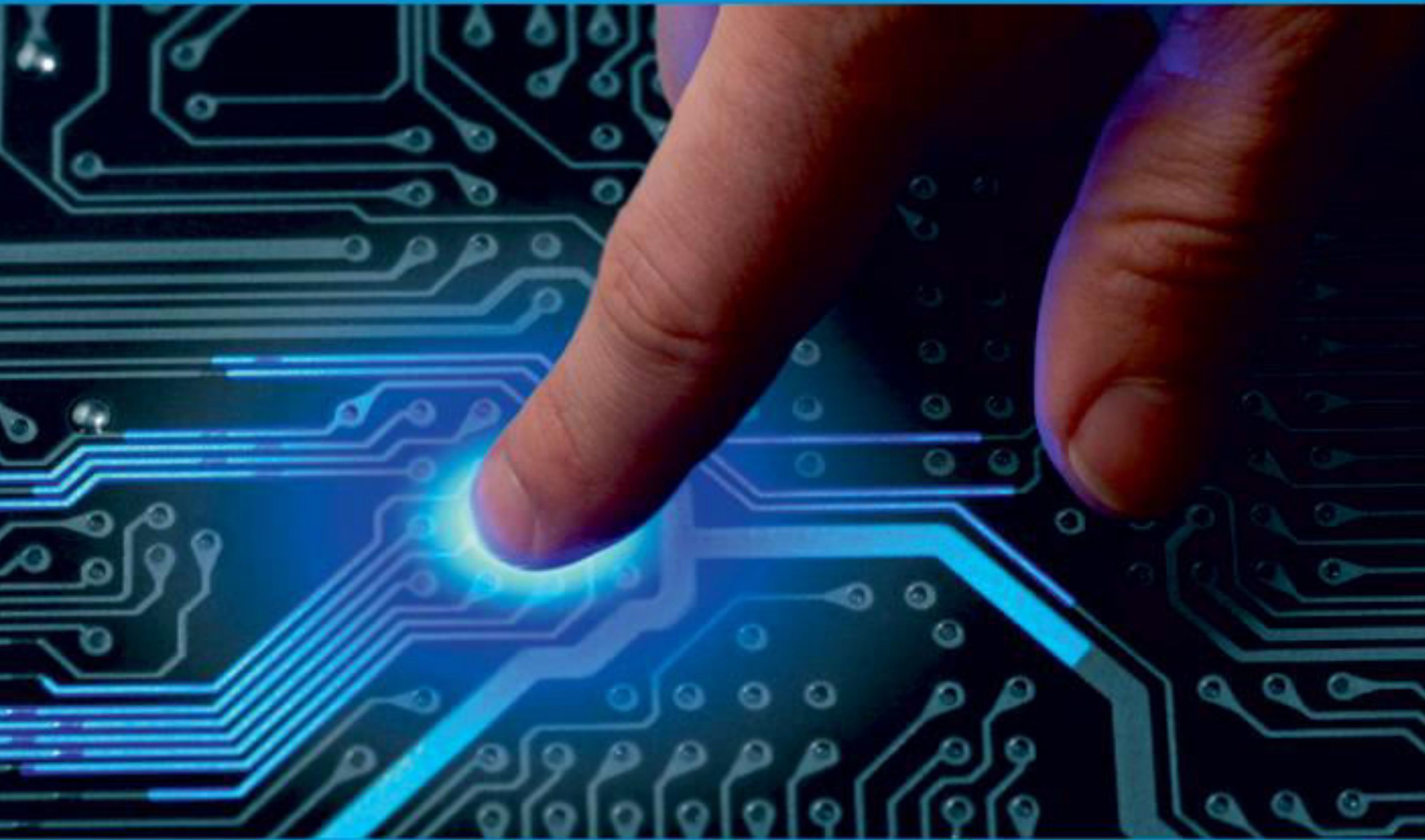
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204118 |

Analysis of Facial Expressions to Estimate the Level of Engagement in Online Lectures

S.NAZEEMA, AKKILI SAI DEEKSHTHREDDY, KANTA RAMCHARAN, SUBEDAR SAIKRISHANA

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: This research shows that it is possible to estimate the affective state of participants as well as their engagement levels. The system performed admirably in recognizing participants' learning affect during a online learning session and subsequently categorized it as negative or positive learning affect. The learning impacts were determined by categorizing participants' degrees of attention and discomfort as negative or positive. These classifications were used together as input for the feedback system. In future work, we would like to build a comprehensive assessment, and to achieve this, we propose to fuse the subunits. Furthermore, tailored feedback based on participants' emotional state and level of attention is also still underway. For instance, when a participant displays negative emotion, affective feedback will be provided to help the participant overcome the issue faced. Lastly, we suggest that a database should be created to help recognize different cognitive and emotional states for validation purposes.

KEYWORDS: Engagement; facial expression; deep network; gaze

I.INTRODUCTION

E-learning is flexible and therefore can meet various challenges posed within the sphere of information technology (IT), and most notably it has the potential to widen people's access to knowledge. The joint impact of communication and IT on learning offers various routes of exploration, such as how best to capture and keep learners' attention, as well as developing active and flexible learning environments that motivate students to learn continuously through the use of a variety of IT tools [1].

With the COVID-19 situation, e-learning has been at the forefront in providing quality education to learners. E-learning refers to using computers and online learning tools in a blended learning environment, focusing on collaborative online learning [2]. E-learning can be advantageous as it allows learners to progress at their own pace and access information at the time and place that suits them best [3, 4]. However, it also presents some challenges. One such challenge has to do with the congruence between the e-learning environments' and students' characteristics, as well as the effect of all users', students', and teachers' cultural-educational background on their personal preferences within the e-learning environment. Currently, rapid social and technological changes emphasize the need for lifelong learning. Due to the enormity of this need, it cannot be met in a traditional classroom setting. Thus, e-learning provides an alternative way to pursue lifelong learning via the internet. This mode of knowledge transmission is rapidly gaining momentum due to recent advances in computer technologies, as well as research into the pedagogical methodologies linked thereto. Online learning has become part of the "normal" landscape in training and education because it widens access to knowledge. Furthermore, it also offers learners and instructors flexibility and convenience, since they can upload/access information when it suits them best. One disadvantage is that some online learning sites provide learners with a wealth of information but do not offer them any support with using the information to construct meaning. In such situations, students are passive receivers of information, which does not promote the principles of lifelong learning [5].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204118 |

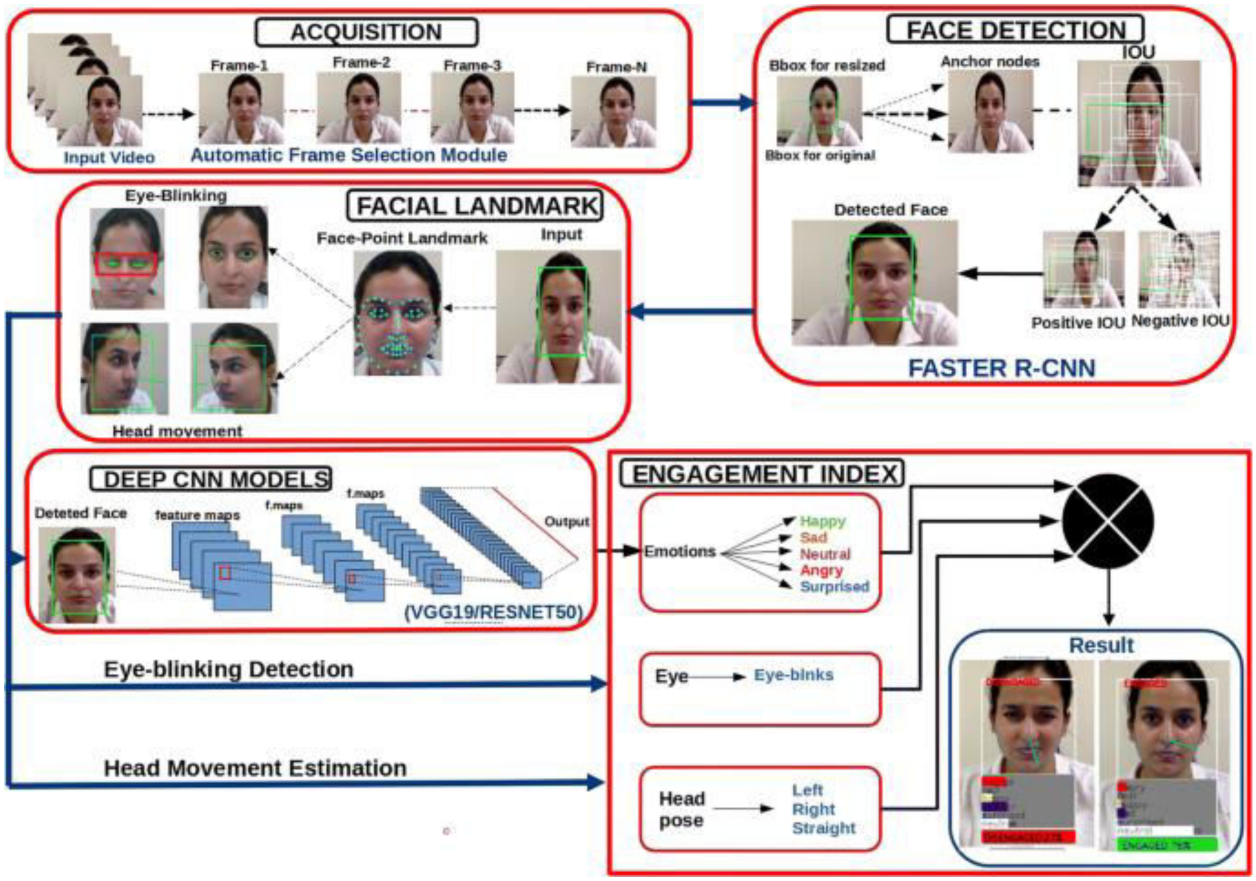


Fig 1: A multimodal facial cues based engagement

In a physical classroom, the instructor and students can send and receive non-verbal communication signals. Instructors who are aware of nonverbal signals sent in the classroom can modify their instruction in response to negative signals (confused looks) from students, as well as send positive nonverbal signals that may aid in student learning. These signals are unfortunately absent in a virtual classroom, which makes it very difficult to gauge students' engagement and boredom levels. Thus, a very informative lecture can completely miss the mark and become ineffective due to the inability of the platform to provide lecturers with students' non-verbal communication signals. For successful online teaching and learning, it is imperative to overcome this drawback. Thus, there is a need for a system that can operate within an e-learning environment and provide real-time feedback to lecturers on their students' levels of engagement in a virtual classroom.

Within the world of educational technology, empathetic interactions between the user and the system are very important. This can be achieved through the use of multiple channels such as text, audio, and visual modalities. This paper proposes and discusses the development of a non-intrusive model that can assess the engagement level and emotional state of the learner and then generate appropriate feedback. In the proposed study, a deep learning-based facial emotion recognition (FER) system, in particular a temporal relational network (TRN) FER system, is used to predict the expressions exhibited by a learner as recorded via a webcam. Furthermore, the predicted expressions from the system are used to estimate the emotional state and engagement level of the learner which allows for generation and distribution of appropriate feedback. For that reason, it is believed that the integration of such a system into the online learning environment can improve the overall platform. Figure 1 illustrates an autonomous e-learning evaluation system based on emotion recognition that can offer lecturers real-time feedback on their students' levels of involvement in a virtual classroom.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204118 |

II.RELATED WORK

Recently, studies have suggested that brain function is affected when basic emotional mechanisms are missing. Thus, emotions are needed for knowledge production. However, it is also indicated that extremely high emotional reactions have an adverse effect on rational thinking [7]. It reasonably follows that positive emotions such as joy, acceptance, satisfaction, and trust can lead to higher levels of creativity and accurate decision-making and problem-solving skills that can enhance learning [8]. Negative emotions such as fear, anger, and sadness have a negative influence on the brain and prolonged emotional distress can adversely affect the learning process and can be demotivating to a learner. For example, a learners' ability to memorize and remember can be affected by depression and anxiety which can lead to frustration and despair which is ultimately expressed in emotional forms such as anger, fear, and/or sadness [9]. In these scenarios, good feedback practices can help get learners back on track as well as motivate them, which can ultimately lead to enhanced learning. Thus, in an e-learning situation, an effective system should be able to read the learners' emotions and measure their attention in order to provide intelligent feedback that will enhance the learning experience of the learners. Effective and intelligent feedback in e-learning can be given to learners using embodied conversational agents (ECAs) [10]. These ECAs can be given a digital persona and can communicate with the user verbally or non-verbally. They can also express emotions that are more effective than a "faceless" computer providing feedback. Unfortunately, building these systems can be quite challenging.
When working with emotion recognition, it can be challenging to map emotional states according to facial expressions. To counteract this, Paul Ekman set out to map universal facial expressions [11] for emotions such as disgust, fear, anger, sadness, surprise, and happiness. Furthermore, this field gained more interest in the late 1990s when affect was successfully recognized by a machine from static images as well as from audio-visual signals. Literature suggests that emotion data should be drawn from observing the entire face and specifically noting the use of certain facial muscles. This is called the sign-judgment approach [12]. Additionally, the Facial Action Coding System (FACS) can be used to classify the action units (AUs) of facial expressions, which can then be labeled according to emotion [13]. Bartlett et al. [14] made use of AUs in their studies and found it to be a robust detection system to accurately label emotions. Other models, such as the geometric feature-based model, trace the variations in the shape and size of facial components (mouth, eyes, and eyebrows) to identify emotions. On the other hand, holistic approaches make use of a variety of machine learning approaches to mine facial features to ascertain emotions. Thus, making use of already available applications, it seems possible to identify emotions based on facial expressions.Employs FACS to differentiate between six basic emotional states. The accuracy of this application is 89%. Happy also reported a substantial accuracy in their work on facial expression recognition by employing local features on a person-specific dataset. It has also been indicated that facial expressions can hint at whether a person is feeling bored or tired; however, this has not been researched in depth as of yet.

Furthermore, posture and gestures can also provide insight into a person's affective state, as well as a level of attention and interest. This means that data based on gestures and body posture can be used to measure the cognitive state of the learner. However, these aspects have not been researched in depth. These aspects can be combined with an affective tutoring system that predominantly relies on identifying emotional states for appropriate feedback during the learning process. In the authors proposed a preliminary architecture that makes use of both emotional responsiveness and personality for the virtual tutor. Claims have been made by that a computer-based tutor should be able to detect and analyze levels of motivation, confidence, boredom, frustration, and fatigue in order to provide relevant feedback for each of these states. Other researchers [3, 4] have looked into detecting basic emotions in an e-learning situation. They made use of ECAs to accomplish parallel empathy and thereafter reactive empathy that expresses emotion through voice and articulation.

III.METHODS

The objective of our FER system was to predict the engagement level of learners using relational reasoning and a deep learning approach. Figure 3 illustrates the system architecture of the TRN-based FER system with adaptive mapping module. The main stages of the system include (i) face detection and extraction of representative frames, (ii) emotion recognition using a pretrained TRN model, (iii) adaptive mapping to estimate learning affect, and (iv) feedback generation.
Face detection and extraction of representative frames steps are required to align and normalize the input samples, allowing the deep neural network to learn meaningful facial features. Aspects that are unrelated to facial expressions, such as the

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204118 |

background or the angle and pose of the head, are relatively similar and should be effectively managed for efficient and accurate estimation of labels.

After retrieving all frames from the 251 video segments in the DISFA + dataset, representative video frames are acquired and meta-files are generated. By expanding the amount of input attributes, underfitting and bias are avoided [51]. Due to minute differences in facial expression, it was determined that the transition between frames is difficult to discern. When multiple filters are used, little changes in the frame are lost, making it impossible to train the dataset on certain emotions [52]. To remedy this, the frames are decreased to emphasize the face exclusively. However, it demonstrates that increasing the number of input features can reduce the high bias. To tackle the model's underfitting issue, we used several training and validation samples. The video frame scale is also trimmed to . Due to the fall in pixel value, it is difficult to notice subtle changes in the frame for long-term videos.

IV.RESULT ANALYSIS

In this research, we offer a process able to recognize representative frames in videos utilizing a relational reasoning approach. Once a representative frame is found, the TRN-based facial emotion recognition technique described in Pise et al. [45] will be executed on the observed representative frames. The proposed scheme offers two advantages. First, the CPU time corresponding to the recognition component will be greatly decreased since only a few representative frames are considered. Second, representative frames are specified by the frames where the facial expressions change significantly. Thus, one can determine a representative frame by scanning the sequence for representative frames that have been modified by the expression frame. Figure 4 illustrates the whole block architecture for representative frame detection utilizing multi-scale TRN by probability score.

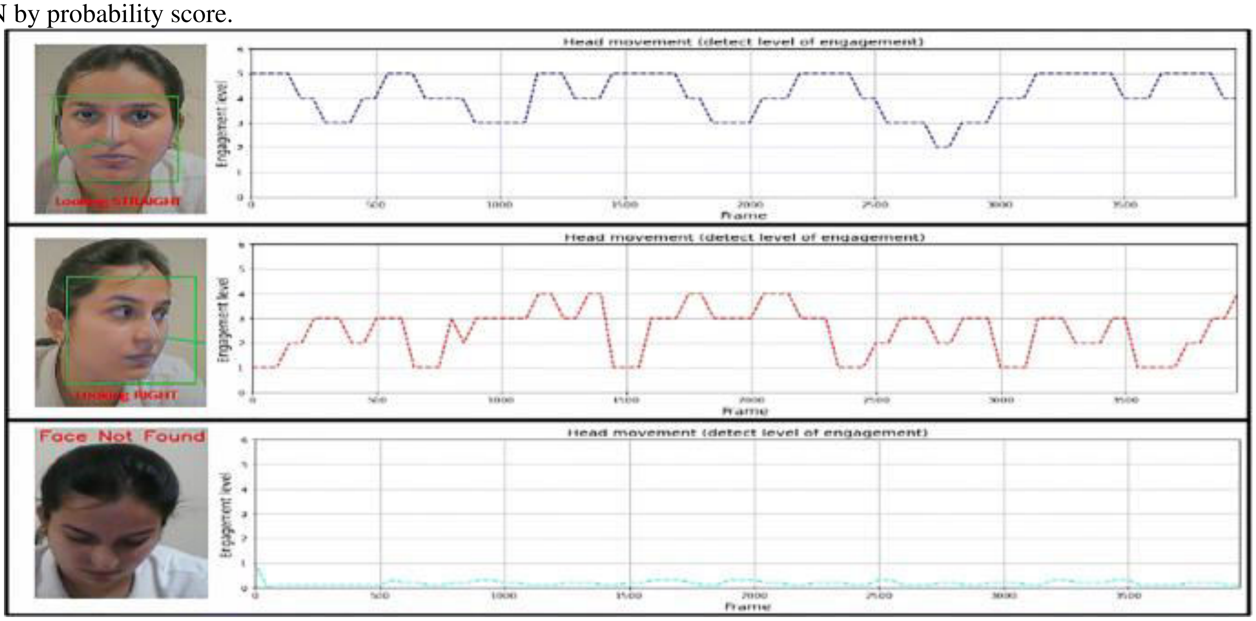


Fig 2: Result of multimodal facial cues based engagement

Users are unaware of the technical aspects of proposed architectures (multi-layer perceptron (MLP), SS-TRN, and MS-TRN), which should be enabled only by researchers. Following model activation and selection, we must press the START button to initiate the live video stream processing. When the user initiates a live test, the window displays the predicted emotion label. Each anticipated label is saved in a text file at the backend, together with the system timestamp and label. When the same emotion occurs more than once, the label automatically keeps it once until it disappears or the label category changes. The label category assigned to the frame that is considered representative changes. This indicates that among the series of these frames, the frame with the highest intensity of features capable of changing the expression label was identified as the representative frame. Also, the frames that represent the changes that took place in the sequence were considered as representative frames.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204118 |

V.CONCLUSION

The current work proposed an autonomous prediction model to measure learning affect and emotion depiction in an online learning environment that has the potential to be used in putting together tailored feedback to both learners and teachers. This was achieved through the use of a non-intrusive deep learning-based system that used visual cues obtained from participants. Based on participants' facial expressions, the system was capable of accurately predicting their cognitive and affective states. The proposed system can be integrated into an existing LMS. We propose that this stand-alone system should work with various LMSs, and we suspect that it will increase participant productivity.

REFERENCES

1. R. M. Bottino, "The evolution of ICT-based learning environments: which perspectives for the school of the future?" *British Journal of Educational Technology*, vol. 35, no. 5, pp. 553–567, 2004.
View at: Publisher Site | Google Scholar
2. S. Tsai and P. Machado, "E-learning basics: essay," *eLearn*, vol. 2002, no. 7, p. 3, 2002.
View at: Publisher Site | Google Scholar
3. C. N. Moridis and A. A. Economides, "Affective learning: empathetic agents with emotional facial and tone of voice expressions," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 260–272, 2012.
View at: Publisher Site | Google Scholar
4. S. Deniz, D. Lee, G. Kurian et al., "Computer vision for attendance and emotion analysis in school settings," in *Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0134–0139, Las Vegas, NV, USA, January 2019.
View at: Google Scholar
5. W. Horton, *E-Learning by Design*, John Wiley & Sons, Hoboken, NJ, USA, 2011.
6. S. M. Mavadati, P. Sanger, and M. H. Mahoor, "Extended DISFA dataset: investigating posed and spontaneous facial expressions," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2016*, pp. 1452–1459, IEEE Computer Society, Las Vegas, NV, USA, June 2016.
View at: Google Scholar
7. R. W. Picard, S. Papert, W. Bender et al., "Affective learning—a manifesto," *BT Technology Journal*, vol. 22, no. 4, pp. 253–269, 2004.
View at: Publisher Site | Google Scholar
8. M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, *Handbook of Emotions*, Guilford Press, New York, NY, USA, 2010.
9. C. N. Moridis and A. A. Economides, "Toward computer-aided affective learning systems: a literature review," *Journal of Educational Computing Research*, vol. 39, no. 4, pp. 313–337, 2008.
View at: Publisher Site | Google Scholar
10. S. ter Stal, L. L. Kramer, M. Tabak, H. op den Akker, and H. Hermens, "Design features of embodied conversational agents in ehealth: a literature review," *International Journal of Human-Computer Studies*, vol. 138, Article ID 102409, 2020.
View at: Google Scholar
11. J. B. Engelmann and M. Pogosyan, "Emotion perception across cultures: the role of cognitive mechanisms," *Frontiers in Psychology*, vol. 118, no. 4, 2013.
View at: Publisher Site | Google Scholar
12. S. L. Happy, A. Dasgupta, P. Patnaik, and A. Routray, "Automated alertness and emotion detection for empathic feedback during e-learning," in *Proceedings of the 2013 IEEE Fifth International Conference on Technology for Education (t4e 2013)*, pp. 47–50, IEEE, Kharagpur, India, December 2013.
View at: Google Scholar
13. T. Skiendziel, A. G. Rösch, and O. C. Schultheiss, "Assessing the convergent validity between the automated emotion recognition software noldus facereader 7 and facial action coding system scoring," *PLoS One*, vol. 14, no. 10, Article ID e0223905, 2019.
View at: Publisher Site | Google Scholar
14. M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 568–573, IEEE, San Diego, CA, USA, June 2005.
View at: Google Scholar

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE

SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com