Archives available at journals.mriindia.com

Open Access International Journal of Science and Engineering

ISSN: 2456-3293

Volume 9 Issue 05, 2026

Deep Learning–Based Emotion Recognition for Monitoring Mental Health in E-Learning Platforms

Vinayak S. Mane
Department of Computer Science and Application, Sadashivrao Mandlik Mahavidyalay, Murgud, Kolhapur, Maharashtra, India
Email: vinayakmane003@gmail.com

Peer Review Information

Submission: 10 April 2026

Revision: 26 April 2026

Acceptance: 05 May 2026

Keywords

Deep Learning, Emotion Recognition, Mental Health Monitoring, E-Learning Platforms, Affective Computing, Educational Technology

Abstract

The intensive development of e-learning has heightened issues surrounding the emotional stability and mental stability of the learners, since there is less interaction between the teacher and the learner, the teacher is unable to detect the affective responses that may be signs of distress, disengagement, and anxiety. Reactionary to this, deep learning-based emotion recognition has also become a potential solution to ongoing tracking of the emotional condition of learners and assisting mental health in an online educational setting. This review is a critical investigation of peer-reviewed literature published within the last five years on deep learning techniques of emotion recognition and how they can be used in mental health monitoring on e-learning platforms in institutions of higher learning, at K-12, and corporate learning scenarios. The review summarizes the progress of convolutional neural networks, recurrent networks, transformer models and multimodal learning models, which combine facial expressions, speech, text, and physiological cues. In addition to the technical performance, the review assesses the use of these systems to deduce engagement, stress, anxiety, and depressive tendencies and how the insights can inform adaptive instruction, early intervention, and learner support. Although recent research findings have shown significant improvements in the level of recognition and real-time viability, there are still critical issues concerning the generalizability, bias, interpretability, privacy, and ethical implementation. The trends mentioned in the review include multimodal fusion, explainable and privacy-preserving learning, and longitudinal affect tracking, with a focus on human-centered design and interdisciplinary collaboration. In general, the discussion indicates that emotion recognition based on deep learning could be used responsibly and beneficially in the context of mental health awareness in e-learning to act as a supplementary component and enhancement of human judgment and care, but not a substitute.

Introduction

Deep learning is transforming emotion recognition in education, enabling continuous monitoring of learners' affective states as a window into their mental health. E-learning platforms have been particularly rampant over the past few years, particularly after the pandemic, yet that growth has been met with an increase in the awareness of the student well-being and psychological strain in online settings

(Khan et al., 2025). Teachers and scholars acknowledge that emotions such as anxiety, confusion, or engagement are central to the outcomes of the learning process and that they can indicate some mental problems (Wang and Dong, 2025). Old techniques of testing the emotions of students, including self- report surveys or teacher observations, are subjective and practical and not always feasible in large virtual classes. It has led to interest in emotion recognition based on deep learning being automated and capable of objectively determining the emotional state of students in real time (Khare et al., 2024). Using artificial intelligence to read faces, voice intonation, written message, and even body language, e-learning platforms have the potential of detecting students who are struggling and offer help before matters get out of control. This critical review aims to summarize the current trends (mostly within the past five years) in deep learning-based emotion recognition and discuss the ways in which these technologies are currently being implemented to track and help regulate the mental health of online learners. We compare the state-of-the-art models, modalities and deployment strategies, and the new outcomes, challenges, and ethical implications of the developments. Instead of just listing algorithms, we adopt a critical approach to the effectiveness with which existing systems identify the appropriate emotional indicators of mental health and how these can be enhanced or used responsibly in education to the advantage of the student.

Emotion Recognition, Mental Health, and Learning:

The rationale behind the connection between the emotional statuses of students and their academic outcomes and mental well-being is solid theoretical and empirical. Positive emotions (e.g. enjoyment, hope) are more likely to increase the cognitive resources and are associated with the improved engagement and learning results in educational psychology, and negative emotions (e.g. frustration, anxiety) may hinder learning and can be a sign of stress or burnout (Walsh, 2025). Unanswered emotional turmoil during the learning process can undermine motivation and lead to issues with mental health depression or persistent anxiety (Wang & Dong, 2025). On the other hand, strong emotional strength and capacity to sustain a favorable affective balance are associated with better well-being (Shehada et al., 2025). These facts explain the importance of tracking the emotional patterns of students: the initial signs of disengagement, confusion, and chronic sadness in an online student may indicate the necessity to intervene (Masud et al., 2025). In virtual classes, however, the same face-to-face signs of distress will not be available to the instructors. This has stimulated the creation of affective computing technology that can automatically recognize moods using computer signals, which in turn serve as a mental health sensor of remote learners (Belludi & Kopackova, 2025). With constant monitoring of student feelings, not only academic results, but also with this, the educator will be able to learn about the needs of each of the learners and help them in a more individualized way. Notably, emotion recognition can not be considered an independent mental health diagnosis since, as Na et al. (2024) note, facial expressions or other indicators only capture a specific part of the psychological condition and should be used in conjunction with other tests. Nevertheless, the established recognition of emotions can be seen as an effective building block in a more extended system that would promote well-being within the e-learning settings, allowing learning processes to be more responsive and emphatic. Altogether, the deep-learning-affective-science convergence can be deemed as the promise to close the empathy gap in online education and help uncover hidden struggles, facilitate mental health and academic success.

Deep Learning based Emotion Recognition: The current generation of emotion recognition systems relies on deep-learning based models with the ability to extract hidden patterns in complex data, such as images, audio, and text. Facial expression recognition is a predominant task in this area, in which convolutional neural networks (CNNs) are used to analyze face images to determine emotions like happiness, sadness, anger, fear, surprise, or neutrality (Na et al., 2024). Facial emotion recognition (FER) based on CNN has been demonstrated to be highly accurate in controlled settings; recently, transfer learning using state-of-the-art CNN models can reach more than 90-95% accuracy on face expression benchmarks (Aly, 2025). New advances comprise attention systems to target significant body parts of the faces and features, which consequently increases micro-expression and subtle indications detection. Making a CNN employ a channel-attention module, Wang and Dong (2025) proved that the recognition rates of children with fleeting expressions increased by more than 14 percent, resulting in the conviction of 86.5 and more advanced cross-age generalization rate. In addition to the stationary images, deep learning models also utilize the time component of video. The archiving of expressions with time combined with CNN-RNN or 3D CNNs helps differentiate between actual emotional response and brief movements of the facial muscles (Lu et al., 2025). In one study, a single framework was offered, which involved a ResNet-50 CNN of their spatial features, a 3D CNN, and temporal modules, to track the sequence of facial expressions of students interacting through online lessons, allowing them to detect a change in engagement or confusion on a moment-to-moment basis.

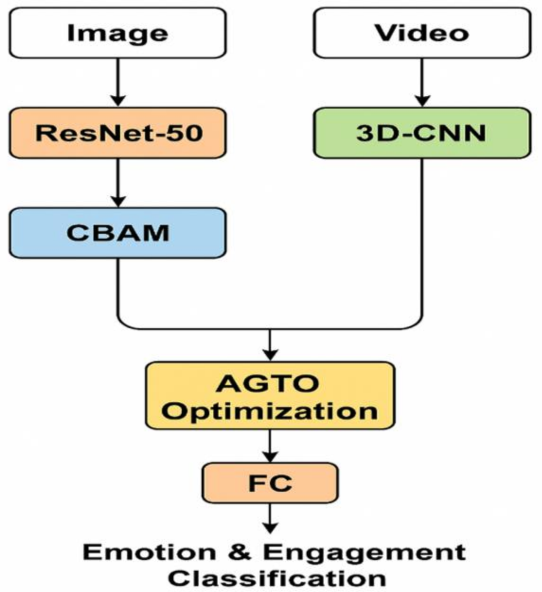

Figure 1: A deep learning framework combining spatial (image-based) and temporal (video-based) facial analysis for emotion and engagement recognition.

This hybrid architecture uses a ResNet+CBAM branch to extract features from static facial frames and a parallel 3D-CNN branch to capture expression dynamics from short video clips. Feature fusion and optimization (e.g., via genetic algorithms) then yield an integrated representation, which is fed into fully-connected layers to classify the learner's emotional state and engagement level. Such multi-branch models can more accurately identify complex states like sustained frustration or attentiveness by considering both immediate facial cues and their temporal context (Aly, 2025). Besides vision, speech is also an abundant source of emotional information: an applied deep learning model (e.g. LSTM or Transformer networks) on audio can help determine affective tone based on such vocal characteristics as pitch, intensity, and rhythm. Jordan et al. (2025) conducted a systematic review and discovered that speech emotion recognition (SER) has been demonstrated to be potentially effective in mental health assessment, with AI-based technology assessing voice recordings to identify depression, anxiety, or even suicidal intentions. An example is that multiple studies have obtained significant levels of accuracy in detecting a depressed and a non-depressed person using speech models, which implies that voice is a promising modality in mental condition monitoring (Jordan et al., 2025). In teaching, SER would be applicable in the verbal answer or conversation of pupils to measure the levels of frustration or excitement, which supplements the visual clues. Natural language processing can also be used to analyze text data, including the posts of students in a forum or chat, or their written comments. Language models that are trained on sentiment analysis (e.g. BERT, RoBERTa) can be fine-tuned with a transformer on sentiments, or can identify a particular mental health condition (Khan et al., 2025). In one recent study, a RoBERTa-Large model correctly predicted the self-reported mental health status of students, with 97% accuracy, by analyzing their language on their social media and forums, essentially detecting those who were at risk of developing problems such as burnout or loneliness based on language (Khan et al., 2025). The findings illustrate the effectiveness of deep learning to interpret subtle indicators of affect in text that a person may overlook, including variation in sentiment polarity, negative emotive word use, or disengagement in communication. Additionally, multi-modal methods are becoming increasingly popular: a combination of facial, vocal, textual, and even physiological cues can be used to give a more complete and accurate assessment of emotion than the one provided by a particular channel (Khare et al., 2024). Multimodal deep learning Multimodal deep learning models have been designed to combine features across sources - e.g. Lu et al. (2025) used a hybrid CNNLSTM architecture to combine EEG brainwave signals, heart rate (ECG) data, and facial videos. This would be able to measure internal physiological arousal and external manifestations at once with high accuracy in detecting when the students were stressed or in complete concentration in case of competitive activities. The approach to this type of system depicts the direction toward the holistic approach to affect detection in which a set of sensors and deep networks collaborates to predict an emotional state of a learner in many different dimensions. To a recurring theme, CNNs (and their adaptations) are used as feature extractors and recurrent networks (such as LSTMs or GRUs) or attention-based Transformers as temporal dependencies or sequence context (Wang et al., 2023). CNNs in particular have become notably salient especially to images, or even to one-dimensional signals or spectrograms, thanks to being able to learn salient features automatically; indeed, a general survey of deep learning in mental health revealed CNNs to be the most common architecture to be used to perform tasks such as mood detection (Arji et al., 2023). Nevertheless, scholars are looking into more sophisticated methods: attention modules serve to narrow in on salient moments or regions (as in facial focus on the eyes/mouth, or self-focus on the speech to realize expressions of emotional salience utterances), and explainable artificial intelligence methods are being integrated to understand what these complex models are learning (Shehada et al., 2025). As an example, the use of gradient-based saliency maps (e.g. Grad-CAM) to show what parts of the face a CNN relied on to decide that a student was "frustrated" has been demonstrated to improve visibility among educators and developers. Overall, the emotional recognition deep learning toolbox in e-learning is abundant and constantly developing - old-fashioned CNNs and hybrid or transformer models all aim to the more accurate, subtle, and real-time perception of the emotions of the learners.

Mental Health Monitoring in E-Learning: The end goal of these technological innovations is to have responsive learning environments that can be used to help students with their mental health. One of the main uses is real time engagement monitoring in online courses. Deep learning models can also be used to give instructors live measurements of the mood and the level of engagement of the classroom by analyzing the video feeds of the students on their webcams (Bhardwaj et al., 2021). As an example, a digital platform to study could have an engagement dashboard based on facial expression, that is, a teacher can see that a certain student does not seem to be engaged or is unhappy in the lesson (Bhardwaj et al., 2021). Bhardwaj et al. (2021) provided an example of such a system, where students were classified as either engaged or not engaged by the use of facial expression and eye gaze (a CNN-based method was used), and timely interventions were given to the student (e.g., calling a distracted student, or providing assistance to the frustrated learner). Expanding this idea, Aly (2025) created a more sophisticated platform which would capture the facial image of students periodically and run through a ResNet50-based deep model and be able to track the progress of emotions in a student; a system with an accuracy of over 95 percent in identifying the main expressions and could constantly record the mood of each learner during a session. These tools are not only able to detect negative conditions such as confusion or boredom that can impede the learning process, but they can equally record positive engagement to present the students with feedback once they are extremely focused on what they are doing or when they like the material. By so doing, emotion recognition will make it possible to have a kind of affective learning analytics that will complement the traditional performance metrics with information about the feelings of students, an equally significant part of successful learning. Along with overall engagement, there are more mental health-specific use cases that are emerging. Other scholars have aimed at identifying academic stress and anxiety in students using physiological and behavioral indicators. In situations of e-learning with a high degree of stress (where you have a timed quiz or a competition on a team) the authors Lu et al. (2025) suggest that a combination of facial strain indicators and biosignals will be able to quantify the level of stress or lack of concentration in the student, and their hybrid model was actually able to detect when subjects experienced stress or lacked concentration, which could be mitigated by stress-reduction help or encouragement. Other aims of work focus on preventing severe mental health risks. To illustrate, Masud et al. (2025) trained text and survey data with deep learning models to detect the symptoms of clinical depression in university students. These models can be included in the e-counseling systems of universities or even learning management forums to identify the students whose language and activity patterns are typical of at-risk individuals (Masud et al., 2025). Likewise, message and post sentiment analysis may indicate that a student is becoming frustrated or desperate - minor shifts that can be detected by the automated algorithms before a human advisor. At a less urgent scale, it is utilizing emotion recognition to customize learning and enhance user experience. When a system is aware that a student is lost or dissatisfied with what he is learning at the moment, it might automatically serve to change the level of difficulty, to offer further clarifications, or to initiate a chat-bot that will be supportive to the student (Aly, 2025; Belludi and Kopackova, 2025). Cases of intelligent tutoring systems which utilize facial emotion input to adjust their teaching approach on-the-fly - such as slowing down to give hints when the learner displays frustration to avoid the negative reactions of negative emotions rising (Khare et al., 2024). Aggregated affective data in the group learning context, such as an online collaborative project, can be used to assist moderators to maintain a healthy group communication; once a student has been identified as either perpetually angry or bored, conflict management or motivation techniques can be employed accordingly. In Table 1, some of the representative recent studies are captured that depict all these diverse applications and their results in contexts of e-learning.

Table 1: Key recent studies on deep learning-based emotion recognition in e-learning and mental health.

Study (Year)	**Context/Participants**	**Modality & Deep Learning Approach**	**Key Findings / Outcomes**
Bhardwaj et al. (2021)	University online classes (India)	Facial expressions via CNN model (webcam video)	Real-time detection of student **engagement**; system flagged inattention reliably, allowing timely instructor intervention (engagement classification accuracy ~92%).
Aly (2025)	Virtual classroom platform (Higher Ed)	ResNet-50 + CBAM network with temporal CNN (facial video)	Achieved **95–98% accuracy** on standard FER datasets; monitors students' moment-to-moment emotions and engagement, enabling adaptive feedback in live online classes.
Wang & Dong (2025)	Primary school classroom (ages 9–12)	Attention-enhanced CNN for facial micro-expression	**86.5% accuracy** in recognizing children's emotions; improved detection of subtle distress signals and cross-age generalization, aiding mental health monitoring for young learners.
Lu et al. (2025)	Competitive e-learning tasks (college)	**Multimodal** CNN–LSTM (EEG, ECG, and facial video)	Integrated physiological and facial cues to identify **stress vs. focus** states; demonstrated high accuracy and provided insights to reduce student anxiety in high-stakes scenarios.
Khan et al. (2025)	Student text data (social media/forums)	RoBERTa-Large transformer for sentiment analysis (text)	**97% accuracy** in classifying mental health status from language use; effectively predicted students experiencing depression or anxiety based on negative sentiment patterns.
Shehada et al. (2025)	Assistive mental health app (adults)	Lightweight CNN for facial emotion + **Federated Learning**	~75% cross-dataset accuracy on FER; introduced privacy-preserving distributed training and **XAI** diagnostics, building trust for deployment in mental health monitoring.
Masud et al. (2025)	University students (Bangladesh)	Ensemble of ML and deep NLP models (survey & text data)	Detected **depression** with 91% accuracy; incorporated explainable AI (SHAP/LIME) to identify key features (e.g., isolation keywords), facilitating understanding and early counseling outreach.

These instances highlight the ways of how emotion recognition technologies are being customized to different educational settings, such as K-12 classrooms to college and more, to support mental wellness. They demonstrate that deep learning is capable of consistently identifying such directly relevant to learning and mental health states as disengagement, confusion, or stress. Notably, not only are a variety of systems monitored but also, the feedback loop is closed: an emotion-aware tutoring system could pause automatically or offer support to a student when it detects that he or she is frustrated (Aly, 2025), avoiding the emergence of negative emotional cascades. Similarly, aggregated data on student emotions in classes can assist instructors to self-reflect and to make changes to their course structure (e.g., recognize lectures that always bored or overwhelmed students). Such interventions eventually are likely to create more positive emotional learning experiences, which are associated with improved mental health outcomes such as lowered anxiety and increased self-efficacy. There is an initial indication that students are responsive to empathetic learning conditions--in a study it was reported that there was a higher level of satisfaction and decreased stress when an online platform reacted dynamically to emotional expression of learners (Khare et al., 2024). However, it has not yet been implemented in practice in the educational domain, where it is usually a pilot study or controlled experiment.

Difficulties and Moral Implications

Although there is enough technical advancement, implementation of deep learning- based emotion recognition into e-learning in the context of mental health creates serious challenges. Generalizability and accuracy still are the major issues. Depending on the person and culture, emotional expressions can differ significantly; what one demographic is used to perceive as a certain signal can be misunderstood by another demographic with an algorithm-trained (Wang et al., 2023). E.g. a model may confuse a culturally neutral expression of a student as negative just due to bias in the training data. In addition, students may occasionally conceal their emotions because they are either intentional (to prevent embarrassment) or accidental (because of camera off or low light) which results in distress being unidentified. Even non-physiological cues such as facial expressions and tone are always indirect proxies of internal states; it is a well-known fact, as Walsh (2025) reminds, that a smiling face does not necessarily translate to a happy brain. Excess dependence on such cues may produce a false reassurance or false alarms. This is the reason why other researchers suggest using physiological indicators (heart rate variability, skin conductance, EEG) that are more difficult to control and can demonstrate undiscovered anxiety that facial or vocal analysis fails (Wang et al., 2023). In fact multimodal systems containing these signals are more resistant to surface-level variability, though at the expense of using more sensors. The other technical challenge is the provision of real time performance and scalability. It can be computationally demanding to run deep neural networks on video streams in real-time over a span of an entire class. Although newer models such as MobileNet or efficient variants of CNNs can be run close to real-time on consumer hardware, it is difficult to be responsive with dozens or hundreds of concurrent feeds of students (Trabelsi et al., 2023). Emotion recognition accuracy could further be impaired by network bandwidth and video quality problems in the home setup. In this way, intelligent tradeoffs between sampling rate and the complexity of model (e.g. the analysis of one frame/s instead of full video) will have to be made to allow the practical systems to be operated in live e-learning without delays or failures. The ethical and privacy issues are even more marked on the ethical and privacy front. The question of privacy is of primary concern: constant video or audio surveillance of students may be perceived as an invasive form of surveillance without attention to the matter (Shehada et al., 2025). Students and parents might be concerned about the purposes of the emotional data and accessibility, as well as the possibility of its misinterpretation and even punitive use. Data protection cannot be based on strict methods, including on-device processing or federated learning that does not concentrate raw sensitive information (Shehada et al., 2025). E.g., a federated learning structure, in which the emotion model is trained on devices of students, without uploading individual videos to a server, was implemented by Shehada et al. (2025), which eliminates the threat of privacy violations. Despite all these measures, transparency and consent are essential: a student has to be informed and preferably given a choice whether their feelings will be monitored. Also a risk is stigma or self-fulfilling prophecy in case a system refers to a student as being at risk. False positives would result in unwarranted alarm or labeling, but false negatives may result in an opportunity to help being missed. In this way, such AI tools are not to substitute human judgment but to complement them, and a counselor or an educator will have to be in the loop to confirm and put the tools into context. Favouritism and equality are other ethical issues. Similar to most AI systems, emotion recognition AI can also be biased by training data. As an example, not all facial expressions datasets are diverse, and the resulting models can be more effective with specific groups of people or age intervals (Walsh, 2025). When applied blindly, these biased models could favor the mental health requirements of a certain group of students compared to another. The solution to this is by treating training data (demonstrating demographic balance) and performing continuous bias checking and correction. Another useful method is explainable AI, which can be used to understand what information the model is using; in case an FER system were dominated by irrelevant features (such as a headscarf or glasses) then designers can re-train or re-configure the model accordingly. Interpretability of emotion AI is a technicality, but an ethical requirement, as to ensure that a system is used ethically, educators must be made aware of why the system is alerting them about a student, and students must be made aware of how they are being evaluated. An increase in explainability leads to trust and gives the opportunity to have a conversation instead of a mystical attitude or just blindly follow the decision made by the AI (Walsh, 2025). Within the educational setting, any intervention message driven by an AI notification must be soft and enabling and should be conveyed in the form of an offer to help and not as a punitive action. Lastly, there are pedagogical and social factors. Other teachers fear that, with automated emotion tracking, the role of the teacher might be marginalized or might be viewed to be too intrusive to an extent that it creates a surveillance-like environment that prevents interaction between teachers and students. It is necessary to stress that these tools do not substitute the emotional labor by teachers, but they are the supplements. But applied wiserly, it can give instructors more time to turn to empathy and support as they should identify students who would otherwise pass unnoticed through the cracks. The technology must be deployed in a clear policy: such as the use of data must be done to support the students and not to punish or grade a student. Even talking to students about these tools can be an educational experience, and it makes them aware of emotional intelligence and self-controlling (and in some systems, the students are even provided with the feedback on their engagement pattern, which encourages meta-cognition and healthy studying behaviors (Aly, 2025)). To conclude, the implementation of emotion recognition based on deep learning in e-learning is a delicate balancing act that needs to prioritize the best use of psychiatric value (early intervention, personalization, enhanced interaction) and the importance of mitigating the risks posed by privacy, bias, transparency, and ethical concern. The only way to navigate this landscape is to continue interdisciplinary collaboration between AI developers, educators, psychologists, and ethicists.

Future Directions

There are a number of directions that can continue to develop this field. Enhancing the granularity and scope of emotion detection is one of the directions. The existing systems often represent large-scale categories of emotions or a basic engaged/disengaged dichotomy, whereas a new generation of models can potentially represent more fine-tuned affective states and changes as well as a more complex construct such as motivation, boredom, or flow. It may include training on increasingly subtle labels and unsupervised or semi-supervised learning to learn a pattern not present in the useful categories (Khare et al., 2024). In addition, the longitudinal tracking of emotion may be used to show trends, such as the change of the emotive tone of a student to be more negative, which would reflect underlying problems. Another frontier is combining academic performance data and context (e.g., deadlines, exam periods) with emotion data to predictive analytics that can identify situations when academic and emotional factors in isolation do not endanger a student, but when combined, do harm him. The modeling side has new possibilities in the proliferation of transformer-based architectures and large multimodal models. As large language models have performed so well in NLP, large pre-trained emotion models are possible, exploiting large datasets of facial, vocal, and physiological data (Walsh, 2025). They can have a vivid depiction of human emotion, and can be personalized to certain educational situations. It is also possible that we will gain increased applications of transfer learning to other domains (e.g. applying models that have been trained on clinical emotion recognition to assist in detecting mental health signs among students). The aspect of cross-cultural adaptability will be addressed, which could be personalized or federated learning strategies that adjust emotion models to every classroom or population but share common knowledge. As far as deployment is concerned, it is important to have a smooth integration into the e-learning platforms. Emotion recognition must be an invisible component to the systems such as Zoom, Microsoft Teams, or LMS platforms, not an isolated application like a noise-cancellation or auto-captioning feature is nowadays. A few researchers have already developed prototype plug-ins that superimpose emotion analytics in video conferencing packages in education (as cited by a study that applied multimodal emotion detection to Microsoft Teams with promising results in gain in engagement). We expect to see more integrations of this kind and we are doing to design the user experiences in such cases to be valuable but not distracting. Emotion AI-based intervention strategies will also change. In addition to teacher notifications, the systems of the future may communicate with students directly: an encouraging chatbot may provide some tips or words of encouragement should the system notice that a learner is getting frustrated, or an intelligent tutor may automatically change the level of difficulty of the material. This can be connected to mental health resources- e.g. in case of a student who has displayed prolonged stress, the system may recommend a brief mindfulness session or alert a school counselor (with their permission). Educational psychology best practice should guide these intervention mechanisms to provide students with real help in order to cope and prosper. Lastly, the focus on ethical frameworks and policy will continue to influence the direction. With the increasing trend of emotion recognition as a marketing tool in the education sector, there is a high probability that there will be standards and regulations on the market that will help to protect the privacy of student data and regulate how it is used. The scientists including Walsh (2025) recommend the use of strong explainability and stakeholder control on any high-risk AI system, including emotion-reading systems. Schools and universities will have to come up with effective methods of consent and open communication with students and parents on how these mechanisms operate and what they (and do not) imply. This discussion may be empowering to students as well, by helping them be more aware of themselves, and lessen the possibility of being misinterpreted. The vision of AI in education of tomorrow is, more simply put, a human-centered technology: AI that is compassionate, involved, and supportive, and created and implemented in a profound respect of the human emotions it aims to comprehend.

Conclusion

Emotion recognition with deep learning has come out as a potent platform to aid mental health in online learning. Over the last five years, the accuracy and viability of automatic emotion monitoring have increased significantly due to the development of neural network models, starting with CNNs that recognize facial expression and ending with transformers that read text sentiment (Khare et al., 2024; Khan et al., 2025). These systems may serve as a constantly attentive companion to teachers: to see when a distance student gets lost, overwhelmed, or uninterested and deliver evidence-based information to personalize learning and to get assistance. The initial research shows that these affect-sensitive methods may improve engagement with students and even identify signs of mental health issues that would not be enacted otherwise (Bhardwaj et al., 2021; Masud et al., 2025). It is a paradigm shift in online learning: i.e., the shift of the one-size-fits-all approach to online education, where cognizational outcomes are the only significant concern, toward a more responsive paradigm, in which emotional well-being is regarded as an important aspect of learning. Simultaneously, the review has demonstrated that limitations and ethical traps are the fundamental issues to be addressed. Emotion AI is not flawless, it has to be constantly improved to be reasonable, transparent, and contextually aware. Motivated by these reasons, developers and educators should remember that the understanding of a smile or a frown by an algorithm is merely an approximation of an otherwise complex interior condition (Walsh, 2025). As such human support must be supported with systems rather than be substituted. Such technologies with appropriate safeguards can be used in a way that does not infringe upon the rights and dignity of students, as it might be done by the use of a privacy-preserving design (Shehada et al., 2025), opt-in usage, and counselor participation when the issues become more serious. Finally, the hope of e-learning that lies at the heart of deep learning-based emotion recognition is that it will assist in producing not only intelligent but also emotional education. Monitoring the emotional pulse of the classroom continuously, educators will be able to make sure that an unspoken request of assistance of any student could not remain unnoticed in the realm of the Internet. The current study direction indicates that, used intelligently, AI-powered emotion tracking will be able to make online education more individualized, humane, and able to support the development of the mind and the mental well-being of all learners. The key to its successful implementation in the changing world of education will be continued interdisciplinary cooperation and ethical attention to the issue to make sure that technology becomes the source of empowerment and care in the changing world..

References

Arji, G., Erfannia, L., Alirezaei, S., & Hemmat, M. (2023). A systematic literature review and analysis of deep learning algorithms in mental disorders. *Informatics in Medicine Unlocked, 40*, 101284. https://doi.org/10.1016/j.imu.2023.101284

Belludi, S., & Kopackova, H. (2025). Adoption of deep learning and recognition of emotion across e-learning platforms with implementation of blockchain across various devices. *Procedia Computer Science, 263*, 90–97. https://doi.org/10.1016/j.procs.2025.07.012

Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhatt, A. (2021). Application of deep learning on student engagement in e-learning environments. *Computers & Electrical Engineering, 93*, 107277. https://doi.org/10.1016/j.compeleceng.2021.107277

Jordan, E., Terrisse, R., Lucarini, V., Alrahabi, M., Krebs, M.-O., Desclés, J., & Lemey, C. (2025).

Speech emotion recognition in mental health: Systematic review of voice-based applications. *JMIR Mental Health, 12*(1), e74260. https://doi.org/10.2196/74260

Khan, H. U., Naz, A., Alarfaj, F. K., & Almusallam, N. (2025). Analyzing student mental health with RoBERTa-Large: A sentiment analysis and data analytics approach. *Frontiers in Big Data, 8*, 1615788. https://doi.org/10.3389/fdata.2025.1615788

Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2024). Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. *Information Fusion, 102*, 102019. https://doi.org/10.1016/j.inffus.2023.102019

Lu, L., Yuan, L., & Chen, L. (2025). Deep learning-based emotion recognition for analyzing students' psychological states during competitions. *Entertainment Computing, 55*, 101005. https://doi.org/10.1016/j.entcom.2025.101005

Masud, G. H. A., Shanto, R. I., Sakin, I., & Kabir, M. R. (2025). Effective depression detection and interpretation: Integrating machine learning, deep learning, language models, and explainable AI. *Array, 25*, 100375. https://doi.org/10.1016/j.array.2025.100375

Na, J., Aldrees, R., Hakeem, A., Mohaisen, D., Umer, M., AlHammadi, D. A., Alsubai, S., Innab, N., & Ashraf, I. (2024). FacialNet: Facial emotion recognition for mental health analysis using U-Net segmentation with transfer learning model. *Frontiers in Computational Neuroscience, 18*, 1485121. https://doi.org/10.3389/fncom.2024.1485121

Shehada, D., Tawfik, H., Bouridane, A., & Hussain, A. (2025). An explainable framework for mental health monitoring using lightweight and privacy-preserving federated facial emotion recognition. *Sensors, 25*(23), 7320. https://doi.org/10.3390/s25237320

Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition. *Big Data and Cognitive Computing, 7*(1), 48. https://doi.org/10.3390/bdcc7010048

Walsh, E. (2025). Does a face speak for itself? Emotion recognition technologies and explainable AI. *Philosophy & Technology, 38*(1), 67. https://doi.org/10.1007/s13347-025-00891-8

Wang, F., & Dong, J. (2025). The application of improved AFCNN model for children's psychological emotion recognition. *Scientific Reports, 15*, Article 24138. https://doi.org/10.1038/s41598-025-10269-7

Wang, X., Ren, Y., Luo, Z., He, W., Hong, J., & Huang, Y. (2023). Deep learning-based EEG emotion recognition: Current trends and future perspectives. *Frontiers in Psychology, 14*, 1126994. https://doi.org/10.3389/fpsyg.2023.1126994

Aly, M. (2025). Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model. *Multimedia Tools and Applications, 84*(13), 12575–12614. https://doi.org/10.1007/s11042-024-19392-5 (in press)