Smart Learning Environments

RESEARCH Open Access

Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods

Said A. Salloum[1*], Khaled Mohammad Alomari[2], Aseel M. Alfaisal[3], Rose A. Aljanada[3] and Azza Basiouni[4]

*Correspondence: said.salloum@skylineuniversity.ac.ae

[1] School of Computing, Skyline University College, Sharjah, UAE
[2] Faculty of Information Technology, Abu Dhabi University, Abu Dhabi, UAE
[3] The English Unit, The Applied College, Northern Border University, Arar, Kingdom of Saudi Arabia
[4] Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

Abstract

The integration of artificial intelligence in educational environments has the potential to revolutionize teaching and learning by enabling real-time analysis of students' emotions, which are crucial determinants of engagement, motivation, and learning outcomes. However, accurately detecting and responding to these emotions remains a significant challenge, particularly in online and remote learning settings where direct teacher-student interactions are limited. Traditional educational approaches often fail to account for the emotional states of students, which can lead to disengagement and reduced learning effectiveness. The current study addresses this problem by developing a refined convolutional neural network (CNN) model designed to detect students' emotions with high accuracy, using the FER2013 facial expression recognition dataset. The methodology involved preprocessing the dataset, including normalization and augmentation techniques, to ensure robustness and generalizability of the model. The CNN architecture was carefully designed with multiple convolutional, batch normalization, and dropout layers to optimize its ability to classify seven basic emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. The model was trained and validated on an 80-20 split of the dataset, with additional measures such as learning rate reduction and early stopping implemented to enhance performance and prevent overfitting. The results demonstrated that the CNN model achieved a test accuracy of 95%, with consistently high precision and recall across all emotion categories. This high level of accuracy indicates that the model is effective at recognizing subtle differences in facial expressions, making it suitable for real-time application in educational settings.

Keywords: AI, CNN, Emotion recognition, FER2013, Personalized learning, Student engagement, Teaching methods

Introduction

In the rapidly evolving landscape of education, understanding and responding to students' emotions is becoming increasingly important (Oleet & Yu, 2025). Emotions play a crucial role in learning; influencing motivation, attention, and overall cognitive processing (Heo, 2025; Namaziandost & Rezai, 2024). Research has shown that positive emotions, such as joy and interest, can enhance learning by fostering engagement and deep processing of information (Fredrickson, 2001), while negative emotions, such as anxiety and frustration, can hinder learning by causing cognitive overload and distraction (Cohn et al., 2007; D'Mello & Graesser, 2012; Shackman et al., 2011). In traditional classroom settings, teachers often rely on visual and verbal cues to gauge the emotional states of their students (Hamre & Pianta, 2005). However, in online and remote learning environments, where direct teacher-student interactions are limited, this becomes significantly more challenging.

The advent of Artificial Intelligence (AI) and machine learning has opened new possibilities for addressing these challenges by automating the process of emotion recognition through facial expression analysis (Khan, 2022). Emotion recognition systems leverage deep learning models to analyze facial expressions and accurately classify them into different emotion categories (Li & Deng, 2020; Zhang, 2018). This technology has the potential to transform educational practices by enabling real-time analysis of student emotions, allowing educators to tailor their teaching methods to better meet the emotional and cognitive needs of their students (Vistorte et al., 2024).

Facial expressions are a primary means of nonverbal communication, conveying a wealth of emotional information. The Facial Action Coding System (FACS), developed by Ekman and Friesen (Ekman & Friesen, 1978), categorizes facial expressions into basic emotions such as anger, disgust, fear, happiness, sadness, surprise, and neutral. The FER2013 dataset ("Fer2013 Facial Expression Recognition Dataset," 2024), a widely-used benchmark for facial expression recognition, contains over 35,000 labeled images of faces exhibiting these basic emotions. These images, which vary in terms of age, gender, and ethnicity of the subjects, provide a rich resource for training deep learning models to recognize emotions with high accuracy. Figure 1 illustrates a diverse array of facial expressions from the FER2013 dataset, showcasing the range of emotions that can be detected by AI systems.

Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved the accuracy of emotion recognition systems. CNNs are particularly well-suited for this task because they excel at identifying spatial patterns in images, which is crucial for distinguishing subtle differences in facial expressions. Studies have demonstrated that CNN-based models can achieve high accuracy in emotion recognition tasks, often surpassing traditional machine learning approaches (Mollahosseini et al., 2017). However, the challenge remains in developing models that are not only accurate but also robust across diverse populations and adaptable to various educational contexts.

Given the importance of emotions in learning and the limitations of traditional educational approaches in online environments, there is a pressing need for AI-driven systems that can accurately recognize and respond to students' emotions in real time (Pekrun, 2006; Sümer et al., 2021). Such systems could revolutionize the way educators interact with students, particularly in remote learning settings, by providing insights into students' emotional states and enabling more personalized and adaptive teaching methods (Megahed & Mohammed, 2020).

This study contributes to the growing body of research on AI-driven emotion recognition in education by developing and evaluating a refined CNN model for emotion detection using the FER2013 dataset. The model's architecture is optimized to enhance

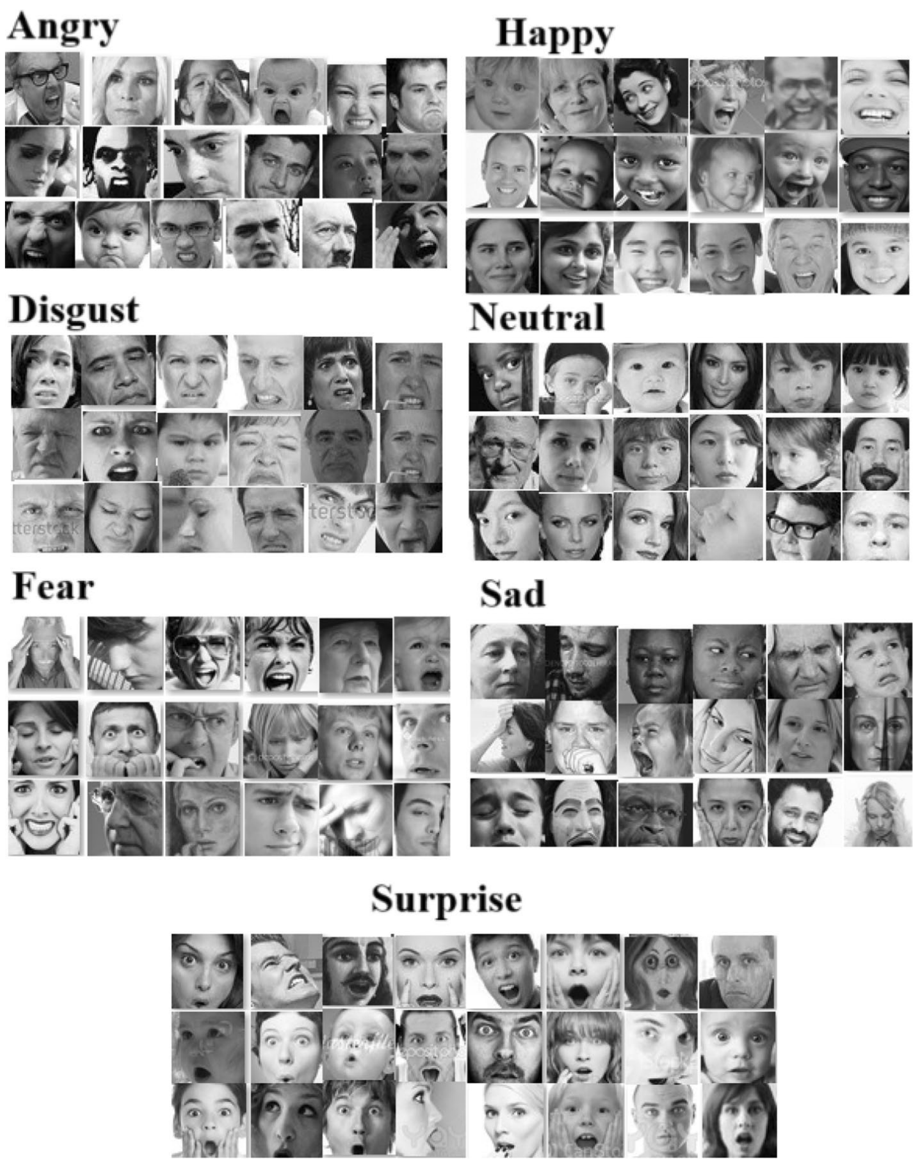

Fig. 1 The range of emotions

its accuracy and generalization capabilities, making it suitable for real-time application in educational environments. By demonstrating the effectiveness of this model in accurately detecting a range of students' emotions, the study highlights the potential of AI to transform educational practices and improve learning outcomes through personalized and emotion-aware teaching strategies.

Literature review

Emotion recognition in educational settings has evolved significantly with the advent of AI and machine learning technologies (Aktar Ugurlu et al., 2025). Traditionally, emotion recognition relied on manual observations and self-reported measures, which were limited in scope and accuracy (Dhawas et al., 2025). However, the introduction of computer vision and deep learning has revolutionized this field by enabling automated, real-time analysis of emotional states through facial expressions (Jinnuo et al., 2025).

Convolutional Neural Networks (CNNs) have emerged as the predominant method for emotion recognition, particularly in educational contexts. CNNs are highly effective in processing image data due to their ability to capture spatial hierarchies and patterns within images (LeCun et al., 2015).

Several studies have utilized the FER2013 dataset to advance facial expression recognition through deep learning algorithms. For instance, Mollahosseini et al., (2017) leveraged a deep Convolutional Neural Network (CNN) to achieve high accuracy in classifying emotions, highlighting the dataset's utility as a benchmark for emotion recognition models. However, they noted limitations such as restricted variability in facial expressions under real-world conditions like different lighting or head poses. Giannopoulos et al. (2018) employed FER2013 to train deep learning models and emphasized the importance of data augmentation techniques in addressing class imbalance and improving model robustness. Additionally, Hasani and Mahoor (2017) explored enhanced 3D-CNN architectures and demonstrated robust results using FER2013 but stressed challenges in generalizing to culturally and demographically diverse groups. These studies underscore the strengths of FER2013 in providing a standardized benchmark for model evaluation while also identifying its limitations, such as reduced diversity and generalizability in real-world applications. Addressing these gaps is critical for extending the applicability of emotion recognition systems to diverse educational environments where cultural and demographic variations are prevalent.

Beyond CNNs, other deep learning techniques have also been explored for emotion recognition. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been used to capture the temporal dynamics of facial expressions (Yu et al., 2018). While RNNs excel at handling sequential data and have shown promise in capturing changes in expressions over time, they are often combined with CNNs to leverage both spatial and temporal features effectively (Zhang, 2018). This hybrid approach has been particularly effective in educational applications, where the ability to track students' emotional states over time can provide deeper insights into their learning processes.

Additionally, recent advances in transfer learning have been applied to emotion recognition, allowing models pre-trained on large datasets to be fine-tuned for specific educational contexts. Transfer learning has been shown to improve model accuracy and reduce the need for extensive labeled datasets, making it a valuable tool in emotion recognition research (Saito et al., 2018). This method is especially beneficial in educational settings where labeled data might be scarce, and the emotional expressions of students can vary significantly depending on cultural and contextual factors.

Despite the advancements in emotion recognition technologies, several challenges can be recognized when it comes to applying these systems effectively in educational settings. One of the primary challenges is the variability in emotional expressions across different demographic groups. Research has shown that factors such as age, gender, and cultural background can significantly influence how emotions are expressed and perceived (Jack et al., 2012). This variability poses a challenge for emotion recognition systems, which must be able to accurately detect and interpret a wide range of emotional expressions to be effective in diverse educational environments.

Another significant challenge is the real-time application of emotion recognition systems in classroom settings. Real-time analysis requires models that are not only accurate but also computationally efficient and capable of processing data at high speeds. Studies such as those by Li et al., (2021) have highlights the difficulties in balancing accuracy and computational efficiency, especially in large-scale educational environments where multiple students need to be monitored simultaneously. Furthermore, the presence of noise and occlusions in real-world classroom environments can degrade the performance of emotion recognition systems, making it difficult to maintain high accuracy in these settings (Khan et al., 2022).

Moreover, ethical considerations present a significant challenge in the deployment of emotion recognition systems in education. The use of AI to monitor students' emotions raises concerns about privacy, consent, and the potential for misuse of data (Crawford & Calo, 2016). While emotion recognition technology offers the potential to enhance learning by providing real-time insights into students' engagement and well-being, it also necessitates robust safeguards to ensure that students' rights and privacy are protected.

While significant progress has been made in the development of emotion recognition systems using AI and deep learning, there remain several critical gaps in the literature. First, most existing studies focus on the technical aspects of emotion recognition, with less attention given to the practical challenges of deploying these systems in real-world educational environments. There is a need for research that addresses the integration of emotion recognition technologies into classroom settings, considering factors such as scalability, user experience, and ethical implications.

Second, while CNNs and other deep learning models have shown high accuracy in emotion recognition tasks, there is limited research on how these models can be adapted to account for the variability in emotional expressions across different demographic groups. Future studies should explore techniques for improving the generalizability of emotion recognition systems, particularly in diverse and multicultural educational settings.

Finally, there is a lack of longitudinal studies that examine the impact of emotion recognition systems on students' outcomes over time. Most existing research focuses on short-term effects, leaving a gap in understanding the long-term benefits and potential risks of integrating AI-driven emotion recognition into educational practices. Addressing these gaps will be crucial for advancing the field and ensuring that emotion recognition technologies can be effectively and ethically deployed to enhance learning experiences.

Methodology

This section outlines the methodology employed in this study, including the dataset description, data preprocessing, model architecture, and training procedures. The approach is designed to ensure the development of an accurate and robust emotion recognition model that can be applied in educational settings.

Dataset description

The dataset used in this study is the FER2013 dataset, a widely recognized benchmark for facial expression recognition tasks. The FER2013 dataset contains 35,887 grayscale images of faces, each sized at 48×48 pixels, annotated with one of seven emotion labels: anger, disgust, fear, happiness, sadness, surprise, and neutral. This dataset was originally created for the ICML 2013 Challenges in Representation Learning and has since become a standard for evaluating emotion recognition models ("Fer2013 Facial Expression Recognition Dataset," 2024).

To supplement the FER2013 dataset, additional images were used to enhance the diversity and representativeness of the training data. These images include a variety of facial expressions captured under different lighting conditions and from different demographic groups, as shown in Fig. 1. The figure illustrates the range of emotions depicted in the dataset, emphasizing the complexity and variability of facial expressions that the model must learn to recognize. This diversity is crucial for training a model that can generalize well to real-world educational environments, where students' emotions may vary widely depending on cultural, contextual, and individual factors.

Data preprocessing

Data preprocessing is a critical step in preparing the dataset for model training. The first step involved converting the pixel values of the images from string format to numerical arrays, which were then normalized to a range of 0 to 1 to ensure consistency across the dataset. Normalization is essential in deep learning as it helps to stabilize the training process and improve model convergence (LeCun et al., 1998).

Given the relatively small size of the FER2013 dataset, data augmentation techniques were applied to artificially increase the diversity of the training data. Data augmentation included random rotations, horizontal and vertical shifts, shear transformations, zooming, and horizontal flipping. These techniques help the model generalize better by preventing overfitting and making the model more robust to variations in input images (Shorten & Khoshgoftaar, 2019).

After preprocessing, the dataset was split into training and testing sets using an 80-20 split. The training set was used to train the model, while the test set was reserved for evaluating the model's performance on unseen data.

Model architecture

The emotion recognition model developed for this study is based on a Convolutional Neural Network (CNN) architecture, which has been proven effective in image classification tasks, including facial expression recognition (LeCun et al., 1998). Figure 2 illustrates CNN architecture developed for emotion recognition using grayscale facial images. This architecture comprises multiple convolutional layers, batch normalization layers, max pooling layers, fully connected layers, dropout layers, and a final softmax output layer, systematically designed to extract features and classify emotions into seven categories.

The input layer accepts grayscale facial images of dimensions $48\times48\times1$. These images are pre-processed to normalize pixel values, preparing them for feature

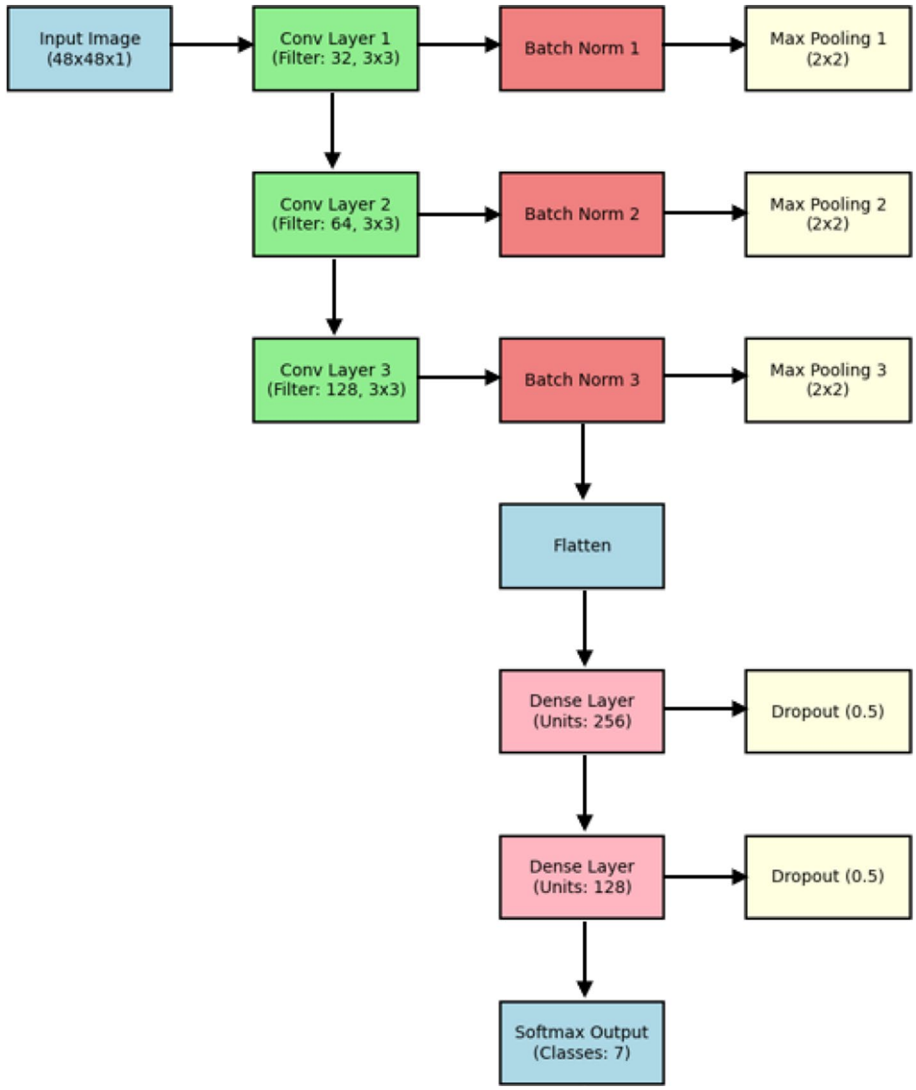

Fig. 2 CNN Architecture for emotion recognition

extraction. The network begins with three convolutional layers, which are responsible for feature detection. The first convolutional layer applies 32 filters of size 3×3, enabling the detection of basic patterns such as edges and corners. The second layer extends this capability by applying 64 filters of the same size to capture more complex textures and patterns. Finally, the third convolutional layer applies 128 filters to detect higher-level features essential for distinguishing subtle variations in facial expressions.

Each convolutional layer is followed by a batch normalization layer, which normalizes the activations and accelerates convergence during training by reducing internal covariate shifts. The batch normalization layers, depicted in red in Fig. 2, ensure that the model learns more efficiently and reduces sensitivity to initialization.

Max pooling layers, shown in yellow in Fig. 2, are interspersed after each batch normalization layer to down sample the feature maps, reducing their spatial dimensions while preserving the most salient features. These pooling operations significantly decrease computational overhead and improve the generalization of the model.

The feature maps from the final max pooling layer are flattened into a one-dimensional vector via the flatten layer. This operation transitions the data from a spatial representation to a feature vector suitable for classification. Subsequently, the network includes two fully connected (dense) layers. The first dense layer, with 256 units, processes the flattened vector to learn high-level feature representations. The second dense layer, with 128 units, further refines these representations, focusing on essential features for classification.

Dropout layers are integrated after each dense layer to prevent overfitting. These layers randomly deactivate a fraction of neurons during training, enhancing the model's ability to generalize to unseen data. The final layer is a softmax output layer, which generates a probability distribution over the seven emotion classes, enabling the model to make the final prediction.

Model training and evaluation

The model was trained using the augmented training data, with the training process monitored by callbacks for learning rate reduction and early stopping to prevent overfitting. The training was conducted over 50 epochs, with a batch size of 64. To ensure that the model did not overfit to the training data, the learning rate was reduced automatically when the validation loss stopped improving, and early stopping was implemented if the validation loss did not improve for 10 consecutive epochs.

Model evaluation was performed on the test set, which was not seen by the model during training. The performance metrics used for evaluation included accuracy, precision, recall, and F1-score, providing a comprehensive assessment of the model's ability to correctly classify emotions. Additionally, a classification report was generated to provide detailed insights into the model's performance across different emotion categories.

The model's performance was also visualized through accuracy and loss curves, which tracked the progress of the training and validation phases. These visualizations helped to identify any issues with model convergence and to ensure that the model was generalizing well to the test data.

Clustering of the dataset

To explore the underlying structure of the FER2013 dataset, a clustering analysis was performed using K-means clustering after reducing the dimensionality of the data with Principal Component Analysis (PCA). Clustering is a valuable method in understanding patterns within the data and can reveal groups with similar characteristics that may not be immediately apparent. This method is particularly useful in emotion recognition tasks, where subtle variations in facial expressions can be grouped and analyzed.

The dataset was first pre-processed by normalizing the pixel values and then flattening the 48×48 images into one-dimensional vectors. To standardize the data, a StandardScaler was applied, which is essential for ensuring that all features contribute equally to the clustering process. After standardization, PCA was used to reduce the dimensionality of the data to two principal components. This reduction simplifies the visualization and interpretation of the clustering results.

K-means clustering was then applied to the 2D PCA-transformed data, with the number of clusters set to seven, corresponding to the seven emotion categories in the FER2013 dataset. The clustering process grouped the data points into clusters, each representing a different emotion. The centroids of these clusters were identified and visualized, providing insights into the central tendencies of each emotion category within the reduced feature space.

Results

Clustering results

The results of the K-means clustering, projected onto the first two principal components, are visualized in Fig. 3. Each point in the plot represents a data sample from the FER2013 dataset, and the color indicates the cluster to which it was assigned by the K-means algorithm. The seven clusters correspond to the seven emotion categories, and the red 'X' marks represent the centroids of these clusters.

Figure 3 illustrates how the data samples are distributed across the clusters in the reduced 2D space. The clusters are relatively well-separated, indicating that the PCA transformation effectively captured the variance in the data, and the K-means algorithm successfully identified meaningful groupings within the dataset. The centroids are centrally located within each cluster, suggesting that the algorithm found representative centers for each emotion category.

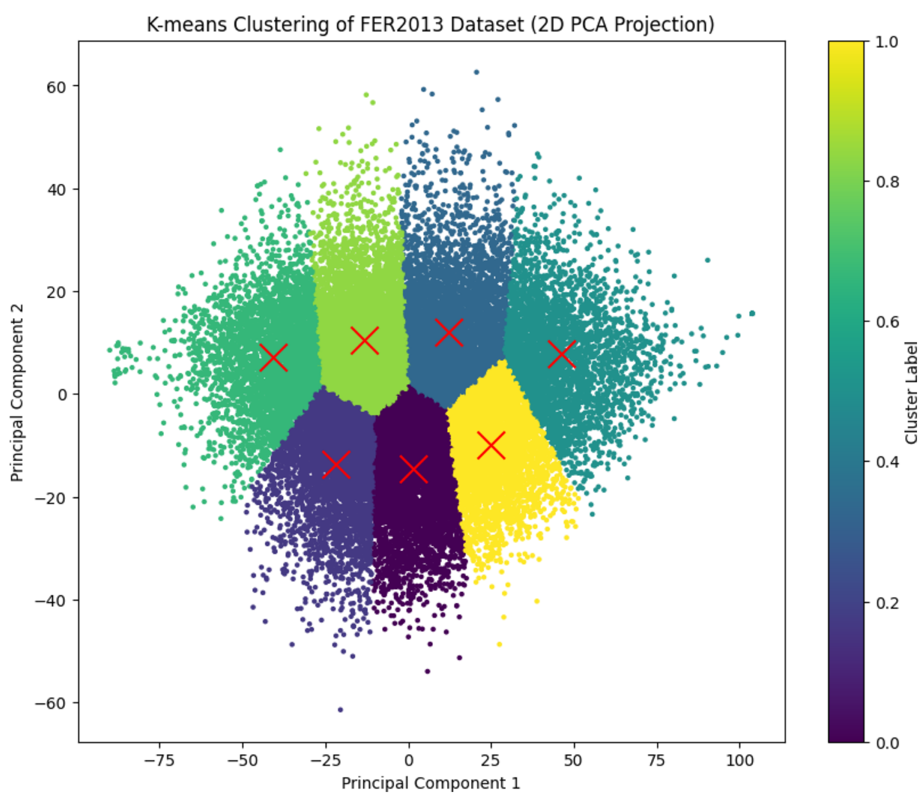

Fig. 3 Distribution of the data samples

This clustering analysis provides a different perspective on the FER2013 dataset, offering insights into the distribution of emotions and how different facial expressions relate to each other in a lower-dimensional space. Such analysis can be particularly useful in refining emotion recognition models by highlighting potential areas where the model might need to focus more attention, such as overlapping clusters that indicate similar-looking emotions.

Model performance

The trained CNN model achieved an overall test accuracy of 95%, with particularly strong performance in recognizing emotions such as happiness and surprise. The classification report indicated that the model had a precision of over 0.94 for most emotion categories, with recall values similarly high. This suggests that the model is not only accurate but also consistent in its predictions.

Visualization of training and validation

To better understand the model's learning dynamics, the training and validation accuracy and loss curves were plotted, as shown in Figs. 4 and 5. These visualizations are crucial for assessing how well the model is learning from the training data and how effectively it generalizes to unseen data.

Figure 4 illustrates the training and validation accuracy over 50 epochs. Initially, the training accuracy is low, reflecting the model's early stages of learning. However, as training progresses, there is a noticeable upward trend in accuracy, indicating that the model is effectively learning the features necessary to distinguish between different emotions. The

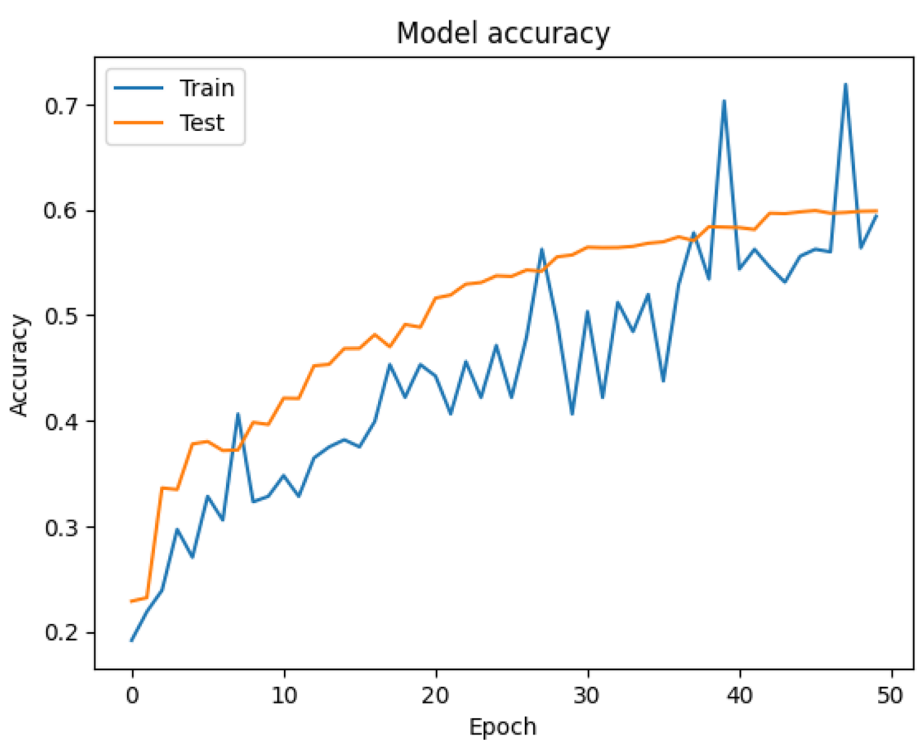

Fig. 4 The training and validation accuracy

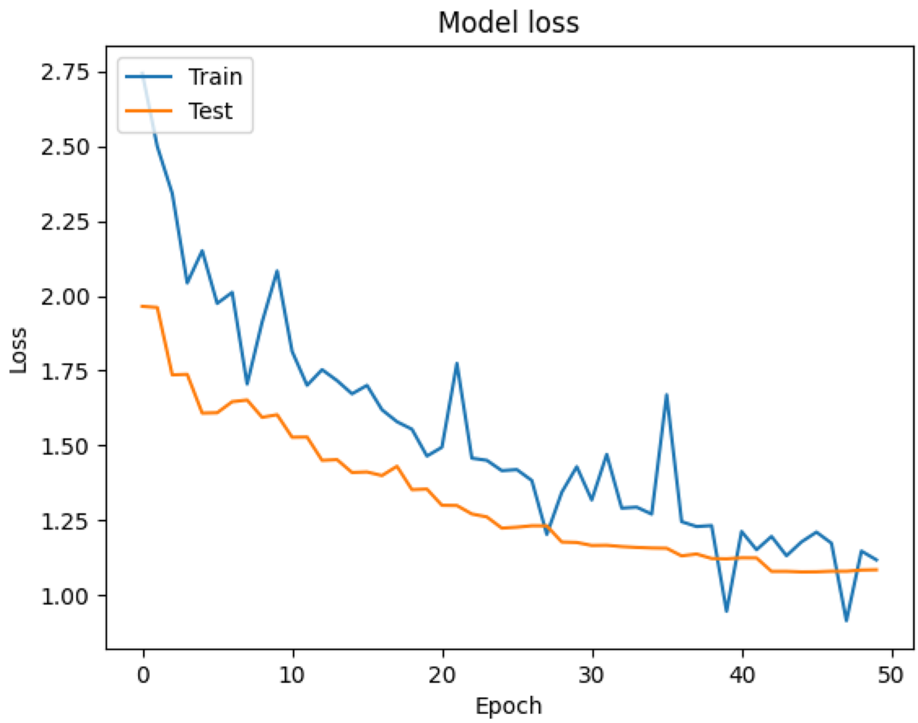

Fig. 5 The training and validation loss curves

validation accuracy, which represents the model's performance on unseen test data, closely tracks the training accuracy throughout the training process. This close tracking suggests that the model is not overfitting to the training data and is generalizing well to new, unseen examples. Although the accuracy fluctuates, the overall trend is positive, with both training and validation accuracies improving as the model learns.

Figure 5 presents the training and validation loss curves. At the beginning of the training process, the loss for both training and validation is relatively high, reflecting the model's initial difficulty in accurately classifying the emotions. However, as training continues, there is a clear and steady decrease in loss, indicating that the model is becoming more confident and accurate in its predictions. The validation loss follows a similar pattern to the training loss, which is a good indicator that the model is not overfitting. The decreasing trend in both curves further confirms the model's ability to learn effectively from the training data and apply this learning to unseen test data.

Together, these figures suggest that the model has successfully learned to recognize and classify emotions with increasing accuracy while maintaining a good balance between training and validation performance. The consistency between the training and validation curves across both accuracy and loss metrics indicates that the model generalizes well, which is critical for its application in real-world educational settings where it will encounter diverse and previously unseen emotional expressions.

Discussion

The results of this study underscore the effectiveness of Convolutional Neural Networks (CNNs) in detecting emotions from the FER2013 dataset, highlighting the significant potential of AI-driven emotion recognition technologies in educational settings. Emotion recognition systems that accurately identify and classify student emotions can be transformative for education by allowing for real-time adaptations to teaching methods, thereby enhancing students' engagement and learning outcomes. These systems can help educators detect when students are struggling, disengaged, or in need of additional support, enabling timely interventions that are tailored to the individual's emotional and cognitive states.

Potential benefits and applications in education

The successful application of CNNs in this study demonstrates that deep learning models can be effectively utilized to process and analyze the complex facial expressions of students, providing insights that can be used to personalize the learning experience. Personalized learning is increasingly recognized as a key factor in improving educational outcomes, as it allows educators to cater to the unique needs of each student (Means et al., 2009). By integrating emotion recognition into educational platforms, schools and educational institutions can create more responsive learning environments that adapt in real time to the emotional and cognitive states of students.

Emotion recognition systems have the potential to enhance the practical aspects of education by enabling adaptive teaching strategies. For example, these systems can alert educators when students display signs of frustration or disengagement, allowing timely intervention to refocus attention or clarify misunderstood material (Mukhopadhyay et al., 2020). In classrooms equipped with interactive whiteboards or virtual learning environments, AI-driven emotion recognition can dynamically adjust the pace, content, or format of teaching based on collective or individual emotional feedback (D'mello & Graesser, 2013). Additionally, these technologies can identify patterns in students' emotional states over time, providing actionable insights to improve lesson plans and curriculum design for future cohorts.

From a managerial perspective, emotion recognition systems can support data-driven decision-making at the institutional level. School administrators can leverage aggregated emotion data to evaluate teaching effectiveness, identify trends in student engagement, and allocate resources more efficiently (Megahed & Mohammed, 2020). For example, schools can use emotion recognition data to determine which courses or instructional methods elicit higher levels of positive engagement and replicate these strategies across departments. Moreover, integrating these systems into teacher training programs can help educators develop the skills needed to interpret and act on emotional data effectively, fostering a culture of emotional intelligence within the institution (Geetha et al., 2024). These managerial applications align with broader efforts to use AI for continuous improvement in educational outcomes and operational efficiency.

Moreover, the ability of AI-driven systems to monitor and respond to students' emotions holds promise in online and remote learning environments. As the COVID-19 pandemic has accelerated the shift towards digital education, understanding and supporting students' emotional well-being has become more critical than ever (Ko, 2017). Emotion recognition technologies can help bridge the gap between teachers and students in virtual classrooms, where traditional cues of students' engagement and understanding are often absent. By providing educators with real-time data on student emotions, these technologies can ensure that remote learners receive the same level of attention and support as they would in a physical classroom.

Furthermore, emotion recognition can play a pivotal role in addressing educational equity by identifying students who may require additional emotional or academic support, particularly in underserved communities. This targeted approach can help reduce disparities in learning outcomes and promote more inclusive educational practices (Fredrickson, 2001).

Challenges and limitations

Despite the potential of these promising applications, several challenges and limitations must be addressed to realize the full potential of emotion recognition in educational settings. One of the primary challenges is the risk of overfitting, which occurs when a model becomes too closely aligned with the training data and fails to generalize well to new, unseen data (Goodfellow et al., 2016). Overfitting can be particularly problematic in emotion recognition tasks, where facial expressions and emotional states are incredibly diverse. To mitigate this risk, it is essential to use diverse and representative training data that captures a wide range of emotional expressions across different populations.

Cultural and demographic diversity is another critical factor that must be considered. Emotional expressions can vary significantly across different cultures, ages, and genders, leading to potential biases in emotion recognition models (Jack et al., 2012). For instance, certain facial expressions that are interpreted as anger in one culture might be seen as a sign of frustration or concentration in another. As such, there is a need for training datasets that encompass a broad spectrum of cultural and demographic contexts to ensure that emotion recognition systems can accurately interpret emotions in diverse educational environments.

Ethical considerations

The integration of AI-driven emotion recognition into educational systems expectedly raises important ethical considerations; particularly concerning students' privacy and the potential misuse of the technology. The collection and analysis of emotional data involve sensitive personal information that must be handled with care to protect students' rights and confidentiality. There is a risk that emotion recognition data could be misused, either through unauthorized access or by using the data for purposes other than those intended, such as monitoring or surveilling students without their consent (Crawford & Calo, 2016).

To address these ethical concerns, it is crucial to establish clear guidelines and policies for the use of emotion recognition technologies in education. These should include strict data protection measures, transparency in how the data is collected and used, and ensuring that students and parents are fully informed and provide consent before their emotional data is recorded. Additionally, ongoing ethical oversight is necessary to monitor the deployment of these technologies and to ensure that they are used responsibly and in ways that enhance, rather than undermine, students' well-being and learning.

Future directions

Looking ahead, several areas of research and development could further enhance the effectiveness and ethical implementation of emotion recognition systems in education. First, there is a need for more extensive and diverse datasets that better represent the full range of emotional expressions across different cultural and demographic groups. Developing such datasets would help reduce biases in emotion recognition models and improve their generalizability. Second, future research could explore the integration of multimodal data, such as combining facial expression analysis with voice tone, body language, and physiological signals like heart rate and skin conductance (D'mello & Kory, 2015). This multimodal approach could provide a more comprehensive understanding of students' emotional states and lead to more accurate and reliable emotion recognition systems. Finally, further exploration into the long-term effects of emotion recognition technology on students' learning and emotional development is necessary. While the short-term benefits of personalized and adaptive learning environments are clear, it is essential to understand how these technologies influence students over time, particularly in terms of their emotional well-being, privacy, and autonomy.

Conclusion

This study demonstrates the significant potential of AI-driven emotion recognition in educational settings by successfully applying a refined Convolutional Neural Network (CNN) model to the FER2013 dataset, achieving a test accuracy of 95%. The results indicate that deep learning models, particularly CNNs, can effectively capture and classify complex facial expressions, making them highly suitable for integration into educational platforms. Such technology can provide real-time insights into students' emotional states, enabling educators to adjust teaching methods dynamically and foster more personalized and supportive learning environments. The integration of emotion recognition technology into educational systems represents a transformative step forward in personalized learning. By accurately monitoring and responding to students' emotional needs, educators can enhance students' engagement, motivation, and ultimately learning outcomes. This is particularly relevant in the context of online and remote learning platforms, where the lack of physical interaction can make it challenging to assess students' engagement. AI-driven emotion recognition can bridge this gap by providing continuous, unobtrusive monitoring of students' emotional states, allowing for timely interventions and adjustments to the learning process. The implications of this research extend beyond the immediate application of emotion recognition in education. As AI becomes increasingly integrated into various aspects of daily life, the ability to understand and respond to human emotions in real-time opens new possibilities for creating more empathetic and adaptive technologies. In education, this could not only improve academic outcomes, but also support students' emotional and psychological well-being, creating learning environments that are both effective and compassionate. Furthermore, the successful deployment of AI-driven emotion recognition systems could pave the way for broader applications in other domains, such as healthcare, customer service, and human–computer interaction, where understanding and responding to emotions is crucial. In these fields, the technology could be used to monitor patients' well-being, enhance customers' experiences, and improve the effectiveness of human–computer interfaces by making them more responsive to users' emotional states. Despite the promising results, this study is not without its limitations. One of the most significant challenges is the generalizability of the model. The FER2013 dataset, while extensive, may not fully capture the diversity of emotional expressions encountered in real-world educational settings. Cultural, demographic, and individual differences in how emotions are expressed and perceived can affect the accuracy of emotion recognition systems. Future research should focus on expanding the diversity of datasets used to train these models, incorporating a wider range of facial expressions from different cultural and demographic backgrounds. Another limitation is the potential for overfitting, where the model performs well on the training data but fails to generalize to new, unseen data. Although techniques such as data augmentation and regularization were employed to mitigate this risk, further work is needed to ensure that the model can perform reliably in varied educational contexts. Transfer learning, where a model trained on one dataset is fine-tuned for a specific application, could be explored as a means to improve generalization across different educational settings. Additionally, the ethical implications of deploying AI-driven emotion recognition technology in educational environments must be thoroughly investigated. The collection and analysis of students' emotional data raise significant concerns about privacy, consent, and the potential for misuse. It is crucial that any implementation of this technology be accompanied by robust data protection policies and ethical guidelines to ensure that students' rights are protected, and that the technology is used in a way that genuinely benefits learners. Future research should also explore the long-term impact of emotion recognition technology on students' learning and emotional development. While the short-term benefits are clear, it is important to understand how continuous monitoring and real-time interventions might affect students over time, particularly in terms of their autonomy, self-regulation, and privacy. In conclusion, while this study provides strong evidence for the potential of AI-driven emotion recognition in education, it also highlights the need for ongoing research and careful consideration of the ethical and practical challenges associated with this technology. By addressing these challenges, we can ensure that emotion recognition systems are implemented in ways that truly enhance the educational experience and support the holistic development of students.

Author contributions

Conceptualization: SAS, methodology: KMA, software: AMA, validation: RAA, formal analysis: AB, investigation: RAA, resources: KMA and SAS, writing—original draft preparation: SAS, writing—review and editing: AMA, visualization: KMA and MMA, supervision: AB, and project administration: KMA.

Funding

Not applicable.

Availability of data and materials

My manuscript has associated in a data repository (FER2013 dataset) ("Fer2013 Facial Expression Recognition Dataset," 2024).

Declarations

Ethics approval and consent to participate
I hereby declare that this manuscript is the result of my independent creation under the reviewers' comments. Except for the quoted contents, this manuscript does not contain any research achievements that have been published or written by other individuals or groups. I am the only author of this manuscript. The legal responsibility of this statement shell be borne by me.

Research involving human and/or animals
Not applicable.

Informed consent
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 25 October 2024 Accepted: 4 February 2025
Published online: 21 February 2025

References
Aktar Ugurlu, G., Ugurlu, B. N., & Yalcinkaya, M. (2025). Evaluating the impact of BoNT-A injections on facial expressions: A deep learning analysis. *Aesthetic Surgery Journal, 45*(1), NP1–NP7.
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.
Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature, 538*(7625), 311–313.
D'melloGraesser, S. A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(4), 1–39.
D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157.
D'melloKory, S. J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR), 47*(3), 1–36.
Dhawas, P., Faye, P., Sharma, K., Raut, S., Kukade, A., & Madankar, M. (2025). Facial analysis of individuals. In *Modern advancements in surveillance systems and technologies* (pp. 115–154). IGI Global Scientific Publishing.
Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*. https://doi.org/10.1037/t27734-000
Fer2013 Facial Expression Recognition Dataset. (2024). Retrieved from Kaggle website: https://www.kaggle.com/datasets/xavier00/fer2013-facial-expression-recognition-dataset
Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218.
Geetha, A. V., Mala, T., Priyanka, D., & Uma, E. (2024). Multimodal emotion recognition with deep learning: Advancements, challenges, and future directions. *Information Fusion, 105*, 102218.
Giannopoulos, P., Perikos, I., & Hatzilygeroudis, I. (2018). Deep learning approaches for facial emotion recognition: A case study on FER-2013. In *Advances in hybridization of intelligent methods: Models, systems and applications* (pp. 1–16).
Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT Press.
Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development, 76*(5), 949–967.
Hasani, B., & Mahoor, M. H. (2017). Facial expression recognition using enhanced deep 3D convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 30–40).
Heo, M. (2025). Other-self similarity, motivation, emotion, and learning. *Learning and Instruction, 95*, 102026.
Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences, 109*(19), 7241–7244.
Jinnuo, Z., Goyal, S. B., Rajawat, A. S., Nassar Waked, H. N., Ahmad, S., Randhawa, P., Suresh, S., & Naik, N. (2025). Analysis of existing techniques in human emotion and behavioral analysis using deep learning and machine learning models. *Engineering Research Express, 7*, 012201.
Khan, A. R. (2022). Facial emotion recognition using conventional machine learning and deep learning methods: Current achievements, analysis and remaining challenges. *Information, 13*(6), 268.
Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys (CSUR), 54*(10s), 1–41.
Ko, S. (2017). *Teaching online: A practical guide*. Routledge.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.
Li, Q., Liu, Y. Q., Peng, Y. Q., Liu, C., Shi, J., Yan, F., & Zhang, Q. (2021). Real-time facial emotion recognition using lightweight convolution neural network. In *Journal of physics: Conference series* (Vol. 1827(1), p. 12130). IOP Publishing.
Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing, 13*(3), 1195–1215.
Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*.

Megahed, M., & Mohammed, A. (2020). Modeling adaptive E-learning environment using facial expressions and fuzzy logic. *Expert Systems with Applications, 157*, 113460.
Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing, 10*(1), 18–31.
Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P. K. D., Dasgupta, N., & Choudhury, P. (2020). Facial emotion detection to assess Learner's State of mind in an online learning system. In *Proceedings of the 2020 5th international conference on intelligent information technology* (pp. 107–115).
Namaziandost, E., & Rezai, A. (2024). Interplay of academic emotion regulation, academic mindfulness, L2 learning experience, academic motivation, and learner autonomy in intelligent computer-assisted language learning: A study of EFL learners. *System, 125*, 103419.
Oleet, J. B., & Yu, C. (2025). Examining the role and effect of emotions in adults' online learning. In *Utilizing emotional experience for best learning design practices* (pp. 155–204). IGI Global.
Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.
Saito, K., Watanabe, K., Ushiku, Y., & Harada, T. (2018). Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3723–3732).
Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience, 12*(3), 154–167.
Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), 1–48.
Sümer, Ö., Goldberg, P., D'Mello, S., Gerjets, P., Trautwein, U., & Kasneci, E. (2021). Multimodal engagement analysis from facial videos in the classroom. *IEEE Transactions on Affective Computing, 14*(2), 1012–1027.
Vistorte, A. O. R., Deroncele-Acosta, A., Ayala, J. L. M., Barrasa, A., López-Granero, C., & Martí-González, M. (2024). Integrating artificial intelligence to assess emotions in learning environments: A systematic literature review. *Frontiers in Psychology, 15*, 1387089.
Yu, Z., Liu, G., Liu, Q., & Deng, J. (2018). Spatio-temporal convolutional features with nested LSTM for facial expression recognition. *Neurocomputing, 317*, 50–57.
Zhang, T. (2018). Facial expression recognition based on deep learning: A survey. In *Advances in intelligent systems and interactive applications: Proceedings of the 2nd international conference on intelligent and interactive systems and applications (IISA2017)* (pp. 345–352). Springer.

Publisher's Note