Ceddi Journal of Education
Vol. 4 No. 1 June (2025)
https://doi.org/10.56134/cje.v4i1.126

ISSN: 2830-0017 (print)
ISSN: 2830-0025 (online)

Student Expression Detection Based On Facial Image Using Convolutional Neural Network (CNN)

Muh. Riyaldi Pratama[1*], Erwin Gatot Amiruddin[2], Kamaruddin[3], Tamus Bin Tahir[4], Muhammad Qadri[5], Kumar Vivek[6]

[1]*Informatics Engineering, Universitas Teknologi Akba Makassar, Makassar, Indonesia*
[2,5]*Information Technology Education, Universitas Teknologi Akba Makassar, Makassar, Indonesia*
[3]*Information Technology, Universitas Teknologi Akba Makassar, Makassar, Indonesia*
[4]*Information Systems, Universitas Teknologi Akba Makassar, Makassar, Indonesia*
[6]*Department of Information Technology, Sathyabama University, India*

Abstract

The widespread adoption of electronic learning (e-learning) in higher education has brought significant changes to how knowledge is delivered. Despite its advantages, many implementations remain focused solely on content dissemination, often neglecting learners' emotional engagement. Emotional states, particularly in academic contexts, influence concentration, motivation, and comprehension. One of the most effective and intuitive indicators of emotion is facial expression. This research investigates the use of Convolutional Neural Networks (CNN), a deep learning approach, to automatically detect student emotions through facial image analysis. A dataset of facial expressions was constructed and divided into training and testing sets, each containing five distinct emotional categories: anger, happiness, fear, neutrality, and surprise. The CNN model was trained for 100 epochs, resulting in a training accuracy of 89% and a testing accuracy of 88%. These results demonstrate that CNN-based emotion recognition has strong potential to enhance e-learning platforms by providing instructors with real-time emotional insights. By integrating emotional feedback, educators can adapt instructional strategies more effectively to improve student engagement and learning outcomes. This study contributes to the growing field of affective computing and emphasizes the importance of emotional awareness in digital learning environments.

Keywords: Affective Computing; Emotion Recognition; Facial Expression Analysis; Deep Learning; Convolutional Neural Network.

Received: 25 February 2025 Revised: 29 April 2025 Accepted: 5 June 2025

Introduction

Education plays a central role in shaping the character, identity, and norms embedded in the lives of individuals and society at large (Tilaar, 2004; Wahyudin & Suryadi, 2022). The function of education extends beyond the mere transfer of knowledge; it also serves as a means to instill social and moral values, especially within higher education institutions that serve as a platform for nurturing critical and integrity-driven academic individuals (Azhary et al., 2018; Azis, 2022). Along with the advancement of time and technological progress, the paradigm of education has undergone a transformation, one of which is marked by the emergence of electronic learning (e-learning) systems that have become increasingly popular and widely implemented across higher education institutions (Mukhtar et al., 2020; G. D. Putra & Purwaningsih, 2021).

Electronic learning offers spatial and temporal flexibility that supports the effectiveness of the teaching and learning process, particularly in the context of distance education (Arkorful & Abaidoo, 2015). However, in practice, the implementation of e-learning is still dominated by conventional approaches that focus solely on the digital delivery of instructional materials without considering the affective aspects of learners, including their emotional states (Hodges et al., 2020). Numerous studies have shown that emotions have a significant impact on learning motivation, attention, engagement, and academic achievement (Pekrun et al., 2017; D'Errico et al., 2016). Positive emotions can enhance thinking capacity, creativity, and problem-solving abilities in the learning process (Isen, 2008; Adesola et al., 2019; MacAllister, 2025; Wang et al., 2023; Wu & Yu, 2022).

*Corresponding author.
E-mail address: muhriyaldipratama9@gmail.com (Muh. Riyaldi Pratama)

In the context of online learning, the lack of physical interaction makes it difficult for educators to directly perceive students’ emotional expressions (Zhu et al., 2021). Therefore, innovative approaches are needed to recognize and understand students’ emotional states as a basis for improving the quality of online learning (X. Li et al., 2021). One such approach is facial expression recognition, which is the most natural, immediate, and honest form of non-verbal communication in expressing a person’s feelings. Facial expressions can even convey deeper meanings than verbal expressions (Shan & Eliyas, 2024; Oliver & Amengual Alcover, 2020).

Facial expressions play an important role in depicting a person's emotional state. When an individual experiences emotional changes, their face will automatically display certain physical changes, such as wrinkles on the forehead, changes in skin tone, or movements in specific facial muscles (Alexander, 2019). For example, anger is characterized by furrowed eyebrows and a reddened face, while fear is indicated by widened eyes and tensed lips. By recognizing these expressions, educators can gain valuable insights into learners’ emotional readiness and engagement in the online learning process (Gupta et al., 2022; Ngo et al., 2024).

Research on facial recognition and expression analysis has rapidly evolved in line with the advancement of computer technology and artificial intelligence. This technology has been applied in various fields, including security systems, attendance tracking, customer service, and healthcare. In the medical field, facial expression recognition technology is used to diagnose patients' psychological conditions based on their facial reactions. Applying similar technology in education can help create learning systems that are more adaptive and responsive to students' emotional states (Y. Li et al., 2020; Zhou et al., 2023).

One of the most popular methods in facial expression recognition is the Convolutional Neural Network (CNN), a deep learning-based machine learning approach. CNN excels in automatically and accurately recognizing visual patterns from image data without requiring manual feature extraction (Grover & Bansal, 2025). This approach is highly relevant for facial expression recognition because it can detect a variety of emotional expressions with high accuracy, even when working with complex and unstructured datasets (Eka Putra, 2016; Putra, 2015; Ullah et al., 2024).

Therefore, the application of CNN in detecting students’ facial expressions during online learning is a strategic step towards promoting a more humanistic and emotion-based learning approach. The designed system is capable of classifying students’ emotions-such as happiness, fear, surprise, anger, and neutrality in real time. This can be utilized by educators as feedback to adjust teaching strategies, instructional delivery methods, or even implement psychological interventions when necessary.

Based on this background, this study aims to develop a CNN-based facial expression detection system for students in the context of online learning. The system is expected to serve as a tool to automatically, accurately, and in real-time understand students’ emotional conditions, thereby supporting the creation of a more effective, empathetic, and adaptive learning process that meets the needs of learners in today’s digital era.

Method

In this study, the method used to detect students' facial expressions is the Convolutional Neural Network (CNN), which is one of the deep learning approaches proven to be effective in image data analysis (Grover & Bansal, 2025). CNN is designed to automatically and adaptively learn hierarchical spatial features from images, making it highly suitable for facial expression classification tasks (Ullah et al., 2024). This method utilizes an artificial neural network architecture consisting of several layers, such as convolutional layers, pooling layers, and fully connected layers, to extract, abstract, and classify patterns from input images (Khan et al., 2021).

Theoretically, CNN operates by convolving input data (images) using a number of filters or kernels to produce feature maps, which are then reduced through a pooling process to decrease dimensionality while preserving essential features (Zhao et al., 2020). This information is further processed through deeper layers to construct more complex and accurate representations. Finally, the last layer (fully connected) is responsible for classifying the images into specific categories, in this case, emotions such as happy, fearful, angry, surprised, and neutral (S. Li & Deng, 2020).

There are several sequential stages involved in the operation of the system designed to detect facial expressions. The working architecture of the Convolutional Neural Network (CNN), which serves as the core method in this facial expression recognition research, is composed of multiple layers, each performing a specific function essential to the overall classification process. These stages are integral in training the system to recognize and differentiate between various facial expressions. The overall workflow of the system is illustrated in Figure 1, which outlines the key stages of the facial expression recognition process using CNN.

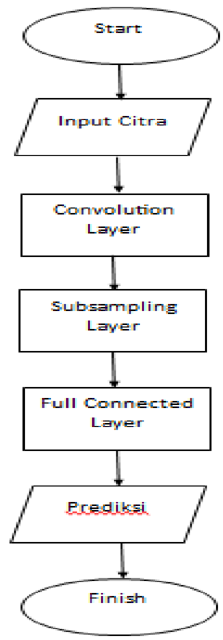

Figure 1. Flowchart CNN

As shown in Figure 1, the process begins with the input of facial images containing emotional expressions. These images are then processed through the CNN framework, starting with convolutional and activation layers, followed by a subsampling (pooling) layer that reduces dimensionality while retaining critical features. The extracted features are subsequently passed into a fully connected layer, where classification and prediction of facial expressions occur. For data acquisition, this study utilized a custom dataset captured using the Vivo V9 smartphone camera. The dataset consists of labeled images representing five distinct emotional states: neutral, happy, angry, afraid, and surprised.


Figure 2. Example of a personalized dataset

Figure 2 illustrates the preprocessing phase applied to the dataset. At the outset, the dataset was relatively limited in size and quality, with several images containing unnecessary background elements that could interfere with the model's ability to accurately recognize facial expressions. To improve the effectiveness of the system, a preprocessing strategy was employed to enhance image clarity and focus. This stage involved two main techniques: cropping, which isolates the facial region from irrelevant surroundings, and augmentation, which expands the dataset by generating varied versions of the original images. These methods collectively help improve the accuracy and robustness of the classification model.

Table 1.Cropping

Before	After
	

As shown in Table 1, the cropping process is carried out to assist the system in more accurately identifying facial expressions depicted in the images. This step plays an important role in enhancing the model's performance during training and validation by removing irrelevant elements from the image. Cropping involves trimming away non-essential parts, such as dominant backgrounds or distracting objects that may interfere with the classification process. By isolating the facial region, the system can focus on the key features needed for accurate emotion recognition. The sample images used for cropping in this study were captured using a smartphone camera.

Table 2. Resize

Before	After
	

As shown in Table 2, the resizing process addresses the inconsistency in image dimensions resulting from facial detection. Since the input images used in this study vary in pixel size, resizing is essential to standardize the dimensions before they are processed during the testing phase. Ensuring uniform image sizes across both training and testing datasets helps maintain consistency and improves computational efficiency. Without resizing, variations in pixel dimensions could negatively impact model performance and extend processing time. Additionally, larger image resolutions require more resources and time to process. An example of an image after undergoing the resizing procedure is provided below. The grayscale conversion stage is also applied, where RGB images are transformed into grayscale format. This conversion is necessary because color information is not a focus in this research. Instead, the study emphasizes facial structure and expression features, which can be effectively extracted using grayscale images.

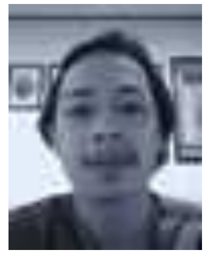

Figure 3. Grayscale

In this study, the classification process is carried out using the Convolutional Neural Network (CNN) method. Several hyperparameters are defined to optimize the performance of the model. A convolutional filter size of 3×3 is applied, while a pooling size of 2×2 is used for the subsampling layers. The stride and padding are set to their default values, 1 and 0, respectively, ensuring standard movement and no additional padding to the feature maps. The activation function plays a crucial role in introducing non-linearity into the network. In this research, the Rectified Linear Unit (ReLU) function is used within the convolutional layers. The ReLU function is defined mathematically as $\sigma(a) = \max(0, a)$, where a is the result from the linear combination of input, weights, and bias. For the output layer, the Softmax function is employed to produce a probability distribution across the predefined emotion categories, as this study deals with multi-class classification.

The batch size, which determines the number of data samples processed before updating the model's weights, is selected based on the available computing resources. Meanwhile, the training process is conducted over 100 epochs, each epoch representing one full pass through the entire training dataset to refine the model's learning.

Results and Discussion

Result

Following the collection of the dataset, the next step involved training and validation processes. The dataset used in this study consists of two main directories: one for training and another for testing. The training set contains 500 facial expression images, while the testing set comprises 100 images, bringing the total to 600 images. All images were captured under favorable conditions, including adequate lighting and an appropriate distance between the camera and the subject, ensuring clarity and usability. Examples of personal dataset images representing different emotional states are shown in the following figures. These images were used to train the model, with each emotion represented by approximately 100 samples:

Figure 4. Angry Condition

This expression is typically characterized by furrowed and drawn-together eyebrows, vertical lines between the brows, tightened lower eyelids, a sharp gaze, tightly pressed lips with downturned corners, flared nostrils, and a protruding jaw.

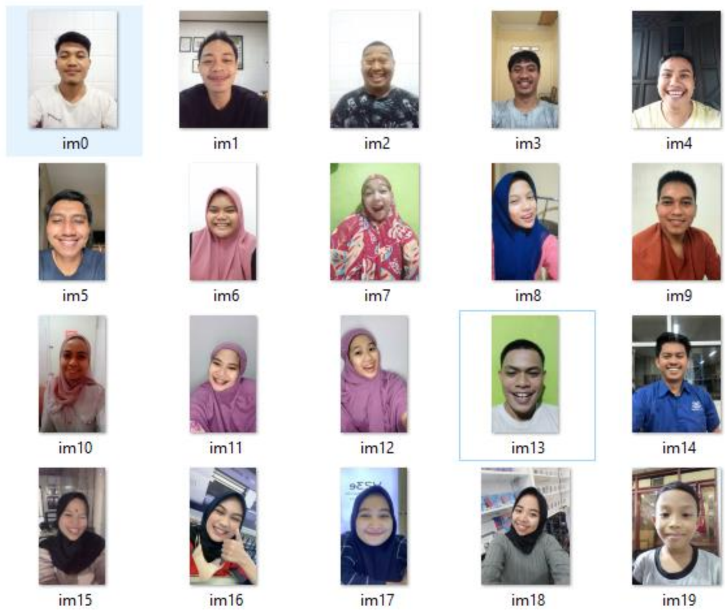

Figure 5. Happy Condition

Indicators of happiness include raised lip corners, visible teeth, wrinkles between the nose and mouth, and possibly wrinkled or tense lower eyelids.

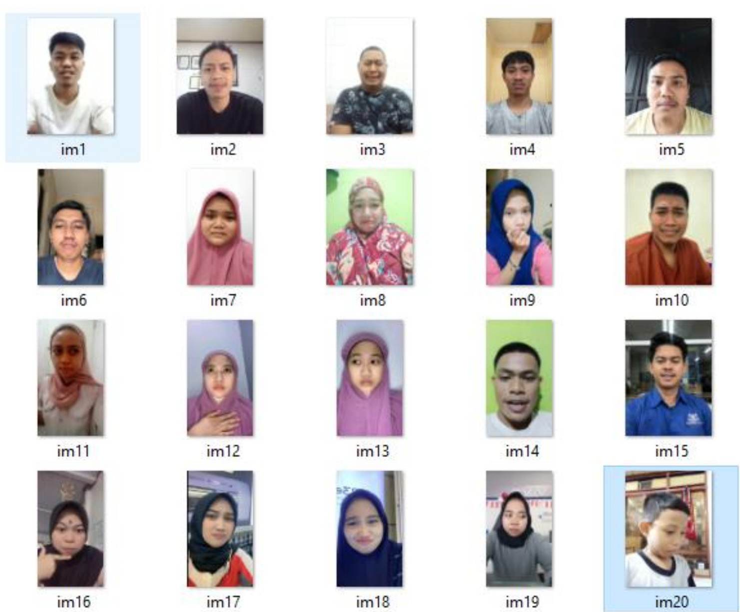

Figure 6. Fear Condition

Figure 6 shows the expression of a person in a state of fear through a training process of 100 images. Features include raised eyebrows, a horizontally wrinkled forehead, widely opened eyes, and an open mouth.

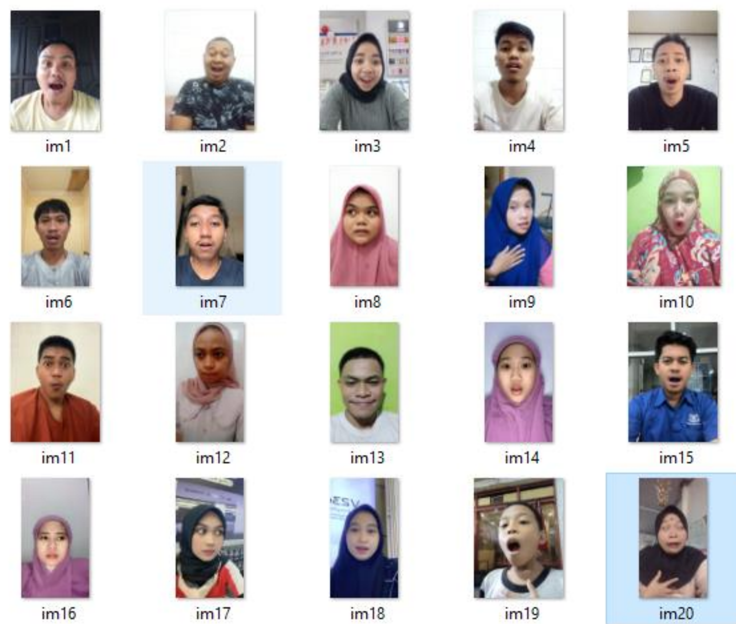

Figure 7. Surprise Condition

Figure 7 shows the expression of a person in a state of surprise through a training process of 100 images. Typical traits include arched and raised eyebrows, horizontal forehead lines, wide-open eyelids, and a lowered jaw.

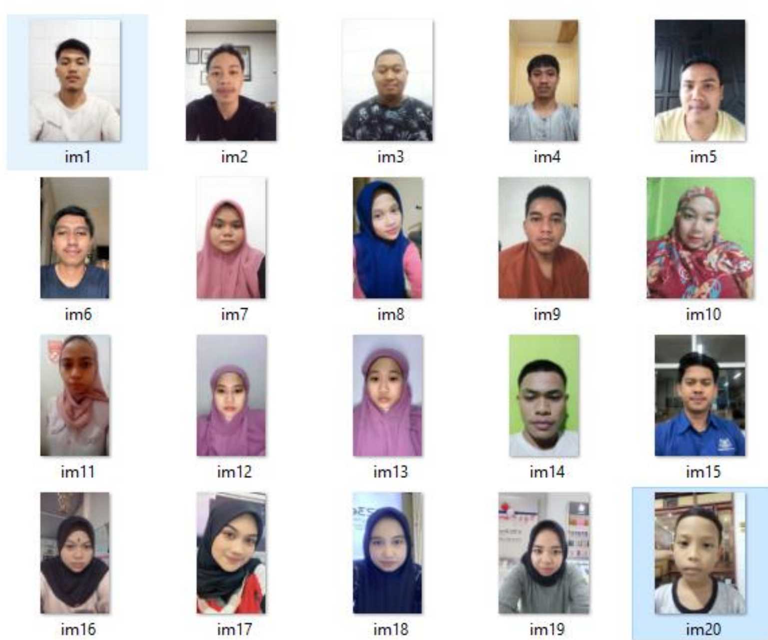

Figure 8. Neutral Condition

Figure 8 shows the expression. The face appears relaxed, with naturally open eyes and a closed mouth, often directed toward the screen. Training was conducted using 128 filters and a 3×3 kernel size over 100 epochs. The training yielded an accuracy of 64.06% with a loss value of 0.8999, as shown in, seen in Figures 18 and 19, Accuracy and validation.

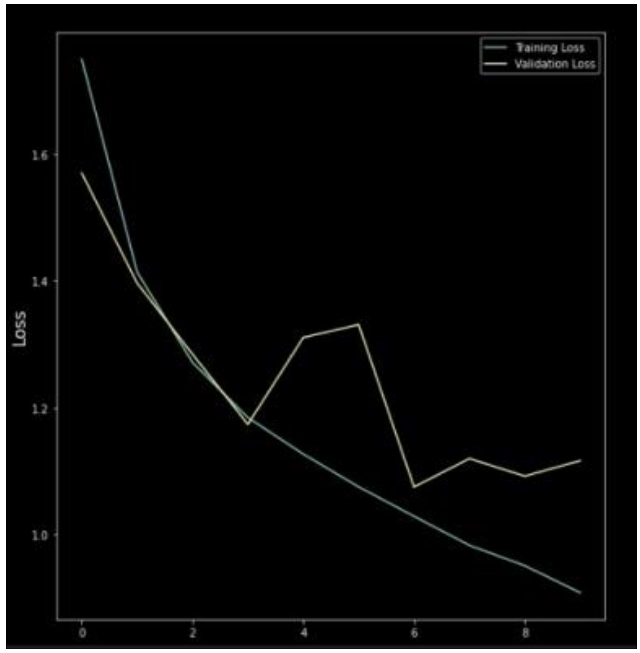

Figure 9. Training Statistik Loss

The experiment used 128 filters and a 3x3 kernel. With 100 epochs, it resulted in a train accuracy of 0.6406; train loss of 0.8999.

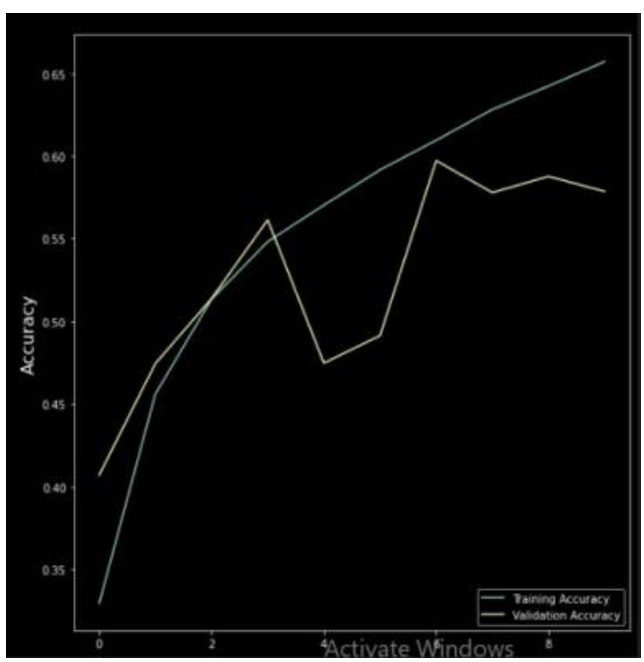

Figure 10. Validation Accuracy

The architecture of the model that will be applied is in the table 3 below.

Table 3. Architecture

Process	Parameter
Convolution	Filter = 64, Kernel = (3,3), Padding = 1, Shape (48,48,1)
Batch Normalization	
Activation	Relu
Max Pooling 2D	Pool_size = (2,2)
Dropout	0.25
Convolution	Filter = 128, kernel= (5,5), Padding = 1 ,Shape (48,48,1)
Batch Normalization	
Process	Parameter
Activation	relu
Max Pooling 2D	Pool_size = (2,2)
Dropout	0.25
Convolution	Filter = 512, Kernel = (3,3), Padding = 1, Shape (48,48,1)
Batch Normalization	
Activation	Relu
Max Pooling 2D	Pool_size = (2.2)
Dropout	0.25
Convolution	Filter = 512, Kernel = (3.3), Padding = 1, Shape (48,48,1)
Batch Normalization	
Activation	Relu
Max Pooling 2D	Pool_size = (2.2)
Dropout	0.25
Flatten	
Dennse	256
Batch Normalization	
Activation	Relu
Dropout	0.25
Dense	Activation = Softmax

To assess model performance, a confusion matrix was generated, as shown in Figure 11, allowing evaluation of classification results across multiple emotion classes.

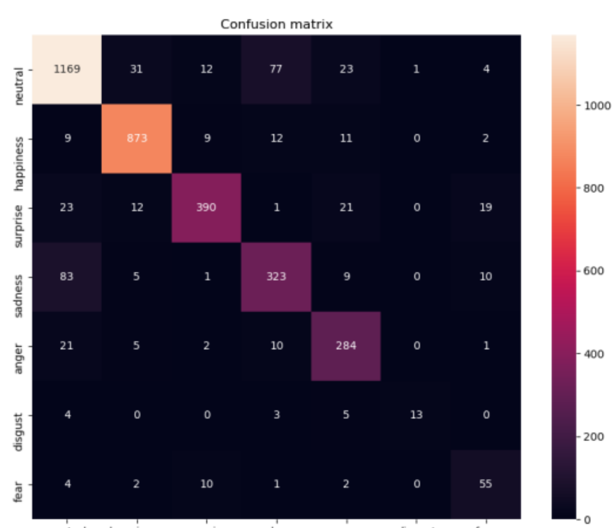

Figure 11. Confusion Matrix

Testing was conducted at three different distances, 30 cm, 50 cm, and 70 cm between the subject and the camera. The objective was to observe the system's ability to detect facial expressions accurately at varying proximities. Detection duration was 3–5 seconds per attempt. The results are detailed in Table 4.

Table 4. Object Distance Testing

Test	Test Distance (cm)														
	30					50					70				
	A	N	H	F	S	A	N	H	F	S	A	N	H	F	S
1	√	√	√	√	√	√	√	√	–	–	–	√	–	–	–
2	√	√	√	–	√	√	–	√	√	√	√	–	√	–	–
3	–	√	√	√	√	√	√	–	–	√	–	–	–	–	–
4	√	√	√	–	√	–	–	√	–	√	–	√	–	–	–
5	√	–	√	√	√	√	–	√	√	√	–	√	–	–	–
6	√	√	√	√	–	–	√	√	–	–	–	√	–	–	–
7	√	√	√	–	√	√	–	√	√	√	–	√	–	–	–
8	√	–	√	√	–	√	√	–	–	–	–	–	–	–	–
9	√	√	√	–	√	–	–	–	–	√	√	√	–	–	–
10	√	√	√	–	√	√	–	√	√	–	–	√	–	–	–
Rata-rata (%)	80 %					56 %					20 %				

Note:

√ = Accurate Detection
– = Inaccurate Detection
A = Angry, N = Neutral, H = Happy, F = Afraid, S = Surprised

From the results, it was observed that detection accuracy decreases as the distance increases. At 30 cm, the accuracy reached 80%, while at 70 cm, it dropped significantly to 20%. This suggests that closer proximity between the subject and the camera enhances facial feature clarity, leading to better recognition performance.

Table 5. Expression Detection Accuracy

Expression	Result
Happy	Successfully Detected
Angry	Successfully Detected
Surprised	Successfully Detected
Afraid	Successfully Detected
Neutral	Successfully Detected

Table 6. Multiple Face Detection Test

Number of People	Result
Two	Successfully Detected
Three	Successfully Detected
Four	Successfully Detected
Five	Successfully Detected

These results indicate that the model performs reliably even when detecting multiple faces within the same frame, provided image quality and lighting conditions are optimal.

Discussion

This study evaluated the ability of a Convolutional Neural Network (CNN)-based system to recognize facial expressions from multiple individuals within a single image frame. The results indicated that the system successfully identified and bounded each face with a detection box and accurately predicted the corresponding emotional state. The tested emotions included happy, angry, surprised, and fearful expressions. This evaluation also included testing on group images to assess the system's capability in recognizing multiple objects simultaneously. The results of the classification predictions are presented in Table 7, which compares the actual emotion labels with the predicted outputs and their respective accuracy scores:

Table 7. Emotion Prediction Results

Real Emotion	Angry	Neutral	Happy	Fear	Surprised	Accuracy
Angry	18	0	0	2	0	90.00%
Neutral	0	19	0	0	1	95.00%

Real Emotion	Angry	Neutral	Happy	Fear	Surprised	Accuracy
Happy	0	0	20	0	0	100.00%
Fear	3	3	0	13	1	65.00%
Surprised	0	1	1	0	18	90.00%
Average Accuracy						88.00%

The Description:
Accuracy is the percentage of the total data that is identified and judged to be correct

$$Acuracy = \frac{TP}{Total\ dataset} \times 100\% \quad (1)$$

$$= \frac{88}{100} \times 100\%$$

$$= 88\%$$

Precision is the data that is retrieved based on less or incorrect, or imprecise information.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Precision Per Class:

$$P(A) = \frac{18}{18+3} = 0{,}86\ P(B) = \frac{19}{19+4} = 0{,}82\ P(C) = \frac{20}{20+1} = 0{,}96$$

$$P(D) = \frac{13}{13+2} = 0{,}87\ P(E) = \frac{18}{18+2} = 0{,}9$$

All Precisions:

$$All\ Precisions = \frac{(P(A) + P(B) + P(C) + P(D) + P(E)}{Number\ of\ Classes} \quad (3)$$

$$= \frac{0{,}86 + 0{,}82 + 0{,}96 + 0{,}87 + 0{,}9}{5}$$

$$= 0{,}89\%$$

Recall is the data that cannot be predicted correctly.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Recall Per Class:

$$R(A) = \frac{18}{18+2} = 0{,}9\ R(B) = \frac{19}{19+1} = 0{,}95\ R(C) = \frac{20}{20+0} = 1$$

$$P(A) = \frac{13}{13+7} = 0{,}65\ R(B) = \frac{18}{18+2} = 0{,}9$$

All Recall:

$$All\ Recall = \frac{(R(A) + R(B) + R(C) + R(D) + R(E)}{Number\ of\ Classes} \quad (5)$$

$$= \frac{0{,}9 + 0{,}95 + 1 + 0{,}65 + 0{,}9}{5}$$

$$= 0{,}88\%$$

The analysis of the confusion matrix and metrics above confirms that the CNN model demonstrated reliable performance in emotion classification, achieving 88% accuracy, 0.88 precision, and 0.88 recall across five emotion classes. Notably, the model showed perfect accuracy in detecting happy expressions and high accuracy for angry, surprised, and neutral emotions. The lowest performance was found in identifying fearful expressions, likely due to subtle facial cues that can resemble other emotional states.

These findings highlight the practical potential of CNN-based facial expression recognition systems in educational environments. The system provides a means for instructors to monitor student emotions in real time, enhancing engagement and enabling adaptive teaching strategies tailored to emotional feedback. This is especially important in online learning contexts where non-verbal cues are limited. Similar studies by Gupta et al. (2022) and Zhou et al. (2023), have also demonstrated the effectiveness of emotion recognition systems in supporting student interaction and psychological monitoring. The integration of such technology into digital learning platforms could help address emotional disconnection, one of the core limitations of e-learning.

In conclusion, this research contributes to the growing field of affective computing in education. By developing a CNN-based emotion recognition system, this study opens pathways for real-time emotion-aware learning tools that support both students and educators. Future developments could explore deeper emotional classification, real-time integration with learning management systems (LMS), and broader datasets to improve generalization and robustness in diverse classroom scenarios.

Conclusions and Suggestions

Conclusions

This study successfully developed a deep learning model based on the Convolutional Neural Network (CNN) to classify facial expressions into five categories: neutral, angry, happy, sad, and surprised. The model was trained using a custom dataset, and evaluation through performance metrics showed that the model achieved an accuracy of 88%. These results indicate that the CNN-based model has a reliable level of accuracy in detecting and classifying emotional expressions, making it a promising approach for emotion recognition in educational and interactive digital environments.

Suggestions

Future research could benefit from incorporating additional comparative methods, such as other neural network architectures or transfer learning techniques, to evaluate and enhance model performance. Improvements can also be made by expanding the dataset to include a more diverse set of facial expressions under various conditions. Additionally, enhancing the quality of the input images, particularly by using higher-resolution cameras, may lead to more precise recognition outcomes. These refinements are expected to improve the robustness, accuracy, and real-world applicability of the facial expression detection system.

References

Adesola, S. A., Li, Y., & Liu, X. (2019). Effect of Emotions on Students Learning Strategies. *Proceedings of the 2019 8th International Conference on Educational and Information Technology*. https://doi.org/10.1145/3318396.3318408

Alexander, D. (2019). *Deteksi Emosi Menggunakan Citra Ekspresi Wajah Secara Otomatis*.

Arkorful, V., & Abaidoo, N. (2015). The role of e-learning, advantages and disadvantages of its adoption in higher education. *Procedia - Social and Behavioral Sciences*, *3*, 29–35. https://doi.org/10.1016/j.protcy.2014.10.067

Azhary, L., Handoyo, E., & Khafid, M. (2018). *The Implementation of Integrated Character Education in Policy Design at SD Muhammadiyah (Plus) Salatiga*.

Azis, M. (2022). The Role of Higher Education in Building Student Character: A Moral and Religious Approach. *Jurnal Pendidikan Ilmu-Ilmu Sosial*, *11*(2), 151–160. https://doi.org/10.15294/jpii.v11i2.38439

D'Errico, F., Paciello, M., & Cerniglia, L. (2016). When emotions enhance students' engagement in e-learning processes. *Journal of E-Learning and Knowledge Society*, *12*, 9–23. https://doi.org/10.20368/1971-8829/1144

Eka Putra, W. S. (2016). Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101. *Jurnal Teknik ITS*, *5*(1). https://doi.org/10.12962/j23373539.v5i1.15696

Grover, R., & Bansal, S. (2025). Efficient Facial Expression Recognition Through Lightweight CNN Technique on Public Datasets. *SN Computer Science*, *6*(15). https://doi.org/10.1007/s42979-024-03557-y

Gupta, S., Kumar, P., & Tekchandani, R. K. (2022). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications*, *81*, 11365–11394. https://doi.org/10.1007/s13278-023-01181-x

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Australasian Journal of Educational Technology*, *36*(3), 20–34. https://doi.org/10.1080/01587919.2020.1756120

Isen, A. M. (2008). Positive affect and decision processes: Some recent theoretical developments with practical implications. *American Psychologist*, *63*(10), 791–803. https://doi.org/10.1037/0003-066X.63.10.791

Khan, M. A., Akram, T., Sharif, M., Javed, K., & Raza, M. (2021). Facial expression recognition using optimized deep CNN features. *Sensors*, *21*(16), 5386. https://doi.org/10.3390/s21165386

Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *Neurocomputing*, *429*, 215–233. https://doi.org/10.1016/j.neucom.2020.01.081

Li, X., Yue, R., Jia, W., Wang, H., & Zheng, Y. (2021). Recognizing Students' Emotions based on Facial Expression Analysis. *2021 11th International Conference on Information Technology in Medicine and Education (ITME)*, 96–100. https://doi.org/10.1109/ITME53901.2021.00030

Li, Y., Song, Y., Luo, Z., Liu, H., & Liu, Q. (2020). Facial Expression Recognition for Intelligent Healthcare: A Review. *Frontiers in Psychiatry*, *11*, 607890. https://doi.org/10.3389/fpsyt.2020.607890

MacAllister, G. P. (2025). The Impact of Emotion-Based Teaching Strategies on Motivation in Collaborative Learning. *Journal of Advanced Research in Education*. https://doi.org/10.56397/jare.2025.01.02

Mukhtar, K., Javed, K., Arooj, M., & Sethi, A. (2020). Advantages, Limitations and Recommendations for online learning during COVID-19 pandemic era. *Heliyon*, *6*(9), e03163. https://doi.org/10.1016/j.heliyon.2020.e03163

Ngo, D., Nguyen, A., Dang, B., & Ngo, H. (2024). Facial Expression Recognition for Examining Emotional Regulation in Synchronous Online Collaborative Learning. *International Journal of Artificial Intelligence in Education*, *34*, 650–669. https://doi.org/10.1007/s40593-023-00378-7

Oliver, M. M., & Amengual Alcover, E. (2020). UIBVFED: Virtual facial expression dataset. *PLOS ONE*, *15*(4), e0231266. https://doi.org/10.1371/journal.pone.0231266

Pekrun, R., Loderer, K., & Götz, T. (2017). Affective science in education: Theoretical contributions and empirical applications. *Contemporary Educational Psychology*, *54*, 1–4. https://doi.org/10.1016/j.cedpsych.2016.10.002

Putra, G. D., & Purwaningsih, Y. (2021). E-learning as a transformation in higher education during the Covid-19 pandemic. *ETERNAL (English Teaching Journal)*, *12*(2), 212–223. https://doi.org/10.26877/eternal.v12i2.8320

Putra, I. N. T. A. (2015). Perancangan dan Pengembangan Sistem Absensi Realtime Melalui Metode Pengenalan Wajah. *JST Jurnal Sains Dan Teknologi*, *3*(2). https://doi.org/10.23887/jst-undiksha.v3i2.4480

Shan, J., & Eliyas, S. (2024). Exploring AI Facial Recognition for Real-time Emotion Detection: Assessing Student Engagement in Online Learning Environments. *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, 1–6. https://doi.org/10.1109/AIIoT58432.2024.10574587

Tilaar, H. A. R. (2004). *Pendidikan, Kebudayaan, dan Masyarakat Madani Indonesia*. Remaja Rosdakarya.

Ullah, S., Ou, J., Xie, Y., & Tian, W. (2024). Compound Facial Expressions Recognition Approach Using DCGAN and CNN. *Multimedia Tools and Applications*, *83*, 85703–85723. https://doi.org/10.1007/s11042-024-20138-6

Wahyudin, A. Y., & Suryadi, A. (2022). Educational values and character building in the Indonesian education system. *Lingua Cultura*, *18*(1), 1–8. https://doi.org/10.23917/lingua.v18i1.18635

Wang, H., Wang, Y., & Li, S. (2023). Unpacking the relationships between emotions and achievement of EFL learners in China: Engagement as a mediator. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1098916

Wu, R., & Yu, Z. (2022). Exploring the effects of achievement emotions on online learning outcomes: A systematic review. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.977931

Zhao, X., Peng, Y., Song, Y., & Song, Y. (2020). Feature extraction and facial expression recognition based on CNN. *Computer Methods and Programs in Biomedicine*, *197*, 105469. https://doi.org/10.1016/j.cmpb.2020.105469

Zhou, Y., Liu, Q., Lin, Y., Yang, H., Tang, Q., & Wang, W. (2023). Emotion Recognition in Education: A Facial Expression Recognition Approach for Real-Time Student Engagement Detection. *Journal of Intelligent and Robotic Systems*, *108*(43). https://doi.org/10.1007/s00542-023-05420-1

Zhu, L., Yan, Y., Liu, S., & Tang, X. (2021). The Research of Online Classroom Status Feedback Based on Facial Expression Recognition Technology. *2021 2nd International Conference on Information Science and Education (ICISE-IE)*, 1329–1332. https://doi.org/10.1109/icise-ie53922.2021.00297