JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 183 – 190 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i1.4354
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

MULTI-FACE EMOTION DETECTION USING CONVOLUTIONAL NEURAL NETWORKS TINY FACE DETECTOR

Jason Istioso[1*], Jeremiah Gerard [1], Marco Marcheleno [1], Muhammad Akbar Maulana[1]
[1]Informatics Engineering, Kwik Kian Gie School Of Business and Informatics
email: jasonistioso@gmail.com

Abstract: Understanding students' emotional conditions is important for evaluating engagement and learning atmosphere in classroom environments. However, conventional evaluation methods are often subjective and difficult to apply in real time. Therefore, this study proposes a real-time multi-face emotion detection system designed for classroom learning environments. The system integrates a CNN-based Tiny Face Detector for multi-scale face localization with a convolutional neural network to classify seven facial emotions: angry, disgust, fear, happy, sad, surprise, and neutral. Experimental evaluation was conducted using classroom video data under varying lighting conditions, face orientations, partial occlusions, and different numbers of detected faces per frame. The proposed system achieves stable real-time performance with processing speeds ranging from 10–20 FPS, depending on face density. The results show higher recognition performance for expressive emotions, while subtle emotions remain more challenging. Overall classification accuracy reaches above 80% when emotion predictions are aggregated across multiple faces and time windows. These results indicate that the proposed system is suitable for objective analysis of emotional dynamics in classroom environments and supports the deployment of lightweight emotion-aware monitoring systems for educational applications.

Keywords: classroom monitoring; convolutional neural network; facial emotion recognition; multi-face detection; tiny face detector.

Abstrak: Pemahaman terhadap kondisi emosional mahasiswa penting untuk mengevaluasi keterlibatan dan suasana pembelajaran di kelas. Namun, metode evaluasi konvensional umumnya bersifat subjektif dan sulit diterapkan secara real-time. Oleh karena itu, penelitian ini mengusulkan sistem deteksi emosi multi-wajah secara real-time yang dirancang untuk lingkungan pembelajaran di kelas. Sistem mengintegrasikan Tiny Face Detector berbasis CNN untuk pelokalan wajah multi-skala dengan jaringan saraf konvolusional untuk mengklasifikasikan tujuh emosi wajah, yaitu marah, jijik, takut, senang, sedih, terkejut, dan netral. Evaluasi eksperimen dilakukan menggunakan data video kelas dengan variasi kondisi pencahayaan, orientasi wajah, oklusi parsial, serta jumlah wajah yang berbeda dalam satu frame. Sistem menunjukkan kinerja real-time yang stabil dengan kecepatan pemrosesan antara 10–20 FPS, bergantung pada kepadatan wajah. Hasil pengujian menunjukkan kinerja yang lebih baik pada emosi ekspresif, sementara emosi dengan ciri halus lebih menantang untuk dikenali. Akurasi klasifikasi keseluruhan mencapai di atas 80% ketika hasil emosi diagregasi berdasarkan banyak wajah dan interval waktu. Hasil ini menunjukkan bahwa sistem yang diusulkan berpotensi digunakan untuk analisis objektif dinamika emosi di kelas serta mendukung pemantauan lingkungan pembelajaran berbasis kecerdasan buatan.

Kata kunci: pengenalan emosi wajah; deteksi multi-wajah; Tiny Face Detector; jaringan saraf konvolusional; pemantauan kelas.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 183 – 190 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i1.4354
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Emotional conditions play an important role in influencing human behavior, interaction, and performance, particularly in learning environments. In a campus setting, students' emotions can reflect engagement, motivation, stress, and overall learning atmosphere. Conventional methods for evaluating learning conditions, such as questionnaires and direct observation, are often subjective, time-consuming, and limited in real-time analysis. Therefore, an automatic and objective approach to emotion detection is required to support more accurate evaluation of learning environments [1], [2].

Recent advances in computer vision and deep learning have enabled the development of facial emotion recognition systems using Convolutional Neural Networks (CNN)[3], [4]. CNN-based methods are capable of extracting high-level facial features and have demonstrated strong performance in emotion classification tasks [5], [6]. However, many existing approaches focus only on single-face detection and are less effective in crowded environments such as classrooms, where multiple faces appear simultaneously with varying scales and lighting conditions [7].

To address this limitation, multi-face detection techniques have gained increasing attention. Tiny Face Detector is a CNN-based face detection algorithm designed to detect faces of various sizes, including small and distant faces [8]. This method is particularly suitable for real-world scenarios where faces may not always appear clearly or at close range [9]. By integrating Tiny Face Detector with a CNN-based emotion classification model, a robust multi-face emotion detection system can be developed.

Several previous studies have explored facial emotion recognition using CNN architectures and face detection techniques. In the study reported in [10], a CNN-based approach was applied to classify facial emotions after detecting faces from input images, showing effective emotion recognition performance. Meanwhile, the research in [11] utilized CNN architectures combined with face detection methods to improve facial expression classification accuracy; however, the study was conducted in controlled environments and focused on single-face scenarios. Research has shown that CNN models outperform traditional machine learning methods in terms of accuracy and robustness. Other studies have highlighted the effectiveness of Tiny Face Detector in handling scale variations and dense face distributions. Nevertheless, limited research has combined Tiny Face Detector with CNN-based emotion recognition specifically for evaluating learning atmospheres in campus environment [12].

Although deep learning-based emotion recognition has been widely studied, most existing works focus on single-face detection in controlled settings and lack real-time multi-face analysis in real classroom environments. In addition, the integration of efficient face detection methods with convolutional neural networks for objective evaluation of learning atmospheres remains limited. To address this gap, this study proposes a real-time multi-face emotion detection system using CNN and the Tiny Face Detector for campus learning environments.

METHOD

This research employs an experimental research methodology to design, implement, and evaluate a multi-face

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 183 – 190 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i1.4354
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

emotion detection system for campus learning environments. The proposed method consists of four sequential stages: data acquisition, multi-face detection, emotion classification, and visualization. The workflow is designed to ensure reproducibility and objective performance evaluation.

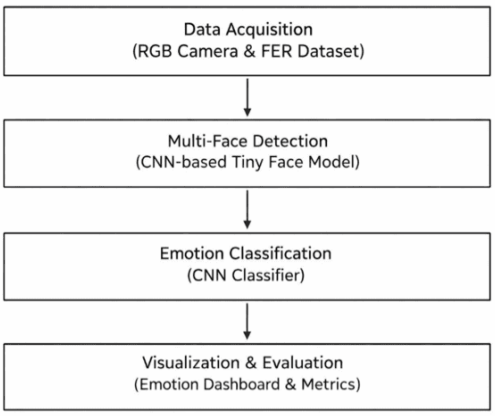

Image 1: Research Stages

Facial image and video data were collected from classroom environments using standard RGB cameras under varying lighting conditions. In addition to the FER-2013 dataset, a limited number (62) of facial images extracted from classroom video recordings were used to complement the dataset. The campus images were manually labeled into predefined emotion categories based on observable facial expressions, following the same emotion classes used in the FER-2013 dataset.

All recorded data were anonymized and used solely for research purposes to preserve participant privacy. The FER-2013 dataset was utilized due to its diversity and its widespread adoption in training and benchmarking CNN-based facial emotion recognition models.

Face localization is performed using a multi-scale CNN-based detection algorithm. The detector used in this study builds upon the Tiny Face Detector concept by aggregating multi-scale features and contextual attention, enabling robust detection of faces of varying sizes in crowded scenes. The face detection process can be mathematically expressed as shown in (1):

$$D(x) = \arg\max \{b \in B\}\ g\varphi(xb) \quad (1)$$

Description:

x_b denotes an image patch corresponding to bounding box b,

B is the set of candidate regions across image scales, and

$g_\phi(\cdot)$ represents the CNN detection function parameterized by ϕ.

Each detected face region is cropped and resized before being forwarded to the emotion recognition module. Emotion classification is conducted using a Convolutional Neural Network (CNN) designed to recognize seven emotional states: angry, disgust, fear, happy, sad, surprise, and neutral. Deep learning models with attention mechanisms have demonstrated improved robustness and accuracy in real-world facial expression tasks. The learned deep features F_k from a face image are classified via (2):

$$\hat{y} = \arg\max\_k\ h\theta(F\_k) \quad (2)$$

The proposed system produces an emotion distribution output that represents the proportion of detected emotional states within a given time window. After emotion classification, the system aggregates prediction results from multiple detected faces and computes the frequency of each emotion class. The distribution is normalized to obtain percentage values, enabling comparative analysis of emotional conditions in the learning environment as classified in (3):

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 183 – 190 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i1.4354
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

$$E\_k = (n\_k / \Sigma\_\{i=1\}^{\{c\}} n\_i) \times 100\% \quad (3)$$

Description:

n_k is denotes the number of detected faces classified into the k-th emotion category,

C is the total number of emotion classes E_k is indicates the percentage of emotion distribution for class k.

This formulation allows the system to quantify the proportion of each emotional state within a given observation period, thereby providing an objective representation of the overall emotional atmosphere in the learning environment.

RESULT AND DISCUSSION

Face Detection Performance

The proposed system was evaluated using real classroom video streams containing multiple students under varying conditions, including illumination changes, face orientation variations, partial occlusions, and different distances from the camera. The CNN-based Tiny Face Detector demonstrated robust performance in detecting multiple faces simultaneously within a single frame. The detector was able to localize both large and small faces effectively, including distant faces commonly observed in classroom environments.

Experimental observations show that when the number of detected faces increases, the processing load also increases, leading to a gradual reduction in frame rate (FPS). However, the system remains stable and capable of real-time operation, indicating that the Tiny Face Detector is suitable for crowded learning environments.

Emotion Recognition Analysis

Emotion classification was conducted for seven basic facial expressions following the standard label order defined in the FER-2013 dataset, namely Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. These emotion categories are widely adopted in facial expression recognition research due to their psychological relevance and clear facial action patterns.

The performance of each emotion class was analyzed based on classification confidence, detection consistency, and robustness under real-world classroom conditions. Variations in lighting, face orientation, partial occlusion, and the presence of multiple faces within a single frame influenced recognition accuracy across different emotion categories. The following subsections provide a detailed discussion of the system's behavior and recognition characteristics for each emotion, highlighting which emotions are more easily detected and which remain challenging in multi-face and real-time scenarios.

Angry Expression

Angry expressions are characterized by distinctive muscle activations in the eyebrow, eye, and mouth regions, making them moderately recognizable by CNN-based models. Lowered and drawn-together eyebrows, narrowed eyes, and pressed lips with jaw tension serve as key indicators of anger. The proposed CNN model effectively captures these localized features, resulting in reliable recognition when facial details are clear as shown in image 2. However, subtle eyebrow movements or lighting variations may cause confusion with neutral or sad expressions.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 183 – 190 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i1.4354
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

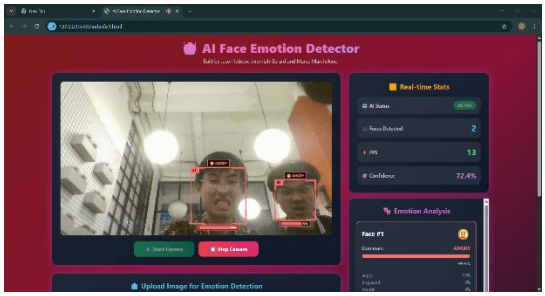

Image 2. Angry Emotion Detection

Disgust Expression

Disgust is characterized by localized muscle activations in the mid-face region, particularly nose wrinkling and upward movement of the upper lip, which form the primary cues for recognition. Eye-related features play a minor role, causing the CNN model to rely mainly on mid-face features. In classroom settings, disgust occurs less frequently and is sensitive to occlusion and lighting variations, which may lead to confusion with anger or surprise. Therefore, disgust is classified as a moderately difficult emotion to detect, as illustrated in Image 3.

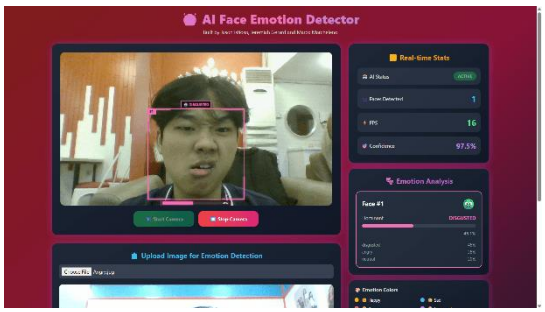

Image 3. Disgust Emotion Detetion

Fear Expression

Fear expressions involve distributed muscle activations across the eye, eyebrow, and mouth regions, making them more challenging to recognize than highly expressive emotions. The most prominent cue appears in the eye region, characterized by widened eyes and raised upper eyelids, while the eyebrow region shows raised brows with slight inward tension that may cause ambiguity with surprise. Mouth features contribute less consistently and often overlap with surprise-related patterns. Consequently, fear remains one of the more difficult emotions to detect due to its visual similarity to surprise and reliance on subtle facial differences, as illustrated in Image 4.

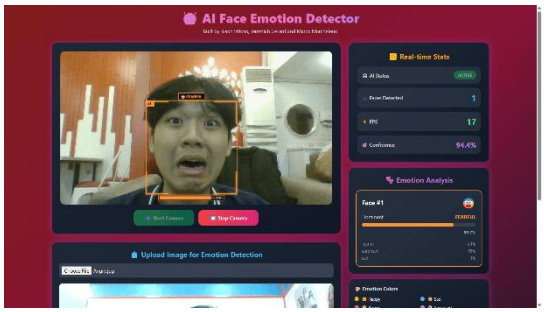

Image 4. Fearful Emotion Detection

Happy Expression

Happy expressions are the most easily recognized due to their strong and globally distributed facial features. The primary cue appears in the mouth region, where upward-curved lip corners forming a clear smile, often with visible teeth, are effectively captured by CNN feature maps. Additional cues from the eye and cheek regions further enhance recognition robustness. As a result, the proposed CNN model achieves high confidence and accuracy in classifying happy emotion, making it the easiest emotion to detect among the FER-2013 classes, as shown in Image 5.

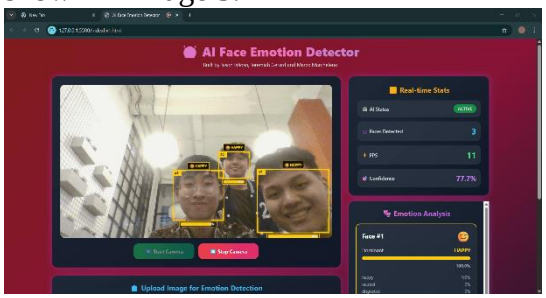

Image 5. Happy Emotion Detection

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 183 – 190 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i1.4354
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

Sad Expression

Sadness is characterized by reduced facial muscle tension and downward-oriented features, making it more subtle and difficult to detect than high-arousal emotions. The primary cues appear in the mouth region as a mild downward pull of the lip corners, while the eye region shows drooping upper eyelids and a relaxed gaze, producing low-contrast facial patterns. Additional eyebrow cues are weak and often inconsistent, causing the proposed CNN model to confuse sadness with neutral expressions. Consequently, sadness is classified as one of the more challenging emotions to recognize in classroom environments, as illustrated in Image 6.

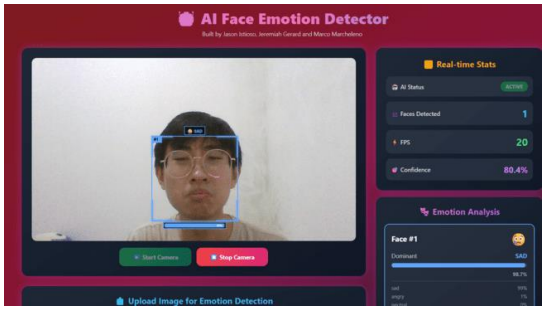

Image 6. Sad Emotion Detection

Surprise Expression

Surprise is characterized by rapid and pronounced facial activations across multiple regions, making it one of the easiest emotions to recognize. Key cues appear in the upper face through sharply raised eyebrows and widened eyes, producing strong visual contrasts captured by CNN filters. The mouth region further contributes through a vertically opened and rounded shape with a relaxed jaw. Due to the simultaneous activation of these features, the proposed CNN model achieves high confidence and accuracy in recognizing surprise expressions, including in multi-face classroom scenarios, as illustrated in Image 7.

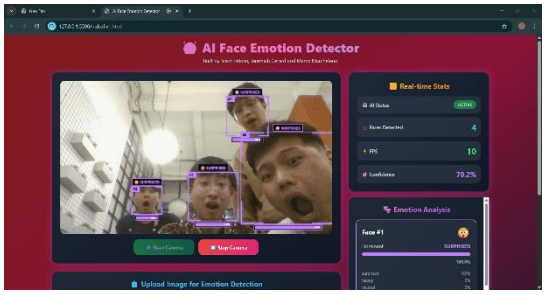

Image 7. Surprise Emotion Detection

Neutral Expression

Neutral expression represents the baseline emotional state and serves as a reference for distinguishing other facial emotions. It is characterized by minimal muscle activation, with relaxed eyebrows, normal eye appearance, and a closed or slightly open mouth without visible tension. In CNN-based emotion recognition, neutral functions as a baseline class from which deviations are classified as other emotions. Although generally easy to recognize under stable conditions, neutral expressions may be confused with subtle emotions such as mild sadness or low-intensity anger. As illustrated in Image 8, the highlighted regions show minimal facial variation, confirming the foundational role of neutral expression in emotion classification.

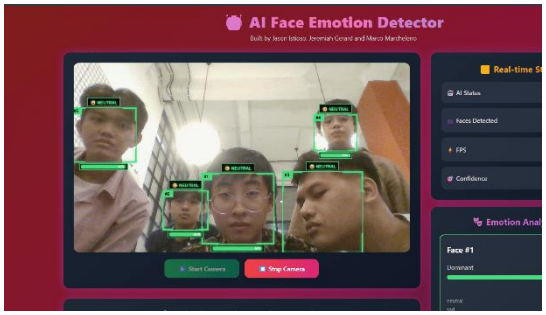

Image 8. Neutral Emotion Detection

Discuss

The experimental results indicate that the proposed multi-face emotion detection system achieves stable real-time performance under varying numbers of

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol XII No 1, Desember 2025, hlm. 183 – 190 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v12i1.4354
Available online at https://jurnal.stmikroyal.ac.id/index.php/jurteksi

detected faces. The system maintains an average FPS of 18–20 when processing one to two faces and remains above 10 FPS for four to five faces, which is sufficient for real-time classroom monitoring, as summarized in Table 1. Compared to previous studies, the proposed system provides additional capability for real-time multi-face analysis while maintaining comparable accuracy**.**

Table 1. Comparison of the Proposed System with Previous Studies

Study	Method	Accuracy (%)
Firmansyah et al. [10]	CNN	78.0
Haider et al. [11]	Triplet CNN + SVM	82.1
Proposed System	CNN + Tiny Face	>80

In terms of emotion recognition, expressive emotions such as happy and surprise achieve higher confidence and accuracy levels, while subtle emotions such as fear, disgust, and sad show lower performance due to less distinctive facial cues. Overall, the system achieves an average accuracy above 80% across seven emotion classes, demonstrating a balanced trade-off between processing speed and recognition accuracy for real-world classroom applications.

CONCLUSION

This research demonstrates the development of a multi-face emotion detection system based on CNN integrated with a Tiny Face Detector for campus learning environments. The proposed system enables real-time detection and emotion analysis of multiple faces under typical classroom conditions, providing an objective representation of students' emotional dynamics and supporting educational monitoring.

However, the system has several limitations. Its performance may decrease under very low lighting conditions, partial face occlusion, or when students wear face masks, which can obscure key facial features. In addition, subtle emotions remain more challenging to recognize accurately. Future work will focus on improving robustness under these conditions by incorporating temporal modeling, enhancing feature extraction for subtle expressions, and integrating multimodal information to further support intelligent emotion-aware learning environments**.**

BIBLIOGRAPHY

[1] R. Pereira *et al.*, "Systematic Review of Emotion Detection with Computer Vision and Deep Learning," *Sensors*, vol. 24, no. 11, p. 3484, May 2024, doi: 10.3390/s24113484.

[2] Z. Shou *et al.*, "A Student Facial Expression Recognition Model Based on Multi-Scale and Deep Fine-Grained Feature Attention Enhancement," *Sensors*, vol. 24, no. 20, p. 6748, Oct. 2024, doi: 10.3390/s24206748.

[3] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, p. 135, Feb. 2024, doi: 10.3390/info15030135.

[4] Y. Chen and M. Zhang, "Research on face emotion recognition

algorithm based on deep learning neural network," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns.2023.2.00533.

[5] F. Fatimatuzzahra, L. Lindawati, and S. Soim, "Development of Convolutional Neural Network Models to Improve Facial Expression Recognition Accuracy," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 10, no. 2, pp. 279–289, Jun. 2024, doi: 10.26555/jiteki.v10i2.28863.

[6] W. Wu, H. Peng, and S. Yu, "YuNet: A Tiny Millisecond-level Face Detector," *Mach. Intell. Res.*, vol. 20, no. 5, pp. 656–665, Oct. 2023, doi: 10.1007/s11633-023-1423-y.

[7] K. Anwar, "'Sistem Deteksi Wajah Berbasis Deep Learning Menggunakan Convolutional Neural Network (CNN),'" *J. Comput. Sci. Inf. Technol.*, vol. 1, no. 2, pp. 46–52, 2025, doi: 10.70716/jocsit.v1i2.258.

[8] R. Leyva, G. Shen, O. Bahadir, V. Sanchez, and T. Guha, "Boosting Tiny Face Detection in Videos with an Integral Score Framework," in *2025 IEEE 19th International Conference on Automatic Face and Gesture Recognition (FG)*, IEEE, May 2025, pp. 1–8. doi: 10.1109/FG61629.2025.11099181.

[9] A. Alshammari and M. E. Alshammari, "Emotional Facial Expression Detection using YOLOv8," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 5, pp. 16619–16623, Oct. 2024, doi: 10.48084/etasr.8433.

[10] I. Firmansyah, D. U. K. Putri, and B. A. A. Sumbodo, "Klasifikasi Ekspresi Wajah Menggunakan CNN Dalam Keadaan Wild Setting Pada Virtual Meeting," *IJEIS (Indonesian J. Electron. Instrum. Syst.*, vol. 14, no. 2, p. 139, Oct. 2024, doi: 10.22146/ijeis.92088.

[11] I. Haider, H.-J. Yang, G.-S. Lee, and S.-H. Kim, "Robust Human Face Emotion Classification Using Triplet-Loss-Based Deep CNN Features and SVM," *Sensors*, vol. 23, no. 10, p. 4770, May 2023, doi: 10.3390/s23104770.

[12] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," *Sensors*, vol. 21, no. 9, p. 3046, Apr. 2021, doi: 10.3390/s21093046.