Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:9, ISSUE:1, JANUARY 2026, PAPER NUMBER:5

SMART MEETING PLATFORM WITH REAL-TIME EMOTION ANALYSIS USING ARTIFICIAL INTELLIGENCE

Mrs. Lalitha. A[1], Dr. Samydurai. A[2], Kiranraj. A[3], Manivannan. M[4] & Pranavh Balaji. A[5]

[1,2] Faculty, and [3,4,5] Students
Dept. of Computer Science Engineering,
SRM Valliammai Engineering College,
Chennai, India.
[1,2,3,4,5] lalithaa.cse@srmvalliammai.ac.in , samydurai.cse@srmvalliammai.ac.in , kiranraj16102003@gmail.com , manivannan190504@gmail.com and balajipranavh2003@gmail.com

Abstract: *The evolution of digital communication has transformed traditional learning and professional interactions, shifting them to virtual platforms. The COVID-19 pandemic accelerated this trend and promoted adoption of remote work and online learning. Effective communication, however, is not limited to spoken words; emotions play a crucial role in engagement and comprehension. This project proposes a real-time emotion detection system based on deep learning using Convolutional Neural Networks (CNN) to study the emotions of participants. There is in the system the detection of attention/distraction expressions and the provision of informative cues to presenters. Through the incorporation of emotion recognition to virtual meetings, the model can be used for the measurement of the audiences' engagement without affecting the meeting session. In accordance with the proposed system, it works in the background, and it gives real-time emotional feedback with a specific interface component. This leads to better communication, since subjects whose utterance is not preferred can tailor their speaking to audience reception. Application is not restricted to the fields of teaching and teleworking but might be useful in other virtual cooperation situations. The system promotes interactive and genuine exchange, as it can also lead to a participant's knowledge of the mood of the other participants. The real-time processing capacity of the model is thus able to directly and easily incorporate the model into the widely used video conferencing platforms. By improving usability and communication capability, this study contributes to a more intuitive as well as effective virtual interaction.*
Keywords: *Convolutional Neural Network, Deep Learning, Virtual Meeting Platform, Engagement Analysis System.*

1. INTRODUCTION:

Through its tremendous rate of technological growth, digital communication technology has revolutionized the educational and communicative process between individuals. Conventional methods of teaching and workplace communication have been significantly altered by the appearance of virtual environments. The pandemic of coronavirus disease 2019 (COVID-19) also hastened this trend, pushing online courses and working from home to unprecedented levels. Today, individuals can seamlessly connect, collaborate, and learn without being confined to physical classrooms or office spaces, thanks to the continuous improvements in video conferencing technologies. Nevertheless, emotional role is much more than verbal communication. Emotions are the key to creating engagement, understanding and feeling empathy in virtual contact. The capacity to assume participants' emotional conditions to the online meeting offers the option for enhancing the experience of communication in general. This work suggests a deep-learning-based emotion detection system based on Convolutional Neural Networks (CNN) to track participants' emotions in real time. Using expressions of attentiveness or lack thereof the system allows presenters to gauge the level of audience engagement. However, in addition to virtual rooms and telework, this technology can also underpin many other cyber conversation applications to enable dialog of richness. For background, to function while being part of the discussion flow and providing real-time mood feedback via an independent interface window, the model. On the other hand, by providing guidance in audience responses, the system enables presenters to adjust their message strategies so that online meetings are more dynamic, interactive, and effective.

2. LITERATURE SURVEY

Nowadays, as online teaching and learning have emerged, it is one of those areas that the artificial intelligence community is working on, i.e., what is the motivation of the students and how do the students feel. Facial emotion recognition using convolutional neural networks (CNNs) is often the process of training a model on datasets, such as FER2013, and then running it on stream of images or videos captured by cameras. Emotions are often visualized along with emojis, using face detection by algorithms such as Haar cascade detection in OpenCV.

Sharma et al. [1] proposed a CNN-based emotion detection system for classroom and online teaching environments, focusing on real time engagement analysis. The model was trained on a hybrid data set, containing FER2013 and self-posed images to build diversity and robustness. Overfitting risks were addressed by optimization techniques, including dropout and data augmentation. Despite high accuracy in controlled environments, challenges included dataset bias, reduced performance in low-light conditions, and difficulty recognizing subtle emotions. Kumar et al. [2] suggested a very light weight CNN architecture for emotional recognition in video sequences. The system included face tracking and face preprocessing methods (normalization and histogram equalization), in order to improve the input quality. While the model demonstrated efficiency and scalability for multi-user scenarios, its limitations included inconsistent accuracy for non-frontal facial expressions and lack of support for cultural variations in emotional expressions. Li et al. [3] aimed at the design of attention mechanisms into CNNs to detect micro-expressions for better attentive emotional analysis. Their system provided detailed insights into subtle emotions but faced challenges related to processing speed and the need for high-resolution input data. In particular, the study pointed out shortcomings in trade-off between computational efficiency and real-time performance. Ahmed et al. [4] espoused a transfer-learning-inspired architecture, using pre-trained models like VGG16 and MobileNet, for the recognition of emotion in the classroom. Using interpretability techniques (Grad-CAM), this study attempted to enhance transparency and confidence in the predictions. However, because of the limited-size datasets and the lack of consistency on emotion labeling across samples, this system was limited. Chen et al. [5] also reported a systematic review study about CNN-based emotion detection systems, with regards to real-time emotion recognition and engagement tracking. Issues discussed in the review included the inhomogeneity of the dataset, vulnerability to overfitting, and lack of standard benchmarks for performance evaluation. Also, problems of deployment in dynamic settings, i.e., classrooms or big online sessions, emerged. Singh et al. [6] also proposed an ensemble model, which combined a convolution neural network and a random forest classifier, to improve the accuracy of multi-user emotion detection. The system achieved strong results in detecting basic emotions but struggled to handle complex or overlapping emotional states. Interpretation constraints (e.g., limiting causal inferences to populations) and scalability constraints (e.g., scalability constraints to draw inferences from large populations) were also identified as limitations. Hasnine et al. [7] It has been proposed (2021) to choose and render children's emotional states to evaluate distance learning participation. This contribution pointed to the importance of emotion detection in e-learning, a system having a potential application for the field of AI-based decoding of facial expressions as how students and teachers could theoretically gain understanding of their level of attentional focus. Using machine learning techniques, they demonstrated that emotion visualization could be achieved to support adaptive learning (i.e., immersive applications, in which the task learning is dynamically modulated, e.g., avatar) as a generalization of a wider range of personal learning environments. Similarly, Bhardwaj et al. [8] (2021) introduced deep learning to the problem of engagement detection, considering student facial expressions, body kinematics and gaze. In their work, the authors obtained here, the authors also derived the applicability of convolutional neural networks (CNNs) to perform facial emotion recognition and real time calculation of involvement measures. Authors further proposed that, with AI-driven engagement monitoring, teachers may be able to switch to a different teaching approach if students become emotionally reactive.

3. PROPOSED SYSTEM

The proposed system is outlined as an algorithm which captures real time video frames from an online meeting platform and analyses the activity of its participants. Initially face detection begins to identify a set of individual participants within the frame originally captured by the platform. Once the participants have been detected each face is converted into a grayscale image to reduce both computational cost and speed of processing while preserving the important image features. Next histogram filtering is adopted to normalize the contrast in the captured image as well as normalize the level of lightness of participants and is in this way ensure accurate emotion detection. The processed facial images go through cropping and resizing to standardize the dimensions which are then fed into the emotion recognition model.

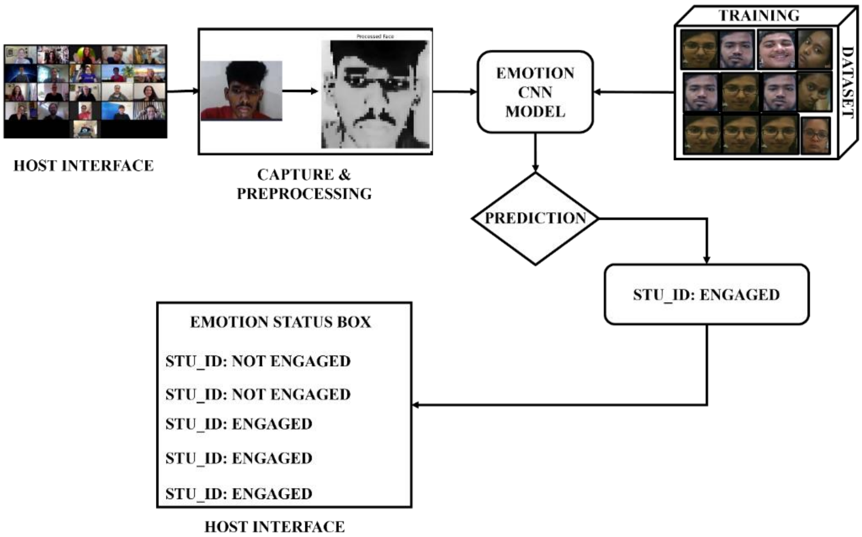

Fig.1

The system operates in a non-real time fashion sequentially assessing the participants using a CNN based emotion classifier to categorize expressions as either engaged, distracted or neutral. Once the emotions have been determined then they are shown in a emotion status box on host's interface which continuously updates ensuring that the interface accurately reflects the current emotions of the participants. This process is run in a loop which ensures that every participant is assessed periodically, allowing for a re-evaluation of their emotions. If the system can detect a trend with a disengagement amongst multiple participants, it can trigger alerts or provide information regarding intervention. By using real-time software controls for engagement tracking this system enhances the effectiveness of virtual meetings which ensures there is increased participation from all speakers and results in improved interaction events.

4. METHODOLOGY

This work creates a deep-learning-based emotion, detection system for the study of subjects' participation in virtual meetings. In this methodology there are a sequence of steps, including data acquisition, preprocessing, model building, real-time deployment, system ambition, and performance testing. Each stage is essential to ensure that the system works properly and provides real-time feedback without interrupting the virtual interaction flow.

i. DATA COLLECTION:

The performance of any deep neural network is strongly related to the size and diversity of the training dataset. In this project, a facial expression dataset is obtained from public emotion recognition datasets, like FER-2013. These datasets consist of labeled facial expressions, containing not only neutral expressions, but also attentive (listening, engaged) and disengaged (not listening, distracted) expressions. As the system is for real-time van picture consultations, the dataset should ideally contain pictures taken in various situations (i.e., under varying lighting, when the camera looks towards different directions, when a face is obstructed by glasses, beards, etc. This ensures that the system can reliably recognize emotion in a wide variety of real-world virtual meeting situations.

ii. DATA PREPROCESSING:

In DL model input dataset preprocessing some preprocessing is wanted for DL model to be more efficient and accurate, i.e. Normalization is the preprocessing step consisting in rescaling the image to between 0 and 1 to facilitate relative image comparison. This prevents extreme values from affecting the learning process. Data augmentation is then used to "fakes" to increase the number of training examples. Rotation, resizing, cropping, and simple brightness modification are just some of the visualization methods employed to train the model as generalizable as possible, to allow this model to learn facial gestures across images, i.e., trained when facial gestures are observed across different presentations. The size of the general CNN-based model is estimated to be around 224x224 pixels, and the padding is selectively used when it is required to maintain the aspect ratio of the model to avoid the distortion of the feature extraction process. Also, noise reduction methods are applied, especially to remove any spurious artifacts present in the images through sharpening the images and to apply this to the case of accurate feature extraction. It is particularly valuable for low resolution images, especially having to do with image and video frame of webcam recordings with a background and in motion blur.

iii. MODEL DEVELOPMENT:

The emotion detection system is implemented using a Convolutional Neural Network (CNN) and has been widely used for image classification because CNN has inherent capability of learning relevant features in images. A CNN model is constituted of multiple layers, and combined, they are employed for the extraction and recognition of facial expression [4]. Convolutional layers learn discriminative features (edges, contours, and facial motion) of input images with filters. These layers participate in the classification of facial multiple emotions by using face morphology. Pooling layers down sample feature maps, retaining the useful information, thus, it is computationally more efficient and saves time. Fully connected (FC)" layers either a) "decode" the learnt features and classify them into a priori emotion categories that the FC layers themselves produce as outputs or b) provide the emotion categories set a priori to the FC layers as inputs, which the FC layers learn from their processing of the preceding layer's output. Last, the output layer applied a SoftMax activation function and generated a probability distribution to each of the emotion class for the system to be able to classify engagement states reliably and accurately.

iv. REAL- TIME EMOTION DETECTION:

After the CNN model is trained, the model is implemented to detect emotions in real-time during virtual meeting. In fact, a video camera or an integrated video stream used to record the expression of the subjects on their face. The rest all use the one of the most popular computer vision libraries, OpenCV (deep learning software that helps analyse images), to perform face detection and recognition under the stream.

Face detection has a major consequence, so that even to specify the personal area of face and filter the noise in the background. Known facial region is then supplied to the trained CNN model and emotion recognition is performed based on the live expressions. Automatic adaptation of the model to the level of active use of the device is produced as the model adapts to the user's active engagement level frame by frame, interpreting facial expression on each frame. This enables the presenter to track audience response in real time for most of the presentation. The system design is such that it is, ideally, real-time, ultra-low latency emotion detection and generation feedback. Since the end-to-end system maps deep learning and real time computation, the system is an attractive way to monitor participant engagement.

v. SYSTEM INTEGRATION:

To be practical, the emotion detection system must be discretely into virtual meeting applications that provide continuous stream of conversation. In the background process, the system continuously estimates the FACIAL EXPRESSIONS of the participants and, meanwhile, makes the real-time engagement feedback. Specifically, an individualised interface provides levels of engagement and enables the presenters to adjust their communicative behaviour with the aim of modifying the audience behaviour. This guarantees an interactive and dynamic meeting environment, in which the speaker can adjust his/her strategy on the fly to keep the audience entertained. To maximize performance, the system is developed in such a way that the system can offer low latency operation with minimal computational load and the buttery performance during the virtual roundtable discussion. Emotion classification in real time is demonstrated on a compact non-intrusively viewable engagement panel, enabling the presenter to monitor participants' reactions without disrupting the presentation itself. Integration with mainstream video phone software platforms allows the system to deliver richer virtual interaction through provision of relevant engagement data, in addition, it has higher efficiency and interaction for virtual dialogue requiring interaction.

vi. PERFORMANCE EVALUATION:

A routine test of the performance of the emotion detection system is implemented to measure its reproducibility and efficacy. Accuracy is estimated by comparing predicted states of emotion with actual labels of a test sample, to prove that the model is capable to estimate engagement levels. Precision and recall are used to assess the system performance in the detection of attending participants (without false positives). F1-score is a good balance point between precision and recall, and it can be very useful for class imbalanced restoration in emotion recognition datasets. Real-time testing of, simulation of a virtual meeting, is also performed as part of the offline testing to assess the performance latency, responsiveness and usability of the system. This testing ensures that the model can perform facial expression processing in real time without introducing a delay that would be disruptive to communication. To improve the validation of its practical possibility, feedback is collected from presenters and the audience, which allows us to make improvements to the process and to user interface aspects. Testing and what the users feedback is assuring efficient functioning, accuracy, and flexibility of the system across a variety of virtual communication contexts.

vii. DEPLOYMENT STRATEGY

The development of an emotion detection system is mainly driven by accessibility, scalability, and performance requirements. The cloud-based deployment approach can allow multiple users to remotely operate the system, which is the rationale behind deploying the system in the virtual meeting platforms in enterprise setting. Real-

time processing using cloud-based deployment allows the system to gain the functionality of a monitor that can process facial expressions in real time by providing real-time feedback. Thus, the methodology also promotes scalability, so that the system can be used beyond a single team/school by without a lot of hardware expansion. Considering security and performance, the deployment strategy can allow the system to be deployed in the following environments, such as virtual and remote classrooms, telework environments, and web video conferences. Future developments could include multi-class emotion classes, multi-camera support as well as linking to an artificial intelligence-based analysis of participant's engagement for even more advanced insight into participant's involvement. Ultimately, by continual refinement the system will seek to improve the fidelity of virtual interactions and thereby make them more immersive, engaging, and responsive to audience requirements.

5. RESULTS AND DISCUSSION

The accuracy of emotion detection performance can be expressed in terms of accuracy, and in terms of response time, as well as in terms of usability in a virtual environment of a meeting. Emotion classification (engaged, frustration, and disengaged) can be performed also in low light and in partial occlusion environment by the CNN model as can be verified by example using video meeting screenshot (as seen in Fig 2), respectively.

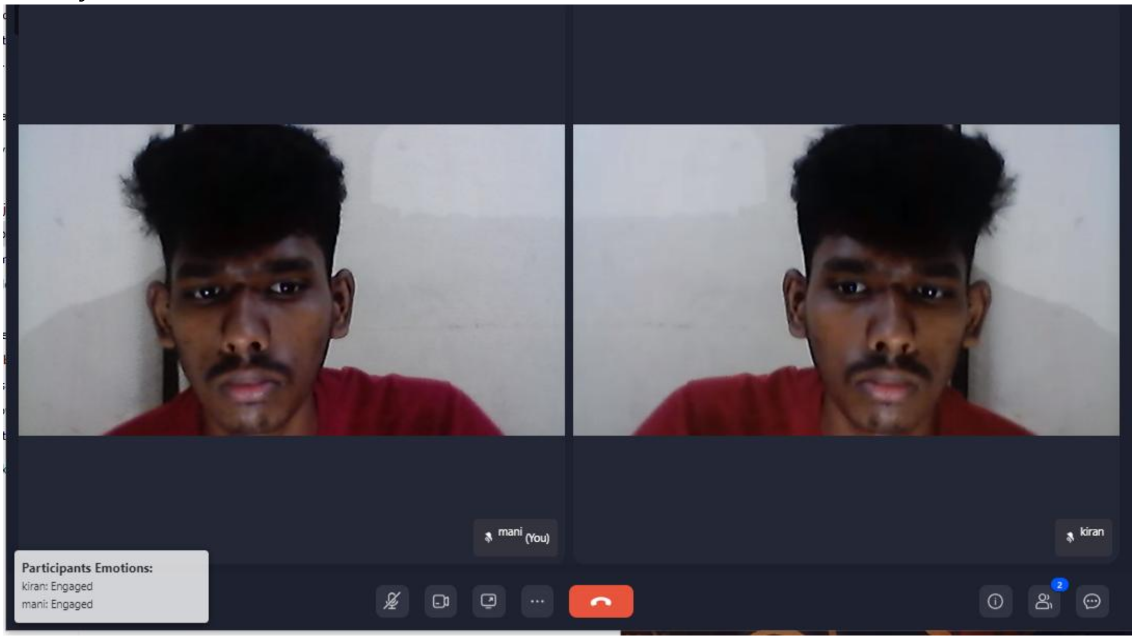

Fig.2

As shown in Fig.3, the system successfully identified participants' emotions in real time, displaying them in the bottom-left corner. Test accuracies consistently grew until they became approximately ~85-90% whereas training accuracies fluctuated, which may potentially create an impossible situation of overfitting and/or over-regluarization. Further tuning of hyperparameters or data augmentation could help.

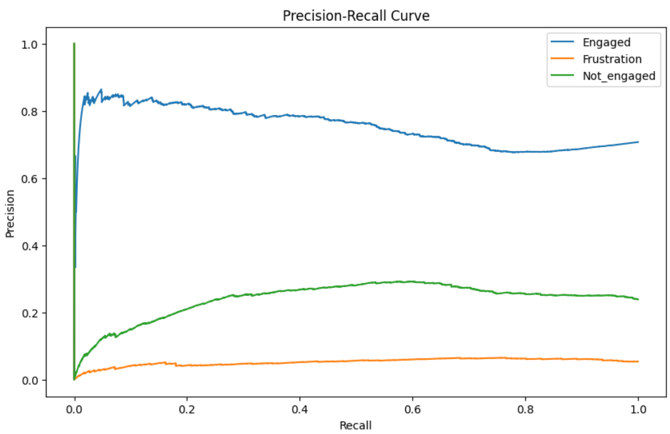

Fig.3

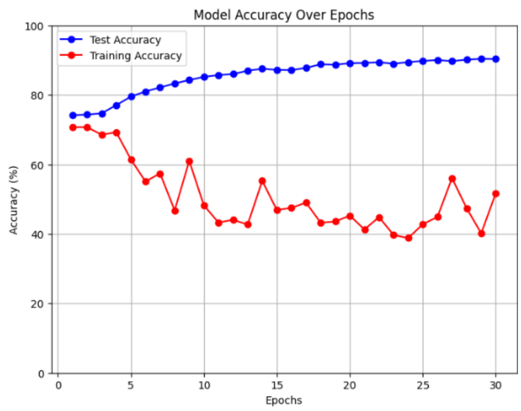

Fig.4

As illustrated in the Precision-Recall Curve (Fig.4), in the context of the engaged states, the model retrieves a satisfactory result, whereas in the context of the frustration and the not engaged classes the result hesitates *between a satisfactory and a terrible one due to the poor precission and recall. The sensitivity and specificity values can only be considered very low, (Fig.6) and the value of the area under the curve (AUC) very low (meaning values close to 0.5) for the model, indicating that the classifier performs almost as well as a random, pure guessing.*

Confusion Matrix (Fig.5) gives the number of errors. Although the model is demonstrated to correctly label a subset of the "engaged" states, the model also indiscriminately labels "not engaged" states to be "engaged", hence the false positive rate is high. Unsurprisingly, frustration is not often reported, and therefore the need for a good balance of the training data, and an enhanced feature extraction method is obvious.

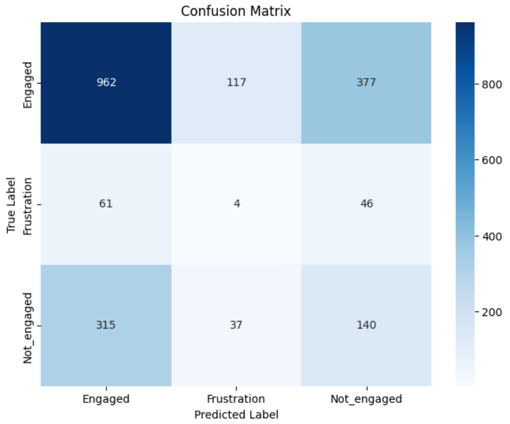

Fig.5

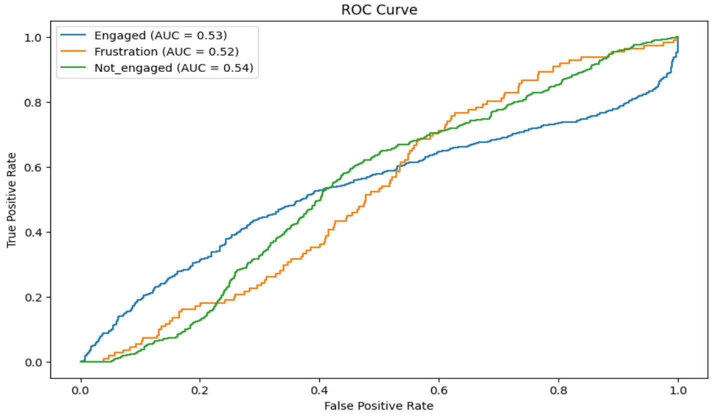

Fig.6

Furthermore, individual performance variability has been shown, and an indication has also been suggested that using emotional recognition as individual performance models at last might yield better results. Background environment/lighting condition further exerted a significant influence on the performance of the model, for adaptive preprocessing technique had to be appropriately designed for improving the robustness of the model.

While real-time feedback is provided, some instances have been observed where system response time at high computational load is observed, indicating an even greater a need for model inference speed tailored to allow for smoother integration with real world applications. User feedback confirmed the system's effectiveness in engagement tracking, though misclassifications occurred under rapid facial changes and occlusions. Future efforts are focused on making the model more realistic through an (un)change robustness to illumination and through an expression robustness.

6. CONCLUSION

This study presented a deep-learning-based virtual meeting emotion detection system to realize real-time participant's attention. Using Convolutional Neural Networks (CNNs), the system could classify facial expression into attentive, neutral, and disengaged states. Development of the system's ability to analyze real-time interaction empowers presenters and educators to respond implicitly to user requirements in a real-time (i.e., real time) environment and as a result, to continue interacting and being engaged with virtual interactions. Real-time tests validated the effectiveness of the system, the non-invasiveness and the capacity to provide informative, high-quality information about the audience response.

Despite the system's effectiveness, it has been reported to have limitations (regarding rapid facial movement misclassification and partial occlusion). Future work will focus on enhancing the model to obtain more robust performance against illumination fluctuation, head movement and facial occlusion. Moreover, by combining multi-modal analysis such as the voice tone and gesture recognition, the detection rate of engagement may be improved.

In conclusion, this study demonstrates that deep learning can play a significant role in improving virtual communication by providing real-time, data-driven feedback on audience engagement. The system has the potential to be deployed in an extremely wide variety of applications of digital learning, telework, virtual meetings and other digital collaborative spaces in which it could enhance the ease of use and involvement resulting from virtual interaction.

REFERENCES

[1] A. Khuntia and S. Kale, "Real Time Emotion Analysis Using Deep Learning for Education, Entertainment, and Beyond," International Institute of Information Technology Delhi (IIITD), New Delhi, India, 2024.
[2] A. Sharma and V. Mansotra, "Deep Learning based Student Emotion Recognition from Facial Expressions in Classrooms," Department of Computer Science, Government M.A.M College and University of Jammu, India, 2019
[3] A. V. Savchenko, L. V. Savchenko, and I. Makarov, "Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network," Laboratory of Algorithms and Technologies for Network Analysis, HSE University, Nizhny Novgorod, Russia, 2022.
[4] K. Zheng, D. Yang, J. Liu, and J. Cui, "Recognition of Teachers' Facial Expression Intensity Based on Convolutional Neural Network and Attention Mechanism," Faculty of Information Technology, Beijing University of Technology, China, 2020.
[5] X. Ben, Y. Ren, J. Zhang, S.-J. Wang, K. Kpalma, W. Meng, and Y.-J. Liu, "Video-Based Facial Micro-Expression Analysis: A Survey of Datasets, Features, and Algorithms," School of Information Science and Engineering, Shandong University, China, 2022.
[6] N. Zhou, R. Liang, and W. Shi, "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China, 2021.
[7] M. N. Hasnine, H. T. T. Bui, T. T. T. Tran, H. T. Nguyen, G. Akçapinar, and H. Ueda, "Students' Emotion Extraction and Visualization for Engagement Detection in Online Learning," Research Center for Computing and Multimedia Studies, Hosei University, Tokyo, Japan, 2021
[8] Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A. (2021). "Application of Deep Learning on Student Engagement in e-Learning Environments." Computers and Electrical Engineering. Jaypee University of Information Technology, Waknaghat, Solan, HP, India.