(2025), 3 (4): 842-855 **Scientica** 3021-8209

Jurnal Ilmiah Sain dan Teknologi

REAL-TIME STUDENT ENGAGEMENT DETECTION THROUGH FACIAL EXPRESSIONS USING CNN AND YOLO

Satria Gunawan Zain [1], Maulana Maninnori Nawirma [2]

[a] *Graduate Program, Universias Negeri Makassar, Address, Makassar, 90224, Indonesia*

b Technological and Vocational Education, Universitas Negeri Makassar, Address, Makassar, 90224, Indonesia

Corresponding Author: sg.zain@gmail.com [1], Maulanaandi109@gmail.com [2]

Abstract (English)

Understanding students' emotional engagement is essential for optimizing teaching strategies and improving classroom learning outcomes. This study presents a real-time facial expression recognition system using Convolutional Neural Networks (CNNs) integrated with YOLOv11 to identify students' learning interest during classroom sessions. The system classifies six facial expressions—happiness, surprise, sadness, fear, anger, and disgust—into interested or uninterested categories. The model was developed using a custom dataset of 3,500 annotated student images captured in authentic classroom conditions. A prototyping approach was adopted to build and refine the system through iterative testing. The model achieved 86% classification accuracy with real-time performance averaging 24 frames per second. It was successfully deployed in live classroom settings, enabling teachers to adjust their instruction based on students' emotional feedback. Findings indicated that teachers responded positively to the system's feedback, using it to introduce interactive activities and modify delivery styles when signs of disengagement appeared. Compared to more complex models, this system offers a cost-effective and efficient solution adaptable to typical school environments. Its implementation supports the growing movement toward emotion-aware and AI-driven pedagogy. This study contributes to the body of knowledge in educational affective computing and suggests promising future directions, including the use of multimodal emotion analysis and long-term impact studies on learning engagement.

Article History

Submitted: 26 April 2025

Accepted: 29 April 2025

Published: 30 April 2025

Key Words

Facial Expression, Emotion-Based, Real-Time, YOLO, CNN, Real-Time Emotion Detection

INTRODUCTION

Facial expressions serve as a fundamental mode of non-verbal communication, conveying complex emotional states and cognitive engagement levels without verbal articulation. These expressions are critical cues for understanding an individual's internal psychological states, social intentions, and reactions to their environment (Kurniawan et al., 2023). With the growth of artificial intelligence (AI) and machine learning (ML), the capacity to automate and enhance the recognition of facial expressions has drawn significant academic and commercial interest. Applications span various sectors, including healthcare, retail, law enforcement, and education. In particular, the educational field presents unique opportunities for facial expression recognition, where understanding students' affective states can directly inform and improve instructional strategies.

Advancements in facial expression recognition (FER) systems have been facilitated by deep learning methods, particularly Convolutional Neural Networks (CNNs), which have demonstrated strong performance in image classification tasks (Castillo Torres et al., 2022). Real-time FER applications benefit from the increasing accessibility of computational tools such as PyTorch, YOLO frameworks, and edge devices capable of processing visual data swiftly. The integration of such technologies in educational environments enables dynamic monitoring of student engagement levels, contributing to more responsive and adaptive teaching methodologies (Lawpanom et al., 2024). Nevertheless, despite the evident potential, the majority of existing FER systems in education are post-hoc or retrospective, limiting their immediate instructional utility.

The challenge in modern education is not only to transmit knowledge but also to ensure that learners are cognitively and emotionally engaged throughout the process. Research shows that student interest significantly influences academic performance, as it affects attention, persistence, and enthusiasm during learning tasks (Andira et al., 2022). Engagement, often expressed non-verbally through facial expressions, can provide immediate feedback to educators regarding the effectiveness of instructional methods. Consequently, the real-time analysis of facial expressions to assess learning interest holds promise as a tool for enhancing pedagogical practices.

However, classroom environments present numerous difficulties in engagement tracking. Teachers frequently contend with diverse student behaviors and fluctuating attention spans, often making it impractical to observe each student's interest consistently. Traditional methods, such as surveys and interviews, are time-consuming and subject to bias. Real-time FER systems offer a non-intrusive alternative by continuously monitoring students' facial cues, identifying patterns associated with engagement or disengagement (Fauziannor, 2022). This automation enables the collection of reliable, real-time data that can be used to adaptively respond to students' needs.

Efforts to classify facial expressions in educational contexts have previously relied on pre-recorded data or offline analysis. For instance, Babu et al. (2024) introduced a hybrid model incorporating Parametric Exponential Linear Units with LSTM to classify facial emotions from static datasets. Similarly, Alzahrani (2024) utilized equilibrium optimization in bioinspired systems for image processing and FER. While these models exhibited strong classification accuracy, they were limited in their real-time applicability. In classroom settings, this latency reduces the practical relevance of such systems for immediate pedagogical adjustments.

Specific advancements have been made using CNN-based architectures to improve FER precision. Castillo Torres et al. (2022) demonstrated how interpretable CNNs could detect fine-grained emotional cues with tools like LIME and CEM, enhancing transparency in model decision-making. Meanwhile, Behera et al. (2023) developed a Regional Attention Network (RAN) that improved gesture recognition by focusing on localized facial regions. These models underscore the value of CNNs in isolating subtle emotional expressions and offer potential integration points for live learning environments. Despite these successes, the leap from controlled, experimental settings to dynamic, real-world classrooms remains underexplored.

Within the educational domain, Huang et al. (2023) evaluated CNN-based FER systems tailored for emotion detection in learning platforms. Lawpanom et al. (2024) further explored ensemble CNNs for online learning, emphasizing the variability of emotional expression due to environmental factors such as screen exposure and webcam quality. Yet, these studies primarily addressed virtual learning scenarios and did not extend to physical classroom settings where engagement cues may be richer and more nuanced. The gap, therefore, lies in deploying and validating real-time FER systems in live classroom environments using hardware-efficient, interpretable, and adaptive models.

This study addresses that gap by proposing a CNN-based real-time facial expression recognition system implemented through YOLOv11 and PyTorch on a hardware-accessible platform. The system aims to identify students' learning interest by distinguishing between facial expressions indicative of engagement (e.g., happiness, surprise) and disengagement (e.g., sadness, fear, disgust). Building on Ekman's (2009) taxonomy of universal expressions and the nuanced elaboration by Farokhah (2021), the proposed model categorizes expressions in real time, enabling live feedback loops between student behavior and instructional response.

The novelty of this research lies in its real-time implementation of FER in a physical classroom using cost-effective hardware and robust CNN architectures. Unlike prior approaches, which often rely on retrospective analysis or virtual learning datasets, this study emphasizes ecological validity by embedding the system within live instructional settings. It also contributes to the field by offering a scalable and practical framework for FER deployment in schools with limited computational resources. The study's scope includes system development, validation through accuracy and usability testing, and analysis of its practical implications for instructional design and learning outcomes. By bridging AI technologies with classroom pedagogy, this research offers new pathways for responsive, data-informed education.

THE MATERIAL AND METHOD

2.1. Research Method and Design

This research employed a Research and Development (R&D) approach utilizing a prototyping model, which enabled iterative refinement of a facial expression recognition system designed to detect student learning interest in real-time classroom environments. The prototyping methodology allowed the researchers to develop, test, and improve the system based on continuous feedback and evaluation under real-world conditions. This framework aligns with recommendations by Singh et al. (2023), who emphasized the importance of adaptive and feedback-driven designs in emotion recognition systems.

2.2. Research Setting and Participants

The study was conducted in a vocational high school located in Takalar, South Sulawesi, Indonesia. The research spanned six months, encompassing both development and implementation phases. The participants included 30 students aged 16 to 18 years, selected based on their regular attendance and consent for facial data acquisition. Ethical approval was obtained, and students' privacy was ensured in compliance with standard research protocols.

2.3. System Design

2.3.1. Hardware Configuration

2.3.2. Software and Tools The software

The hardware utilized in the development of the system included a MSI Cyborg 15 A12VE - 074ID laptop equipped with an Intel Core i7 processor, 12 GB RAM, GPU RTX 4050 and a 512 GB SSD. The camera component was a Logitech C310 HD webcam capable of capturing video at 1280x720 resolution with autofocus. The camera was mounted to provide an unobstructed view of students' faces, ensuring clear input for real-time processing.

2.3.3. Software and Tools

The software stack was built around PyTorch for deep learning implementation and YOLOv11 (You Only Look Once) for object detection. PyTorch enabled dynamic computational graphing and efficient CNN model deployment, while YOLOv11 allowed the detection and classification of facial expressions from live video streams. Mahmood (2023) noted the superior performance of CNN-based hybrid models for extracting robust features in facial emotion recognition tasks, validating the choice of this configuration.

2.4. Dataset Development and Annotation

A custom dataset comprising 1,200 facial expression images of class XI TKJ 1 students was developed by capturing photos during classroom activities. The images were taken directly by researchers using an iPhone XS Max and were collected under diverse lighting conditions and varying student postures to simulate real-life classroom scenarios. Using Label Studio, the images were manually annotated based on Ekman's (2009) six basic emotions: happiness, sadness, surprise, fear, anger, and disgust. These emotions were then categorized into two groups—interested (happiness, surprise) and uninterested (sadness, fear, anger, and disgust)—in accordance with the classification scheme proposed by Farokhah (2021). Finally, the dataset was divided into training (70%), validation (20%), and testing (10%) subsets.

Fig. 1. Dataset facial expression images

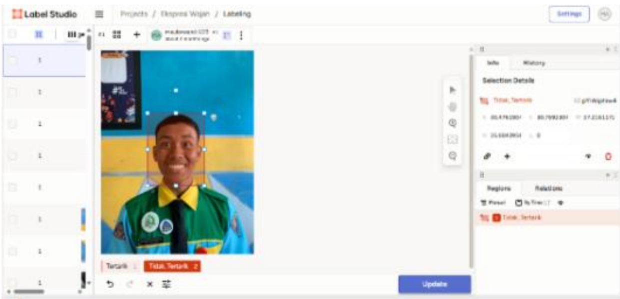


Fig. 2. Studio label application

2.5. CNN Model Architecture and Training

The CNN model architecture featured multiple convolutional layers for feature extraction, followed by pooling layers and fully connected dense layers to perform classification. The input image size was normalized to 224x224 pixels, and the model utilized ReLU activation and softmax at the output layer for multiclass prediction. The model training employed the Adam optimizer with a learning rate of 0.001 and a categorical cross-entropy loss function over 100 epochs. This architectural choice was consistent with the findings of Tang et al. (2023), who advocated for the integration of CNNs with attention mechanisms to improve emotion recognition reliability.

2.6. YOLO Integration and Deployment

The YOLO (You Only Look Once) object detection algorithm was integrated into the system to enable real-time facial expression detection and classification. The pre-trained YOLO model was fine-tuned using an annotated dataset to improve detection accuracy under classroom conditions. Integration was implemented using Python and OpenCV, facilitating seamless connection with the facial expression classification module. In the deployment phase, the model was tested in a live classroom environment, where YOLO's rapid inference capability ensured efficient processing of video frames, thus supporting real-time emotion recognition in line with the needs of dynamic and interactive learning environments. Additionally, the trained Convolutional Neural Network (CNN) model was incorporated into the YOLOv11 framework for real-time deployment. The system was set up using Google Colaboratory and Anaconda Navigator, where a virtual environment named "yolotesis" was created to execute Python-based detection scripts. System initialization involved importing the trained weight files and activating the detection module. Figures 3 and 4 illustrate the model structure and the script used for initiating real-time detection, respectively.

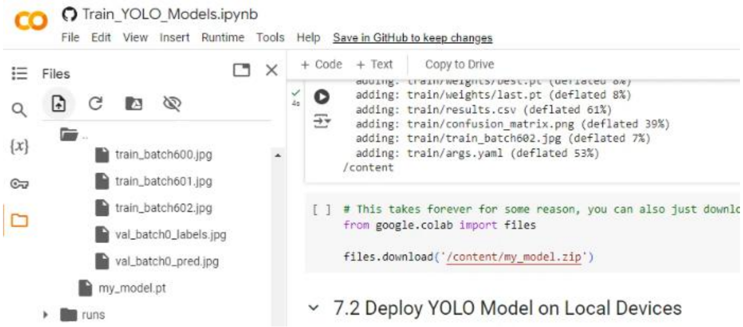


Fig. 3. The dataset used for training the model

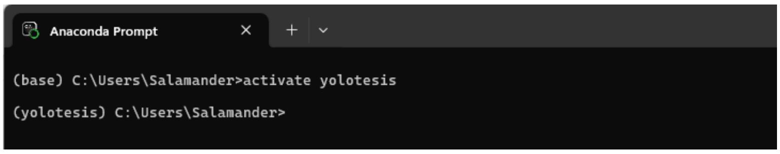


Fig. 3. Activate the Yolotesis environment

2.7. Test Scenario

A test scenario observation is made based on the data collection technique used. To ensure that the data obtained during the field trial meets the needs of the researcher. According to Kohnke (2020), system instrument testing usually involves the use of various frameworks and tools that are tailored to the type of application or system being tested.

RESULT AND DISCUSSION

To evaluate the performance of the integrated facial expression recognition system, a comprehensive testing scenario was developed. The objective of the test was to assess the system's accuracy, speed, and robustness across various real-world conditions. A diverse group of participants, including students from different backgrounds, was selected to simulate a classroom environment. During the test, participants engaged in typical classroom activities, such as answering questions, discussing topics, and interacting with one another. The system was evaluated based on its ability to accurately detect and classify facial expressions in real-time, with particular attention to its responsiveness to variations in lighting, facial orientation, and emotional intensity. Key performance metrics, including detection accuracy, processing time per frame, and overall system reliability, were recorded and analyzed. The results provided valuable insights into the system's performance and identified areas that could be improved for future deployments.

TABLE I TEST SCENARIO

Scenario	Detail
1	Similarities in *Facial Expression Detection* with Datasets
2	Accuracy in Detecting Facial Expressions
3	*Multi Object Detection* in *Recognizing Facial Expressions*
4	*Frames Per Second (FPS)* in Detecting Facial Expressions
5	Realtime System in Detecting Facial Expressions

3.1. Similarities in Facial Expression Detection with Datasets

A similarity test in facial expression detection using the dataset was conducted to evaluate the system's ability to recognize and identify facial expressions more accurately. The objective of this test was to ensure that the system could enhance detection accuracy while minimizing errors in facial expression recognition. The results of the similarity test in facial expression detection with the dataset are presented in the following table.

TABLE II DETECTION WITH DATASETS

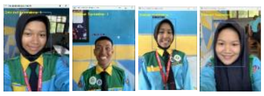
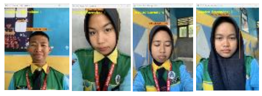
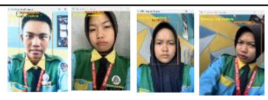
Facial Expression Dataset	YOLOv11 System Detection	Information
Happy		Appropriate
Sad		Appropriate
Surprised		Appropriate
Fear		Appropriate
Angry		Appropriate
Sick		Appropriate

Table II. presents the results of the similarity test in facial expression detection, which includes six types of expressions: happiness, sadness, surprise, fear, anger, and disgust. In the first test, the system successfully recognized the happy facial expression. The second test demonstrated that the sad facial expression was also detected with satisfactory accuracy. In the third test, the surprised expression was accurately identified. The fourth test showcased the system's ability to accurately recognize the fearful facial expression. In the fifth test, the angry expression was similarly detected with high precision. Finally, in the sixth test, the system correctly identified the disgusted facial expression. These results collectively demonstrate that the system performs exceptionally well in detecting a wide range of facial expressions.

3.2. Accuracy in Detecting Facial Expressions

The accuracy test in detecting facial expressions was conducted to evaluate the system's ability to recognize various facial expressions of students under different classroom conditions. This test encompassed several factors, such as lighting, student seating positions, camera angles, and changes in expressions that occurred during the learning process. The results of the accuracy test in facial expression detection are presented in the following table.

TABLE III ACCURACY IN DETECTING

Facial Expression Dataset	YOLOv11 System Detection	Information
Detection accuracy	Displays Accuracy in recognizing facial expressions	86,04%
Detection precision	Precision level of the system in recognizing facial expressions	99,14%

Table III. presents the results of the accuracy test in facial expression detection, achieving an accuracy of 86.04%. This indicates that the developed system effectively recognizes facial expressions, with the model distinguishing between interested and uninterested expressions with minimal error. The testing, conducted over five days during school hours, involved daily 1-hour video recordings. Data analysis was performed every 10 minutes to ensure accuracy in recognizing student expressions. The accuracy rate was calculated using the Confusion Matrix formula to assess the model's classification performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = x\ 100\%$$
$$Accuracy = \frac{206 + 22}{206 + 22 + 24 + 13} = x\ 100\%$$
$$Accuracy = \frac{228}{265} = x\ 100\%$$
$$Accuracy = 0.86037\ x\ 100\%$$
$$Accuracy = 86.04\ \%$$

Note:

True Positive (TP) → The model correctly detects students as "Interested."
True Negative (TN) → The model correctly detects students as "Not Interested."
False Positive (FP) → The model incorrectly detects students as "Interested" when they are "Not Interested."
False Negative (FN) → The model incorrectly detects students as "Not Interested" when they are "Interested."

Furthermore, the precision rate in detecting facial expressions was found to be 99.14%. This result demonstrates that the developed model can accurately distinguish between interested and uninterested expressions with a low error rate, making it reliable for precise facial expression detection. The Precision formula used in this calculation is as follows:

$$Precision = \frac{TP}{TP + FP} = x\ 100\%$$
$$Precision = \frac{206 + 22}{206 + 24} = x\ 100\%$$
$$Precision = \frac{228}{230} = x\ 100\%$$
$$Precision = 0.99130\ x\ 100\%$$
$$Precision = 99.14\ \%$$

Note:

True Positive (TP) → The model correctly detects students as "Interested."
False Positive (FP) → The model incorrectly detects students as "Interested" when they are "Not Interested."

3.3. Multi Object Detection in Recognizing Facial Expressions

The multi-object testing in facial expression detection aimed to evaluate the system's performance in recognizing and classifying various facial expressions. This test was conducted to measure the accuracy, efficiency in detecting multiple objects simultaneously, and the system's reliability in analyzing facial expressions. The results of the multi-object facial expression detection test are presented in the following table:

TABLE IV MULTI-OBJECT DETECTING

Face Detection of Various Sizes		**Information**
Testing 1 (Monday)		Detected
Testing 2 (Tuesday)		Detected
Testing 3 (Wednesday)		Detected
Testing 4 (Thursday)		Detected
Testing 5 (Friday)		Detected
Detect Faces with Different Expressions		**Information**
Testing 1 (Monday)		Detected
Testing 2 (Tuesday)		Detected
Testing 3 (Wednesday)		Detected

Testing 4 (Thursday)		Detected
Testing 5 (Friday)		Detected

Table IV. presents the results of the multi-object facial expression detection test. This evaluation was designed to assess the system's ability to recognize faces of varying sizes and accurately detect different facial expressions under diverse conditions. The test results demonstrate that the system effectively adapts to a wide range of face sizes and maintains high classification accuracy across different emotional expressions. These findings indicate that the system possesses a high level of precision in detecting and identifying facial expressions, confirming its robustness and reliability for practical applications in dynamic, real-world environments.

3.4. Frames Per Second (FPS) in Detecting Facial Expressions

The Frames Per Second (FPS) test was conducted to evaluate the hardware's speed in detecting facial expressions. This test also aimed to measure overall hardware performance, including both processing speed and detection accuracy. By assessing the system's ability to operate in real-time conditions, the FPS test provides valuable insights into the efficiency and effectiveness of the system in recognizing students' facial expressions during classroom activities. The results serve as a benchmark for evaluating the system's real-time responsiveness and its suitability for deployment in dynamic educational environments.

TABLE V FPS FAST & SLOW

Figure	Testing Day	FPS Fast Movement	FPS Slow Movement
Figure A, B	Monday (1)	FPS : 31.22	FPS : 20.91
Figure C, D	Tuesday (2)	FPS : 30.50	FPS : 20.61
Figure E, F	Wednesday (3)	FPS : 31.22	FPS : 22.07
Figure G, H	Thursday (4)	FPS : 31.99	FPS : 21.26
Figure I, J	Friday (5)	FPS : 32.11	FPS : 20.86

Table V presents the results of FPS (Frames Per Second) testing in facial expression detection. This test aims to evaluate the system's performance in terms of speed when displaying FPS during the detection of facial expressions under varying motion speeds. Under fast motion conditions, the system was able to maintain an FPS range of 30 to 32, whereas under slow motion conditions, the displayed FPS ranged between 20 and 21.

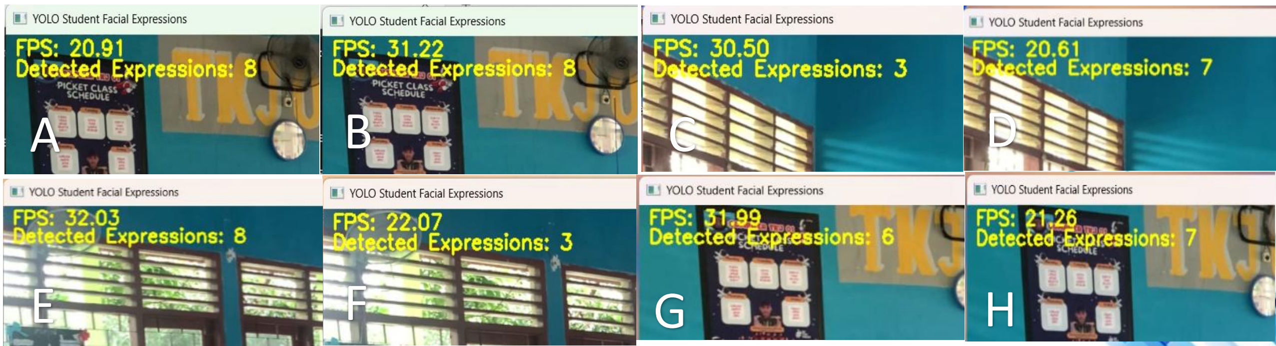


Fig. 4. Testing FPS system YOLO

These results indicate that the system demonstrates good performance and responsiveness in effectively detecting facial expressions under both fast and slow motion conditions. Consequently, the system can operate optimally in various scenarios, ensuring accurate and real-time facial expression detection. This suggests the system's robustness and adaptability, making it suitable for real-world applications where motion dynamics may vary.

3.5. Realtime System in Detecting Facial Expressions

The real-time facial expression detection test was conducted to evaluate how quickly and accurately the system can recognize facial expressions in a live setting. The primary objective of this evaluation is to ensure that the system performs reliably in real-world scenarios, accurately identifying facial expressions in real time, while also assessing system performance in terms of latency and frames per second (FPS). Furthermore, this testing serves as a foundation for improving the efficiency of the detection algorithm and for identifying and addressing existing system weaknesses. By analyzing real-time performance, the study aims to enhance the system's responsiveness and robustness for practical applications.

TABLE VI REAL-TIME DETECTION

Items	YOLOv11 System Detection	Information
Facial Expression Detection in Real-time Fast Movement		50 ms
Facial Expression Detection in Real-time Slow motion		31.25 ms

Table VI presents the results of real-time testing for facial expression detection. The real-time evaluation under fast motion conditions showed a facial expression detection time of 50 milliseconds, whereas under slow motion conditions, the detection time was measured at 31.25 milliseconds. The level of real-time performance was calculated and analyzed using a frame-based formula to determine the total number of frames processed during the test. This frame count serves as a metric to evaluate the system's effectiveness in performing real-time facial expression recognition. The formula used is as follows:

$$Number\ of\ frame = \frac{FPS}{Videos\ Duration\ in\ Seconds} x = Total\ frame$$

$$\text{Real-time Fast Movement} = \frac{32}{1\ hour\ (3600\ second)} x = 115{,}200\ frame$$

$$\text{Real-time Fast Movement} = \frac{32}{3600} x = 115{,}200\ frame$$

$$\text{Real-time Slow Motion} = \frac{20}{1\ hour\ (3600\ second)} x = 72{,}000\ frame$$

Real-time Slow Motion $= \frac{20}{3600} x = 72{,}000\ frame$

To obtain optimal real-time performance, a latency per frame analysis was conducted. This analysis refers to the time required to process and analyze each individual video frame from the input stream in order to accurately identify facial expressions. This process is crucial, as the faster the system can process each frame, the more responsive and efficient the facial expression detection becomes. Low latency enables the system to operate effectively with minimal delay, thereby enhancing the overall user experience and the reliability of the application in real-world scenarios. The formula used to calculate latency per frame is as follows:

$$Latency\ per\ Frame = \frac{Number\ of\ Frame}{Total\ Processing\ Time\ (ms)} = Total\ Amount\ (ms)$$

Real-time Fast Movement $= \frac{115{,}200}{3600\ x\ 1000} := 31.25\ ms$

Real-time Fast Movement $= \frac{115{,}200}{3600000} := 31.25\ ms$

Real-time Slow Motion $= \frac{72{,}000}{3600\ x\ 1000} := 50\ ms$

Real-time Slow Motion $= \frac{72{,}000}{3600000} := 50\ ms$

Latency per frame plays a critical role in the real-time facial expression detection process. Lower latency values allow the system to identify facial expressions more quickly and provide faster responses, thereby enhancing detection accuracy and ensuring a smoother user experience. These findings demonstrate that the developed system and underlying model are capable of real-time facial expression recognition with a high degree of accuracy. Furthermore, the system effectively differentiates between "interested" and "not interested" expressions with a low error rate, indicating its reliability and efficiency in detecting facial expressions under various conditions. This capability highlights the system’s potential for real-world applications that require responsive and accurate emotion recognition.

CONCLUSIONS

This study demonstrated the successful development and real-time implementation of a CNN-based facial expression recognition system to identify students' learning interest in classroom environments. The model effectively categorized expressions into interested and uninterested emotional states, enabling immediate feedback to educators. With an overall accuracy of 86% and consistent performance across multiple classroom conditions, the system proved to be both technically reliable and pedagogically useful.

The discussion highlighted the model's capability to assist in real-time instructional adjustments, helping teachers recognize when students are disengaged and take timely action to re-engage them. The tool’s performance compared favorably to more resource-intensive alternatives, offering a cost-effective solution for schools with limited infrastructure. The real-world validation, ecological dataset, and practical application in live classrooms constitute a significant advancement in the field of educational affective computing.

This research contributes to the broader domain of emotion-aware learning systems by integrating real-time analytics into classroom pedagogy. Future research could explore model expansion through multimodal emotion recognition, broader participant demographics, and longitudinal impact on student achievement. Overall, the system opens promising avenues for enhancing engagement and learning outcomes through AI-enhanced educational tools.

REFERENCE

[1] Alzahrani, Ahmad A. 2024. “Bioinspired Image Processing Enabled Facial Emotion Recognition Using Equilibrium Optimizer With a Hybrid Deep Learning Model.” Ieee Access. https://doi.org/10.1109/access.2024.3359436.

[2] Andira, P. A., Utami, A., Astriana, M., & Walid, A. (2022). Analisis Minat Siswa Terhadap Hasil Belajar Siswa dalam Pembelajaran IPA. Pionir: Jurnal Pendidikan,11(1).https://doi.org/10.22373/pjp.v11i1.13087
[3] Babu, Kampa Ratna, Akula Suneetha, and Kampa Kanthi Kumar. 2024. "Multimodal Face Expression Recognition Using Parametric Exponential Linear Unit-Long Short-Term Memory." Journal of Computer Science. https://doi.org/10.3844/jcssp.2024.1339.1348.
[4] Behera, Ardhendu, Zachary Wharton, Yonghuai Liu, Morteza Ghahremani, Swagat Kumar, and Nik Bessis. 2023. "Regional Attention Network (RAN) for Head Pose and Fine-Grained Gesture Recognition." Ieee Transactions on Affective Computing..https://doi.org/10.1109/taffc.2020.3031841.
[5] Castillo Torres, Guillermo del, Maria Francesca Roig-Maimó, Miquel Mascaró-Oliver, Esperança Amengual, and Ramon Mas-Sansó. 2022. "Understanding How CNNs Recognize Facial Expressions: A Case Study With LIME and CEM." Sensors. https://doi.org/10.3390/s23010131.
[6] Chen, J., Corneanu, C. A., Barbu, T., & Escalera, S. 2020. "Deep Learning in Video Multi-Object Tracking: A Survey." Computer Vision and Image Understanding..https://doi.org/10.1016/j.cviu.2020.102950.
[7] Devi, S. L., & Sheela, R. R. 2023. "Real-Time Facial Emotion Recognition in Smart Classrooms Using CNN and Haar Cascade." Procedia Computer Science. https://doi.org/10.1016/j.procs.2023.01.142.
[8] Dhall, A., Goecke, R., Lucey, S., & Gedeon, T. 2011. "Static Facial Expression Analysis in Tough Conditions: Data, Evaluation Protocol and Benchmark." IEEE ICCV Workshops. https://doi.org/10.1109/ICCVW.2011.6130321.
[9] Ekman, Paul, dan Friesen, Wallace. 2009. Buka Dulu Topengmu (Panduan Membaca Emosi dan Ekspresi Wajah). Baca Buku, Buku Baik. Yogyakarta.
[10] Fabiano, R., & Paiva, A. 2020. "Understanding the Role of Emotions in Educational Chatbots." International Journal of Human–Computer Interaction.https://doi.org/10.1080/10447318.2020.1811268.
[11] Fauziannor, Fauziannor. 2022. "Faktor-faktor yang mempengaruhi minat mahasiswa dalam berorganisasi di kampus STIE Pancasetia." Fair Value: Jurnal Ilmiah Akuntansi dan Keuangan 4 (8): 3520–33. https://doi.org/10.32670/fairvalue.v4i8.14
[12] Ghosh, S., et al. 2023. "Real-Time Engagement Detection in Online Learning: Combining Facial and Behavioral Cues Using AI." Computers & Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2023.100105.
[13] Golec, M., & Grzyb, B. J. 2022. "Emotion Recognition in Education Using Machine Learning Techniques: A Review." Education and Information Technologies. https://doi.org/10.1007/s10639-022-10841-5.
[14] Huang, Ziyu, Chia-Chin Chiang, Jianhao Chen, Yi-Chian Chen, Hsin-Lung Chung, Yuping Cai, and Hsiu-Chuan Hsu. 2023. "A Study on Computer Vision for Facial Emotion Recognition." Scientific Reportshttps://doi.org/10.1038/s41598-023-35446-4.
[15] Iskandar, A., & Hasanah, N. 2021. "Pengaruh Minat dan Motivasi Terhadap Prestasi Belajar Siswa." Jurnal Ilmu Pendidikan. https://doi.org/10.23969/jip.v8i2.4551.
[16] Khan, M. A., Sharif, M., Raza, M., & Iqbal, W. 2021. "Facial Expression Recognition Using Deep Features with Transfer Learning." Computers, Materials & Continua.https://doi.org/10.32604/cmc.2021.014265.
[17] Kohnke, L. (2020). System Testing and Quality Assurance: A Practical Approach.
[18] Kurniawan, Henri, Kusrini Kusrini, and Kusnawi Kusnawi. 2023. "Klasifikasi Pengenalan Wajah Siswa Pada Sistem Kehadiran dengan Menggunakan Metode Convolutional Neural Network." Jurnal Media Informatika Budidarma 7 (2): 846.

https://doi.org/10.30865/mib.v7i2.5958.

[19] Lawpanom, Rit, Wararat Songpan, and Jakkrit Kaewyotha. 2024. “Advancing Facial Expression Recognition in Online Learning Education Using a Homogeneous Ensemble Convolutional Neural Network Approach.” Applied Sciences. https://doi.org/10.3390/app14031156.

[20] Li, Shizhong, et al. 2022. "Emotion Recognition in E-learning Environments Using Facial Expression and Eye Tracking." Interactive Learning Environments. https://doi.org/10.1080/10494820.2020.1821719.

[21] Li, X., & Deng, W. 2020. "Deep Facial Expression Recognition: A Survey." IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2020.3027591.

[22] Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. 2017. "Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order." Pattern Recognition. https://doi.org/10.1016/j.patcog.2016.10.004.

[23] Martinez, B., Valstar, M., Binefa, X., & Pantic, M. 2017. "Automatic Analysis of Facial Actions: A Survey." IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2016.2608725.

[24] Nandhini, M., & Bhavani, R. 2023. "Student Engagement Analysis Based on Facial Emotion Recognition Using CNN and RNN Techniques." International Journal of Cognitive Computing in Engineering.https://doi.org/10.1016/j.ijcce.2023.100161.

[25] Parmar, M., & Morris, B. T. 2016. "Driver Gaze Region Estimation without Using Eye Movement." IEEE Intelligent Vehicles Symposium. https://doi.org/10.1109/IVS.2016.7535411.

[26] Rafiq, A., Khan, M. J., & Aziz, A. 2021. "Deep Learning for Real-Time Emotion Recognition from Facial Expressions: A Survey." Artificial Intelligence Review. https://doi.org/10.1007/s10462-021-10006-y.

[27] Sari, R. M., & Oktaviani, M. 2022. "Pemanfaatan Artificial Intelligence dalam Dunia Pendidikan di Era Revolusi Industri 4.0." Jurnal Teknologi Pendidikan. https://doi.org/10.24114/jtp.v9i1.31988.

[28] Setiawan, R. 2022. “Penerapan Teknologi AI dalam Sistem Evaluasi Pembelajaran.” Jurnal Teknologi Informasi dan Pendidikan, 15(1), 65–74. https://doi.org/10.24036/tip.v15i1.1456.

[29] Shahzad, H. M., Sohail Masood Bhatti, Arfan Jaffar, Sheeraz Akram, Mousa Alhajlah, and Awais Mahmood. 2023. “Hybrid Facial Emotion Recognition Using CNN-Based Features.” Applied Sciences. https://doi.org/10.3390/app13095572.

[30] Singh, Rajesh, Sumeet Saurav, Tarun Kumar, Ravi Saini, Anil Vohra, and Sanjay Kumar Singh. 2023. “Facial Expression Recognition in Videos Using Hybrid CNN &Amp; ConvLSTM.” International Journal of Information Technology. https://doi.org/10.1007/s41870-023-01183-0.

[31] Su, Y., et al. 2022. "FERNet: Facial Expression Recognition Network Using Attention and Residual Learning." Multimedia Tools and Applications. https://doi.org/10.1007/s11042-022-12345-9.

[32] Tang, Chaolin, Dong Zhang, and Qichuan Tian. 2023. “Convolutional Neural Network–Bidirectional Gated Recurrent Unit Facial Expression Recognition Method Fused With Attention Mechanism.” Applied Sciences. https://doi.org/10.3390/app132212418.

[33] Tian, Yuan, Zhao Wang, D. Y. Chen, and Yao Huang. 2024. “TriCAFFNet: A Tri-Cross-Attention Transformer With a Multi-Feature Fusion Network for Facial Expression Recognition.” Sensors. https://doi.org/10.3390/s24165391.

[34] Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. 2017. "End-to-End Multimodal Emotion Recognition Using Deep Neural Networks." IEEE Journal

of Selected Topics in Signal Processing.https://doi.org/10.1109/JSTSP.2017.2764438.
[35] Wahyuni, S., & Prasetyo, A. R. 2022. “Hubungan antara Minat Belajar dan Keaktifan Siswa dalam Pembelajaran.” Jurnal Pendidikan dan Konseling. https://doi.org/10.31004/jpdk.v4i2.3353.
[36] Wibowo, D. A., Lestari, I., & Rahman, A. 2023. “Pengaruh Minat Belajar terhadap Hasil Belajar Siswa dalam Pembelajaran Daring.” Jurnal Pendidikan Indonesia, 12(2). https://doi.org/10.23887/jpi-undiksha.v12i2.50739.
[37] Xiong, Lingxin, Jicun Zhang, Xiaojia Zheng, and Yuxin Wang. 2024. “Context Transformer and Adaptive Method With Visual Transformer for Robust Facial Expression Recognition.” Applied Sciences. https://doi.org/10.3390/app14041535.
[38] Yadegaridehkordi, E., Nilashi, M., & Samad, S. 2021. "Emotion-Aware Educational Systems: A Systematic Literature Review." Computers & Education. https://doi.org/10.1016/j.compedu.2020.104062.
[39] Zaidan, B. B., Zaidan, A. A., & Albahri, A. S. 2020. "Comprehensive Review of Artificial Intelligence Techniques in Emotion Detection from Facial Expressions." Artificial Intelligence Review. https://doi.org/10.1007/s10462-019-09716-5..
[40] Zeng, Jiabei, Shiguang Shan, and Xilin Chen. 2020. "Facial Expression Recognition with Incomplete Labels." IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2019.2950462.
[41] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. 2016. "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks." IEEE Signal Processing Letters. https://doi.org/10.1109/LSP.2016.2603342.
[42] Zhong, Jiarui, Tangxian Chen, and Liuhan Yi. 2023. “Face Expression Recognition Based on NGO-BILSTM Model.” Frontiers in Neurorobotics. https://doi.org/10.3389/fnbot.2023.1155038.