanced experiments in order to draw generalized conclusions about the usability of the considered CMSG.

## 6. Conclusion and Future Works

Serious games are proved to be an efficient tool to incentivize people to im-  
580 prove their lifestyle and especially to help people with special needs [36]. In this article, we use AI technologies to investigate the role of emotions in improving the usability of CMSGs. In fact, this global objective is composed of three sub-objectives. The first is to evaluate the temporal dynamics of learners’ affective states during playing at fine-grained level (every 20 seconds) in order to adapt  
585 the game challenge to their current profiles. The evaluation of this affective dynamics relies on detecting basic emotions from facial features and studying the impact of stress, a specificity of CM domain, on transitions between affective states. Our second sub-objective is to evaluate their final affective states at the end of training process by selecting the dominant emotion based on a new map-  
590 ping between affective states and basic emotions. The third sub-objective aims

to explore the final individuals' affective profiles to generate the group emotion at a collective level.

Our experimental results confirm the major predictions of the cognitive disequilibrium model to assess learners' affective dynamics during learning. Our findings also suggest refinements for the model during crisis management training due the stress impact on affective transitions, especially the presence of boredom $\Rightarrow$ frustration and frustration $\Rightarrow$ confusion oscillations. These findings confirm that it is possible to detect emotions in a less invasive way and without the need of physiological sensors or complex algorithms.

600

In future works, we plan to test the generalizability of our contribution on new students' populations within another multi-players CMSG which is currently under development using Unity3D game engine. In particular, we will develop a game scenario that takes into account the needs of people with mental and physical disabilities in case of disaster, and we want to extend this study to a larger sample of participants based on this scenario. For example, we can simulate a behavior of a player character with a special need who is supposed to manage his/her situation in case of emergency. Furthermore, we plan to exploit our evaluation results to adapt various aspects of the game to the detected learners' affective profiles to regulate their emotions when they are feeling some negative emotions. In fact, recognizing the players' affective states allows the researcher/instructor to adjust the challenge level of the game according to their competencies level basing on their current emotional states (engaged, confused, bored or anxious) [37]. For example when players feel anxious, a decrease of the challenge level seems to be effective in helping them to transit into the Flow state again.

610

615

## 7. Acknowledgment

The authors would like to thank EVERSIM (Simulation & Serious Games) society for providing the iScen platform and assisting in its use.

## 620 References

- [1] I. Daoudi, R. Chebil, E. Tranvouez, W. L. Chaari, B. Espinasse, Towards a grid for characterizing and evaluating crisis management serious games: A survey of the current state of art, *International Journal of Information Systems for Crisis Response and Management* 9 (2017) 76–95.
- 625 [2] A. Solinska-Nowak, P. Magnuszewski, M. Curl, A. French, A. Keating, J. Mochizuki, W. Liu, R. Mechler, M. Kulakowska, L. Jarzabek, 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* 31 (2018) 1013–1029.
- 630 [3] W. E. Walker, J. Giddings, S. Armstrong, Training and learning for crisis management using a virtual simulation/gaming environment, *Cognition, Technology & Work* 13(3) (2011) 163–173.
- [4] M. A. Oulhaci, E. Tranvouez, S. Fournier, B. Espinasse, Improving players’ assessment in crisis management serious games: The simfor project, *Information Systems for Crisis Response and Management in Mediterranean Countries*. Springer International Publishing (2015) 85–99.
- 635 [5] Assisting People with Disabilities in Case of Disaster. European Network for Psychosocial Crisis Management Recommendations for Preparedness, Response and Recovery, Federal Office of Civil Protection and Disaster Assistance, 2017.
- 640 URL [www.bbk.bund.de](http://www.bbk.bund.de)
- [6] S. Sauvagnargues, D. Lapierre, P. L. and Noémie Fréalle, F. Tena-Chollet, P. Ayral, A. Bony-Dandrieux, J. Tixier, Concepts, Tools and Methods for Crisis Management training, 2018, Ch. Decision-making in Crisis Situations: Research and Innovation for Optimal Training.
- 645 [7] H. Reis, D. Alvares, P. Jaques, S. Isotani, Analysis of permanence time in emotional states: A case study using educational software, in: Nkambou

R., Azevedo R., Vassileva J. (eds) Intelligent Tutoring Systems. Lecture Notes in Computer Science, vol 10858. Springer, Cham., 2018.

- 650 [8] M. Csikszentmihalyi, Flow: The psychology of optimal experience, Harper & Row, 1990.
- [9] R. Tadayon, A. Amresh, T. McDaniel, S. Panchanathan, Real-time stealth intervention for motor learning using player flow-state, in: IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)., 2018.
- 655 [10] N. Haferkamp, N. C. Kraemer, C. Linehan, M. Schembri, Training disaster communication by means of serious games in virtual environments, Entertainment Computing 2 (2011) 81–88.
- [11] G. Rebolledo-Mendez, K. Avramides, S. D. Freitas, K. Memarzia, Societal impact of a serious game on raising public awareness: the case of floodsim, in: the ACM SIGGRAPH Symposium on Video Games, 2009, pp. 15–22.
- 660 [12] V. S. R. Silva, A. R. Dargains, S. P. de Almeida Silva Felício, P. R. de Azevedo Souza, F. Sampaio, C. L. R. da Motta, M. R. da Silva Borges, J. O. Gomes, P. V. R. de Carvalho, Stop disasters: serious games with elementary school students in rio de janeiro, in: 8th International Technology, Education and Development Conference, 2014, pp. 1648–1659.
- 665 [13] V. J. Shute, Stealth assessment in computer-based games to support learning, Computer games and instruction 55(2) (2011) 503–524.
- [14] R. W. Picard, Affective Computing, MIT press, Cambridge, 1997.
- 670 [15] R. C. D. Reis, S. Isotani, C. L. Rodriguez, K. T. Lyra, P. A. Jaques, I. I. Bittencourt, Affective states in computer supported collaborative learning: Studying the past to drive the future, Computers & Education 120 (2018) 29–50.

- [16] B. Guthier, R. Dörner, H. P. Martinez, Affective Computing in Games, Vol. 9970 of Lecture Notes in Computer Science, 2016, Ch. Entertainment Computing and Serious Games, pp. 402–441.
- [17] E. Paul, Basic emotions, T. Dalgleish & M. J. Power (Eds.), Handbook of cognition and emotion, 45-60. New York, NY, US: John Wiley & Sons Ltd, 1999.
- [18] J. A. Russell, Emotion, core affect, and psychological construction, *Cognition and Emotion* 23(7) (2009) 1259–1283.
- [19] A. Ortony, G. L. Clore, A. Collins, The Cognitive Structure of Emotions., Cambridge: Cambridge University Press, 1988.
- [20] S. Poria, E. Cambria, R. Bajpai, A. Hussain, A review of affective computing: From unimodal analysis to multimodal fusion, *Information Fusion* 37 (2017) 98–125.
- [21] S. K. D’mello, J. Kory, A review and meta-analysis of multimodal affect detection systems, in: *ACM Computing Surveys (CSUR)*, 47(3), 43, 2015.
- [22] S. D’Mello, A. Graesser, Dynamics of affective states during complex learning, *Learning and Instruction* 22(2) (2012) 145–157.
- [23] S. D’Mello, A. Graesser, Modeling cognitive-affective dynamics with hidden markov models, in: the Annual Meeting of the Cognitive Science Society (Vol. 32, No. 32)., 2010.
- [24] C. A. Cruz, J. A. R. Uresti, Player-centered game ai from a flow perspective: Towards a better understanding of past trends and future directions, *Entertainment Computing* doi:<http://dx.doi.org/10.1016/j.entcom.2017.02.003>.
- [25] T. Baltrusaitis, A. Zadeh, Y. C. Lim, L.-P. Morency, Openface 2.0: Facial behavior analysis toolkit, in: 13th IEEE International Conference on Automatic Face & Gesture Recognition, 2018, pp. 59–66.

- [26] W. V. Friesen, P. Ekman, Emfacs-7: Emotional facial action coding system, university of California. <https://www.paulekman.com/> (1983).
- [27] C. T. Tan, D. Rosser, S. Bakkes, Y. Pisan, A feasibility study in using facial expressions analysis to evaluate player experiences, in: The 8th Australasian Conference on Interactive Entertainment: Playing the System, New York, NY, USA, 2012.
- [28] L. Kuncheva, A theoretical study on six classifier fusion strategies, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2) (2002) 281–286.
- [29] A. Alberdi, A. Aztiria, A. Basarab, Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review, Journal of biomedical informatics 59 (2016) 49–75.
- [30] H. Gao, A. Yüce, J.-P. Thiran, Detecting emotional stress from facial expressions for driving safety, in: IEEE International Conference on Image Processing (ICIP), 2014, pp. 5961–5965.
- [31] M. Feidakis, A Review of Emotion-Aware Systems for e-Learning in Virtual Environments, Academic Press, 2016, Ch. Formative Assessment, Learning Data Analytics and Gamification. ICT Education Intelligent Data-Centric Systems, pp. 217–242.
- [32] S. Drazin, M. Montag, Decision tree analysis using weka, machine Learning – Project II (2012).
- [33] S. A. Kumar, D. V. M.N, Efficiency of decision trees in predicting student’s academic performance, Computer Science & Information Technology (CS & IT).
- [34] J. H. Brockmyer, C. M. Fox, K. Curtiss, E. McBroom, K. Burkhart, J. N. Pidruzny, The development of the game engagement questionnaire: A measure of engagement in video game-playing, Journal of Experimental Social Psychology 45(4) (2009) 624–634.

- [35] D. Johnson, M. J. Gardner, R. Perry, Validation of two game experience scales: The player experience of need satisfaction (pens) and game experience questionnaire (geq), *International Journal of Human-Computer Studies* 118 (2018) 38–46.
- [36] V. Wattanasoontorn, I. Boada, R. García, M. Sbert, Serious games for health, *Entertainment Computing* 4 (2013) 231–247.
- [37] A. Stein, Y. Yotam, R. Puzis, G. Shani, M. Taieb-Maimon, Eeg-triggered dynamic difficulty adjustment for multiplayer games, *Entertainment Computing* 25 (2018) 14–25.