Emerging opportunities for education in the time of COVID-19: Adaptive e-learning intelligent agent based on assessment of emotion and attention

Marko Horvat, Tomislav Jagušt
Faculty of Electrical Engineering and Computing, University of Zagreb
Department of Applied Computing
Unska 3, HR-10000 Zagreb, Croatia
{Marko.Horvat3, Tomislav.Jagust}@fer.hr

Abstract. *The COVID-19 pandemic is a disastrous and rapidly evolving situation. It brings changes to many aspects of life, including teaching. A very important area for improvement of e-learning pedagogies is the ability to assess student engagement and the learning curve objectively and continuously. To this regard, we propose a novel procedure for personalized and adaptive assessment of learning performance based on methods for automated estimation of affective states. In this preliminary report we envision an intelligent agent which constantly monitors students' behaviour parameters during online learning classes. Using unobtrusive video surveillance and machine learning the agent appraises key psychophysiological features related to emotion and attention.*

Keywords. digital learning, adaptive learning, technology enhanced learning, intelligent algorithms, emotion recognition

1 Introduction

Coronavirus disease 2019 (COVID-19) is a global pandemic unprecedented in modern times. It is forcefully changing all aspects of everyday life. However, in this adverse course of events, new opportunities arise for adaptation and improvement. In this regard, online learning or e-learning is quickly emerging as the predominant education paradigm. Schools are literally forced to shift academic activities online and away from customary methods of teaching. It has previously been shown that technology can enhance learning in and out of the classroom, especially by impacting student interest, motivation, and engagement (Jagušt, Botički & So, 2018). In traditional education paradigms feedback on education, quality is obtained through various knowledge assessments and student questionnaires. However, although commonplace and straightforward, these methods are also generic and largely unsuitable for learning personalization (Jagušt & Botički, 2019).

Adaptive learning provides personalized learning experience with the goal of optimizing learning efficiency, increasing student engagement, and addressing specific needs, interests, or preferences of each learner (Hsiao, Sosnovsky & Brusilovsky, 2010). Adaptive pedagogies are of the most discussed topics in the field of technology-enhanced learning (Kirschner, 2017). It is advocated to help balancing different abilities, learning speed, skills, and to improve student engagement, interest, and knowledge (Shute & Zapata-Rivera, 2012).

New unobtrusive video sensor technologies enable monitoring and qualitative measurements of learners' facial expression features and gaze points, which can uncover nonverbal cues related to the success of e-learning education paradigms (Li et al., 2016). These traits related to emotional and cognitive functioning are involuntarily expressed during learning sessions. Detection of such indicative features for all students in a class may be very time-consuming in practice, and thus exceptionally difficult even for the most experienced educators or almost impossible for a novice. However, these visual features may be extremely valuable indicators for the successful learning paradigm, adaptation, and personalization. The additional information could help in avoiding potential misinterpretation, and even misunderstanding of students' behaviour, better evaluation of the acquired knowledge, and achieved levels of proficiency during a series of online learning sessions.

For this paper, a software application has been developed written in C#.NET computer language for emotion detection from facial features using the Active Shape Model (ASM) method. Face localization and extraction is accomplished with Microsoft Azure Cognitive Services API. Although a number of technologies for estimation of emotion and attention are available as commercial-of-the-shelve (COTS) components, an integrated recommendation system for e-learning conceived as an adaptive intelligent agent – and described in this preliminary report – cannot be commercially acquired.

2 Related work

In the literature, there are several previously published research papers on the topic of student engagement and attention in classrooms, such as (Henrie, Halverson & Graham, 2015) (Thomas & Jayagopi, 2017). It has been long established that estimation of long-term student engagement in the learning process is needed in order to evaluate courses and improve learning results (Fredricks et al., 2011). This evaluation is usually done through questionnaires, but with the proliferation of modern e-learning, it became possible to collect implicit usage data to estimate the activity and engagement of students or children within learning activities (Martinez et al., 2015). A review of research on the measurement of student engagement in technology-mediated learning (Henrie, Halverson & Graham, 2015) has provided a review of quantitative and qualitative observational measures (instruments) to measure behavioural, cognitive, and emotional indicators of student engagement. Attention was classified as one of the factors of cognitive engagement, while interest, anxiety, and boredom contributed to emotional engagement. Attention is best described as the sustained focus of cognitive resources on information while ignoring distractions (Myers, Stokes & Nobre, 2017). For a recent review of pedagogical guidelines for the creation of adaptive digital educational resources see (Rozo & Real, 2019).

In the field of education, the terms of sustained attention or vigilance describe the ability to maintain concentration over prolonged periods of time, such as during lectures in the classroom. Pedagogical research is often focused on maintaining student attention (i.e. concentration, vigilance) during lectures because sustained attention is recognized as an important factor of the learning success (Al-Shargie et al., 2019). However, tracking of individual students' attentive state in the classroom by using questionnaires is difficult and interferes with the learning process, which is also the case for using psychophysical data sensors (Chen, Wang & Yu, 2017).

Besides the estimation of engagement and attention, automated measurement of affective parameters in the general public has also been intensively researched. Importantly, it has been proven that the automated estimation of emotional states is feasible in practice and can be used to make valid conclusions on perception, behaviour, attention, and emotion (Ćosić et al., 2013).

Non-intrusive visual observation and estimation of affective parameters is commonly using recorded video (RGB) signal, for example, to estimate student engagement from facial expressions (Monkaresi et al., 2016). A number of studies have been reported on automatic emotion estimation methods using various physiological channels (video, EKG, EMG, eye tracking, ...) in off-line and real-time settings (per example (Shu et al., 2018) (Lim, Mountstephens & Teo, 2020) (Kukolja et al., 2014) (Horvat et al., 2018). Eye tracking devices are very successful in measuring affective parameters such as concentration in the computerized learning environments. In (Tonguç & Ozkara, 2020) computer vision methods are used to register faces and estimate student emotions from facial expressions during a lecture. This study deals with the engagement of a single student during computerized learning which exactly corresponds to the types of pedagogies used during the COVID-19 pandemic. The above cited reviews of affect estimation methods emphasize the importance of face gaze and facial expression as clues to assess cognitive engagement or inattention of students.

In addition to a low-cost HD camera, Microsoft Kinect One is potentially very beneficial type of sensor because it provides advanced capabilities to detect facial features and face gaze. However, this device is much rarer and more expensive than nowadays ubiquitous built-in cameras. In a thorough and methodical study, Kinect sensor has been used to predict students' attention in the classroom (Zaletelj & Košir, 2017. Attention of students has been successfully modelled using eye tracking sensors (Veliyath, 2019), and types of attention have been classified with trackers operating in the thermal part of the spectrum (Abdelrahman, 2019). Also, in a recent research, visual attention has been unambiguously correlated to the observed visual features (Moroto, 2019). The gaze direction has been detected from combined video and depth signals (Hong et al., 2018) and utilized in the visual attention model to estimate human-to-human interaction. In one system using artificial neural networks, student-student behaviour has been classified using a spatio-temporal model from sequences of digital images (Jalal & Mahmood, 2019).

3 Estimation of emotions and attentions based on facial expressions

Facial expressions can be considered as aspects of both an emotional response to certain stimuli and of social communication. Facial expressions are shaped through muscular activity, which is driven by complex underlying neurobiological mechanisms that are still being intensively researched (Ćosić et al., 2015). The process of facial expression recognition involves linking visual representation of the face, reflecting a generation of knowledge about the category of emotion that it belongs to.

Facial expressions are a rich source of information in the process of understanding emotions, which can be represented as a complex psycho-physiological and mental state that consists of various multimodal responses (physiological, vocal and acoustic, facial, etc.). Based on the similarity of facial responses, induced by a group of stimuli with similar emotional content, they can be divided into different disjoint

categories (discrete emotional states). The categorization process is of paramount importance in a social communication environment, due to the high informational complexity and rapid behavioural response to facial emotions.

The first step in the estimation of emotions from facial features using visual sensors is to detect a human face in a video. To do this, the incoming video stream is sampled into frames. Individual static pictures are then evaluated separately. The relation of two pictures in terms of detected objects or emotions is not necessary. The Viola-Jones method is efficiently used for face detection in the literature (Deshpande & Ravishankar, 2017). The result of Viola-Jones detector is a rectangle that envelopes the region of the sampled video frame, most probably, containing a human face. After faces in a scene have been identified, the system can proceed to recognize facial features and estimate emotions from each sampled frame. An example of facial detection is in Fig. 1. Microsoft Azure Cognitive Services (Microsoft Azure Services, 2020) offers computer software developers a convenient tool for facial detection that can be easily integrated with different application architectures.

Figure 1. Face detection using Microsoft Cognitive Services API with the accompanying JSON code defining one of the face boxes in the picture.

Facial feature extraction is based on relative displacements of main anatomical region that are coded by the Active Shape Model (ASM) method (Le & Savvides, 2016). ASMs are statistical models that iteratively deform according to the current shape of object in the picture. The ASM method has two steps: 1) training based on examples, and 2) model fitting. The training is usually done by methods such as Principal Component Analysis (PCA) (Deshpande & Ravishankar, 2017). The variation of the shape of the model depends on variations of the position of facial points in the learning set. In doing so, the shapes created by the variation of a neutral face are in accordance with the contribution of each eigenvector in the total variance of the learning set. Fitting of the ASM is an iterative procedure that attempts to find a set of transformation parameters (e.g. scaling, translation, rotation) so the initial form of the ASM is as close as possible to the defined facial points. The procedure stops if at least one halting condition has been met: 1) the difference in the parameters of two consecutive iterations is below of a certain threshold, or 2) the maximum number of iterations has been achieved.

Based on their visual appearance, face emotions can be represented by a set of geometric properties (i.e. spatial features), which describe both position, movement direction, and temporal dynamics of main anatomic structures of the face. A technique for measurement of facial movements which is still used today has been described in (An et al., 2017). In this system, human facial expressions, better known as Facial Action Coding System (FACS) have been taxonomized (Rosenberg & Ekman, 2020). The FACS encodes various movements of individual facial muscles which are called Action Units (AUs). Action Units are a standardized form of describing visual movements of particular anatomical face regions. The presence of specific AU can be detected by observing the spatial characteristic of a certain related morphological structure of the face expression with regard to the corresponding spatial characteristic in the neutral face. Neutral face is a special representation of spatial properties of main anatomical regions of the face where the mouth is closed, the gaze is directed perpendicular to the screen plane, eyes are open and the eyelids are tangent to the iris (Sun et al., 2016). In total there are 46 main AUs which are used for coding movements of facial muscles.

Facial expressions of six basic emotions (or norms as they are also called) are demographically and culturally independent (Ekman & Friesen, 1971). A set of AUs describing spatial characteristic of face regions for basic emotions has also been established. The list of AUs relevant to the six basic norms adapted from (Ekman & Friesen, 1971) is listed in Table 1. Primary AUs must be present for an expression to be classified. Auxiliary AUs do not have to be detected but they will increase belief in the classification.

Table 1. List of AUs relevant for estimation of the six basic emotions. Adapted from (Ekman & Friesen, 1971).

Norm	Primary AUs	Secondary AUs
Happiness	6, 12	25, 26, 16
Sadness	1, 15, 17	4, 7, 25, 26
Disgust	9, 10	17, 25, 26
Surprise	5, 26, 27, 1+2	
Anger	2, 4, 7, 23, 24	17, 25, 26, 16
Fear	20, 1+5, 5+7	4, 5, 7, 25, 26

As an example, the difference between spatial features of a neutral face and expression of happiness is shown in Fig. 2. As can be seen in Fig. 2, and from Table 1, the expression of happiness is identified with two primary action units AU6 ("Cheek raiser") and AU12 ("Lip corner puller"), and one secondary action unit AU16 ("Lower lip depressor"). Apart from detecting specific action units, Euclidian distance between corresponding points on the face can be measured. By doing this it is possible to measure intensity of facial expressions and related emotion with a larger distance being positively correlated with an increased emotion intensity (Ćosić et al., 2013).

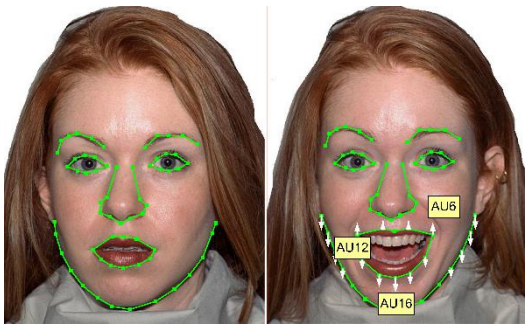

Figure 2. Comparation of spatial features between a neutral face (left) and expression of happiness (right) with affiliated AUs.

Students' attention may be estimated through observation of eye movements during e-learning course. It has been established, per example (Ono & Taniguchi, 2017), that the number and duration of fixations are larger and longer for attention grabbing targets than for non-targets. Further, the probability of first fixation is higher for a target than provokes an emotional reaction than for a neutral target. The exact relationship of gaze patterns and displayed course content must be established once the e-learning intelligent agent has been developed and tested. Gaze estimation tools exist today that are already quite capable (Lasa et al., 2017). Using commonplace visual sensors novel tools can discriminate facial features and gaze direction in high resolution, and in ambient lightning conditions. A limiting factor to these methods is that they recognize facial expressions only within certain limits (the so-called headbox). Still, the freedom of movement is still sufficiently large for students to be relatively unrestricted – at least in terms of a normal range of movement while watching an online class on the screen.

4 Adaptive e-learning intelligent agent based on assessment of emotion and attention

Captured video stream transmitted to other participants is continuously analysed by the intelligent agent for recognition of facial expressions. For ease of use it is vital to use existing infrastructure, instead of requiring to equip each user with specialized devices. Such equipment may be high cost, fragile, cumbersome, or difficult to acquire. Fortunately, today all notebooks and smartphones are equipped with a camera facing the user (the so-called "selfie camera" on mobile communication devices, and "webcam" on computers).

As a student participates in an e-learning session, the intelligent agent is receiving video signal from his computer in real time. First, the agent processes raw image and segments the video feed into frames. Then a set of features is extracted from each picture. In the third step the agent makes statistical inferences about the student's emotion and attention based on visual features. The student's emotional state is evaluated for intensity of six basic emotions (happiness, sadness, surprise, fear, anger, and disgust) on the scale 0 – 100, with 0 indicating total absence of an emotion and 100 the most intensive sensation of that particular emotion. Student attention level and target is inferred with eye gaze estimation. Rapidly shifting focus to the lectured topic or current assignment will indicate responsiveness and lecture awareness. Furthermore, prolonged periods of low arousal, if preceded with high valence or arousal may be indicative of low interest in the conveyed subject. Specially if they are accompanied with low body movements and slow or stationary gaze point.

The integrated results are displayed on the teacher's terminal. The functioning of the e-learning intelligent agent is illustrated in the figure below (Fig. 3).

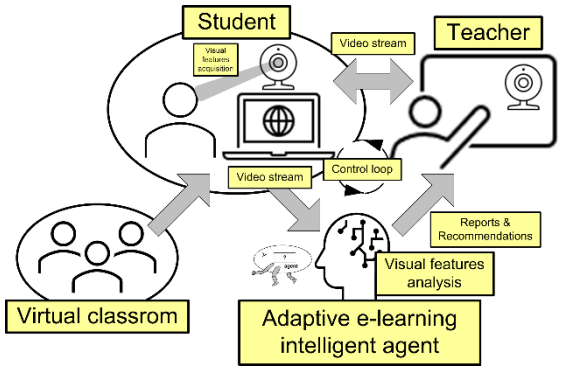

Figure 3. Diagram of adaptive e-learning intelligent agent based on assessment of emotion and attention.

The psychophysiological impact of a lecture or exam is evaluated for each student separately by the intelligent agent. The teacher is notified of extreme levels (low or high), and sudden and sharp variations of emotion and attention, in relation to a specific class content. Also, student's values are compared with performance in previous classes, and with other students in the same class and school. Thus, the most appropriate next step in their e-learning process may by indicated to the teacher. This adaptive process is illustrated in UML activity diagram (Fig. 4).

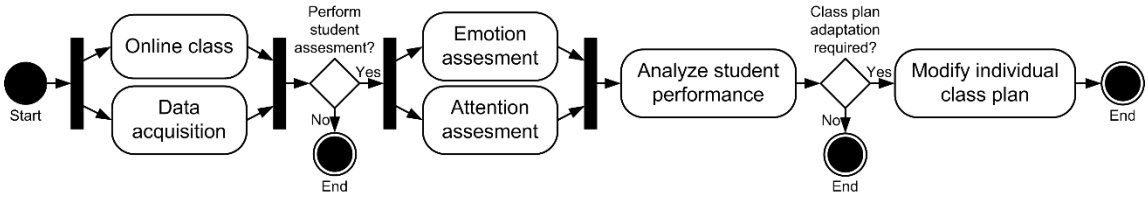

Figure 4. UML activity diagram describing the dynamic aspects of the proposed concept. Class plan is independently conformed to each student based on emotion and attention assessments by the intelligent agent, and the teacher's decision.

In later versions of the intelligent agent, the intelligent agent may not only report but also recommend on the best actions to improve pedagogic efficiency.

It is very important to objectively measure and evaluate students' overall facial and eye movement activity throughout the learning sessions. There are a variety of options for evaluating cognitive engagement, like the sum of all relative spatial offsets of facial characteristic points, e.g. ASM characteristic points, from the baseline neutral face point set. Information regarding differences between the estimation result of student's personalized estimator and the expected emotion estimation value, e.g. by using referent estimator learned over the general student population for the same class, can also be integrated in the estimation process of a student's acquired knowledge levels during online learning sessions.

The online course is segmented into lectures and exams, allowing the course to be dynamically and modularly constructed thereby optimally improving adoption of lecture materials for each student.

The adaptive process is repeated until the end of all lectures has been reached. The system is not capable of independent decision making and only acts act in a decision-support role letting an expert to decide on the course of study. Importantly, because of the unobtrusive nature of the video sensors, the student is unaffected by monitoring procedures, learns and acts spontaneously.

5 Discussion

In our approach, the key visual sensor is a high-definition camera. The same camera is used for e-learning, i.e. additional sensors are not required. Such sensors are ubiquitous and integrated into all mobile computers. They are easy to manage and operate by end-users and software developers alike.

The biggest concern for the practical application of the proposed intelligent agent is the reliability of the information on emotion and attention of students from consumer-grade visual sensors. The increased quality and abilities of such sensors, the popularity of mobile computers and smartphones contribute to the prospect of using this paradigm virtually anywhere. Another potential concern is the usefulness of obtained information for improving learning pedagogies. However, a great body of previously published research give a strong reason to believe that the adaptive e-learning intelligent agent based on assessment of emotion and attention is a workable concept. Intelligent systems and computer vision technologies are mature and already proven in practice.

Because of the limitations of employed algorithms, estimation of emotions from facial expression will depend on the possibility of determining a neutral face. The ground-truth may be either a universal neutral face (as in Section 3), or it can be determined before the start of the class for each student or a group of students.

Information security and privacy concerns are a very important issue that must be considered. Legal frameworks and the appropriate legislation must be respected to protect the privacy of personal data. The use of the system should be optional and not mandatory. Also, students and teachers must be informed before on the purpose of the intelligent system, and how the collected data is utilized. The collected data should not be misused. In this respect, it would be best if personal information is not permanently stored, and only used during sessions within a predetermined timeframe.

Also, it would be beneficial if the proposed paradigm could be integrated into a popular Learning Management System (LMS) as an optional feature. For example, as a Moodle plugin[1]. In this way, features of both systems would be immediately available from a single standardized user interface to a wide audience. Also, it would not be necessary for users to set up and manage the two systems separately.

6 Conclusion and future work

The COVID-19 pandemic is a tragic, sudden, and rapidly changing situation greatly affecting global health and economy, but it also reshapes all aspect of life including e-learning. Perhaps paradoxically, this also creates opportunities to improve learning processes comprehensively and permanently. The change is brought on not because of the choice, but out

[1] https://moodle.org/plugins/

of urgency. New visual pedagogies and intelligent algorithms are crucial in enabling this transformation.

To this regard, a new assistive paradigm for objective assessment of learners' success in acquisition of curriculum has been proposed. The intelligent agent delivers potentially useful information to teachers allowing them to adapt the teaching process for each student. Based on information importance, the instructor may decide on the actions to be takes, if any. Also, the educator can decide to what extent he wants to rely on information and recommendations that the intelligent agent provides. Data on student attention and emotion regarding class and exams units, depending on the context, may be invaluable for personalization of pedagogies. This paradigm can help students who experience difficulties in knowledge acquisition or ability to apply their knowledge into different learning settings.

It must be pointed out that a computer system with the described features currently does not exist. A number of enabling technologies for assessment of emotion and attention are indeed commercially available, but an integrated system incorporating an adaptive intelligent agent for e-learning represents a new concept.

We believe that the proposed concept could facilitate team-facilitated, active, and self-directed learning, and promoting individualized and interprofessional education. Importantly, it can be used on different platforms since visual sensors are present with a wide range of personal electronics, computers and smartphones.

Recognition of emotion and attention with commonplace visual sensors is the key technology for the proposed concept to work in practice. However, based on previously published research, intelligent systems paired with computer vision technologies are mature and already proven in practice. Hence, we believe that the described system is entirely feasible and could be realistically used to improve efficiency of e-learning paradigms.

In the future we envision the possibility of evolving the system to a truly multimodal estimation paradigm with the usage of numerous additional information channels such as pupil dilation, eye blink rate, body position, body posture, gestures, head orientation, voice recognition, keyboard usage, and mouse dynamics.

References

Abdelrahman, Y., Khan, A. A., Newn, J., Velloso, E., Safwat, S. A., Bailey, J., ... & Schmidt, A. (2019). Classifying Attention Types with Thermal Imaging and Eye Tracking. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(3), 1-27.

Al-Shargie, F., Tariq, U., Mir, H., Alawar, H., Babiloni, F., & Al-Nashash, H. (2019). Vigilance decrement and enhancement techniques: a review. Brain sciences, 9(8), 178.

An, S., Ji, L. J., Marks, M., & Zhang, Z. (2017). Two sides of emotion: exploring positivity and negativity in six basic emotions across cultures. Frontiers in psychology, 8, 610.

Chen, C. M., Wang, J. Y., & Yu, C. M. (2017). Assessing the attention levels of students by using a novel attention aware system based on brainwave signals. British Journal of Educational Technology, 48(2), 348-369.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. The handbook of emotion elicitation and assessment, 1(3), 203-221.

Ćosić, K., Popović, S., Horvat, M., Kukolja, D., Dropuljić, B., Kovač, B., & Jakovljević, M. (2013). Computer-aided psychotherapy based on multimodal elicitation, estimation and regulation of emotion. Psychiatria Danubina, 25(3), 0-346.

Ćosić, K., Popović, S., Kukolja, D., Dropuljić, B., Sedmak, G., & Judaš, M. (2015, January). Neural background of multimodal startle-type responses. In 5th Croatian Neuroscience Congress: Book of Abstracts (p. 83).

Deshpande, N. T., & Ravishankar, S. (2017). Face Detection and Recognition using Viola-Jones algorithm and Fusion of PCA and ANN. Advances in Computational Sciences and Technology, 10(5), 1173-1189.

Ekman, P. & Friesen, W. V. (1971). Constants Across Cultures in Face and Emotion. Journal of Personality and Social Psychology, 17, 124-129.

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. Regional Educational Laboratory Southeast.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. Computers & Education, 90, 36-53.

Hong, C., Yu, J., Zhang, J., Jin, X., & Lee, K. H. (2018). Multimodal face-pose estimation with multitask manifold deep learning. IEEE Transactions on Industrial Informatics, 15(7), 3952-3961.

Horvat, M., Dobrinić, M., Novosel, M., & Jerčić, P. (2018, May). Assessing emotional responses induced in virtual reality using a consumer EEG headset: A preliminary report. In 2018 41st International Convention on Information and Communication Technology, Electronics and

Microelectronics (MIPRO) (pp. 1006-1010). IEEE.

Hsiao, I. H., Sosnovsky, S., & Brusilovsky, P. (2010). Guiding students to the right questions: adaptive navigation support in an E-Learning system for Java programming. Journal of Computer Assisted Learning, 26(4), 270-283.

Lasa, G., Justel, D., Gonzalez, I., Iriarte, I., & Val, E. (2017). Next generation of tools for industry to evaluate the user emotional perception: the biometric-based multimethod tools. The Design Journal, 20(sup1), S2771-S2777.

Le, T. H. N., & Savvides, M. (2016). A novel shape constrained feature-based active contour model for lips/mouth segmentation in the wild. Pattern Recognition, 54, 23-33.

Li, J., Ngai, G., Leong, H. V., & Chan, S. C. (2016). Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics. ACM SIGAPP Applied Computing Review, 16(3), 37-49.

Jagušt, T., Botički, I., & So, H. J. (2018). A review of research on bridging the gap between formal and informal learning with technology in primary school contexts. Journal of Computer Assisted Learning, 34(4), 417-428.

Jagušt, T., & Botički, I. (2019). Mobile learning system for enabling collaborative and adaptive pedagogies with modular digital learning contents. Journal of Computers in Education, 6(3), 335-362.

Jalal, A., & Mahmood, M. (2019). Students' behavior mining in e-learning environment using cognitive processes with information technologies. Education and Information Technologies, 24(5), 2797-2821.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. Computers & Education, 106, 166–171.

Kukolja, D., Popović, S., Horvat, M., Kovač, B., & Ćosić, K. (2014). Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications. International journal of human-computer studies, 72(10-11), 717-727.

Lim, J. Z., Mountstephens, J., & Teo, J. (2020). Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges. Sensors, 20(8), 2384.

Martinez-Maldonado, R., Clayphan, A., Yacef, K., & Kay, J. (2014). MTFeedback: providing notifications to enhance teacher awareness of small group work in the classroom. IEEE Transactions on Learning Technologies, 8(2), 187-200.

Microsoft Azure Services. (2020). Retrieved from https://azure.microsoft.com/en-us/services/cognitive-services/

Monkaresi, H., Bosch, N., Calvo, R. A., & D'Mello, S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Transactions on Affective Computing, 8(1), 15-28.

Moroto, Y., Maeda, K., Ogawa, T., & Haseyama, M. (2019, March). Estimation of Visual Attention via Canonical Correlation between Visual and Gaze-based Features. In 2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech) (pp. 229-230). IEEE.

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing information during working memory: beyond sustained internal attention. Trends in Cognitive Sciences, 21(6), 449-461.

Ono, Y., & Taniguchi, Y. (2017). Attentional Capture by Emotional Stimuli: Manipulation of Emotional Valence by the Sample Pre-rating Method. Japanese Psychological Research, 59(1), 26-34.

Rosenberg, E. L., & Ekman, P. (Eds.). (2020). What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS). Oxford University Press.

Rozo, H., & Real, M. (2019). Pedagogical Guidelines for the Creation of Adaptive Digital Educational Resources: A Review of the Literature. Journal of Technology and Science Education, 9(3), 308-325.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., ... & Yang, X. (2018). A review of emotion recognition using physiological signals. Sensors, 18(7), 2074.

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. Adaptive Technologies for Training and Education, 7(27), 7–27.

Sun, W., Sun, N., Guo, B., Jia, W., & Sun, M. (2016). An auxiliary gaze point estimation method based on facial normal. Pattern Analysis and Applications, 19(3), 611-620.

Thomas, C., & Jayagopi, D. B. (2017, November). Predicting student engagement in classrooms using facial behavioral cues. In Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for education (pp. 33-40).

Tonguç, G., & Ozkara, B. O. (2020). Automatic recognition of student emotions from facial expressions during a lecture. Computers & Education, 148, 103797.

Veliyath, N., De, P., Allen, A. A., Hodges, C. B., & Mitra, A. (2019, April). Modeling Students' Attention in the Classroom using Eyetrackers. In

Proceedings of the 2019 ACM Southeast Conference (pp. 2-9).

Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. EURASIP journal on image and video processing, 2017(1), 80.