Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10, 832-853
2025
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v9i10.10555

Development of a real time facial emotion recognition system based on deep learning and CCTV integration for smart surveillance applications

Wijitra Montri[1*], Sumran Chaikhamwang[2], Srinuan Fongmanee[3], Chalida Janthajirakowit[4]
[1,2,4]Information Technology, Faculty of Digital Technology, Chiangrai Rajabhat University, Thailand;
sc_wijittra@crru.ac.th (W.M.) em_sumran@crru.ac.th (S.C.) chalida.jan@crru.ac.th (C.J.).
[3]Computer Science, Faculty of Digital Technology, Chiangrai Rajabhat University, Thailand; sc_srinual@crru.ac.th (S.F.).

Abstract: The purpose of this research is to develop a real-time facial emotion recognition (FER) system that integrates deep learning techniques with closed-circuit television (CCTV) for intelligent surveillance applications. The system aims to overcome technical and environmental challenges often encountered in real-world CCTV environments, such as low image resolution, illumination instability, and varying facial orientations. The study adopts a systematic methodology that includes a literature review, research framework design, dataset preparation, model training using convolutional neural networks (CNNs), and web-based system implementation for real-time monitoring and alerting. The proposed model, developed on the TensorFlow platform and fine-tuned with FER2013, RAF-DB, and AffectNet datasets, achieved an overall accuracy of 80.24%, with precision, recall, and F1-score values of 80.59%, 79.54%, and 80.05%, respectively. The web application allows seamless integration with CCTV feeds, enabling real-time emotion detection, alert notifications via screen and email, and historical data analysis for behavioral trend evaluation. The findings indicate that the proposed FER system can be effectively incorporated into existing surveillance infrastructures to enhance situational awareness and proactive decision-making. This contributes to improving public safety in sensitive environments such as schools, public facilities, and government institutions. The practical implications suggest that the system provides a scalable framework for emotion-aware surveillance, which can be extended to multimodal emotion recognition and edge computing to improve responsiveness, privacy, and scalability in future research.

Keywords: *Deep learning, CCTV integration, Facial emotion recognition, Intelligent surveillance, Real-time processing.*

1. Introduction

In recent years, the rapid advancement of Artificial Intelligence (AI) and Deep Learning has led to significant progress in the domains of image processing and pattern recognition. These developments have substantially improved the accuracy and efficiency of FER systems [1-4]. FER facilitates the automated analysis of human emotional states through visual cues, which has enabled its deployment in a variety of practical applications, including customer behavior analysis in smart retail environments, public safety monitoring, and preliminary emotion assessment in healthcare and educational settings.

In response to these emerging demands, the present study focuses on the development of a real-time FER system utilizing video streams from CCTV cameras. The system employs deep learning models implemented via the TensorFlow framework to detect faces in live video feeds, process facial features, and recognize emotional states instantaneously. This approach represents a substantial improvement over traditional FER systems, which generally rely on static images or pre-recorded high-quality videos for emotion classification [5, 6].

A notable advantage of the proposed system is its ability to operate effectively under the practical

History: Received: 23 July 2025; Revised: 19 September 2025; Accepted: 23 September 2025; Published: 14 October 2025
* Correspondence: sc_wijittra@crru.ac.th

constraints of CCTV environments, which are often characterized by low-resolution footage, unpredictable camera angles, and variable lighting conditions. To mitigate these challenges, the algorithm was designed with adaptive mechanisms that ensure robust emotion recognition across diverse and dynamic scenarios. Examples include the integration of super-resolution-based FER frameworks [7], emotion-aware super-resolution networks that preserve expression fidelity [8] and adaptive recognition pipelines that optimize performance according to input resolution [9].

Furthermore, the system was implemented as a web-based platform, enabling users to access its functionalities through standard web browsers without the need for additional software installation [10]. This design supports convenient, real-time emotion monitoring through an intuitive user interface, while simultaneously facilitating long-term data storage and trend analysis [11]. Such capabilities provide actionable insights that can inform strategic decision-making in both operational and service development contexts [12, 13].

2. Literature Review

2.1. Significance of Facial Emotion Recognition

Facial expressions constitute one of the most natural and nonverbal forms of human communication, effectively conveying emotional states without the use of spoken language. Consequently, FER has garnered increasing attention, particularly in light of advancements in deep learning techniques that have significantly enhanced the accuracy of facial analysis in real-world environments [14-16].

2.2. Technologies and Deep Learning Models in FER

In recent years, CNNs, such as ResNet and EfficientNet, have demonstrated strong performance in emotion classification tasks using static images and video frames [17]. However, hardware limitations in CCTV systems have prompted the development of lightweight models, such as MobileNet, as well as the adoption of model compression techniques like pruning and quantization to reduce memory usage and improve processing speed [18, 19]. More recently, Vision Transformers (ViTs) have gained popularity due to their superior ability to capture the spatial context of facial features, particularly under challenging conditions such as low-resolution images or poor lighting environments [20-23].

2.3. Datasets and Limitations

Commonly used benchmark datasets for training and evaluating FER models include FER2013, AffectNet, and RAF-DB [15, 16]. Despite their widespread adoption, these datasets present several limitations, such as class imbalance in emotional categories and insufficient cultural diversity. These issues may result in biased models that perform poorly when deployed in heterogeneous real-world scenarios [24, 25].

2.4. Application of FER in Real-Time CCTV Systems

Integrating FER into CCTV systems holds considerable potential for behavioral monitoring, public safety, and emotion analysis in public spaces. Nonetheless, real-time video processing from CCTV feeds poses several challenges, including low resolution, variable camera angles, and inconsistent lighting conditions [26]. Research by Alshammari and Alshammari [27] has demonstrated that YOLOv8 is capable of accurately detecting faces and classifying emotions in real-time video streams, especially when combined with edge computing techniques that minimize data transmission latency [22, 28, 29].

Furthermore, the integration of Multi-Person FER and multimodal emotion recognition, incorporating data from additional sensors such as audio or gestures, is emerging as a promising approach to improving the accuracy of emotion assessments in real-world contexts [30]. Despite the substantial progress in FER technologies, challenges remain, such as achieving robust performance under constrained conditions, enhancing model inference speed, and mitigating dataset-induced biases [15]. Moreover, there is a growing need for explainable AI (XAI) models that provide interpretable outputs, particularly for applications in security and mental health domains where transparency is critical [25, 31, 32].

3. Methodology

This study was conducted through a structured sequence of six methodological steps, as outlined below:

3.1. Literature Review and Related Work Analysis

A comprehensive review of relevant literature, academic documents, and prior research was undertaken, focusing on FER technologies. Special emphasis was placed on studies utilizing artificial intelligence, particularly those involving deep learning techniques and the TensorFlow framework. The review encompassed recent trends, technological advantages and limitations, and applications of FER in conjunction with CCTV systems for real-time emotion processing [15, 33-35].

3.2. Definition of Research Framework and Objectives

This phase involved defining the research problem, objectives, scope, hypotheses, and variables associated with the development of a FER model capable of processing real-time CCTV footage. A technical conceptual framework was designed, incorporating CNN architectures for image processing and a web application interface for result visualization [36].

3.3. Prototype System Design and Planning

A prototype system was designed to perform facial emotion detection using CCTV images. The design considered both hardware components (e.g., CCTV IP cameras) and software infrastructure (e.g., TensorFlow and web technologies). The system architecture was structured to handle image acquisition, face detection, emotion analysis, and real-time emotion display [31, 37].

3.4. Dataset Selection and Preparation

Publicly available and well-established FER datasets, including FER 2013, RAF DB, and AffectNet, were selected due to their clearly labeled emotional categories. The data were split into training, testing, and validation sets. Data augmentation techniques, such as image rotation, brightness adjustment, and horizontal flipping, were applied using TensorFlow libraries to reduce overfitting and improve the model's generalization capability [16, 26].

3.5. Model Development and Training

The TensorFlow platform was employed to develop and train a deep learning model for emotion classification. Key tasks in this phase included:

- Designing a CNN-based architecture;
- Training the model with the prepared dataset;
- Performing hyperparameter tuning to optimize learning performance;

Evaluating the model using standard metrics, including accuracy, precision, recall, and F1 score [14, 24].

```
# 7. Train Model
history = model.fit(
train_generator,
epochs=150,
validation_data=validation_generator,
Callbacks=[early_stopping, reduce_lr]
)
Epoch 1/150
234/234[ ==================] - 30s 99ms/step - 1.5475 - accuracy : 0.4343 - val_lossuracy 0.2701
Epoch 2/150
234/234[ ==================] - 28s 94ms/step - 1.3109 - accuracy : 0.5296 - val_lossuracy 0.9222
Epoch 3/150
234/234[ ==================] - 28s 10ms/step - 1.1972 - accuracy : 0.5762 - val_lossuracy 0.6280
Epoch 4/150
234/234[ ==================] - 30s 122m/step - 1.0371 - accuracy : 0.6225 - val_lossuracy 0.6409
Epoch 5/150
234/234[ ==================] - 30s 122m/step - 1.0111 - accuracy : 0.6318 - val_lossuracy 0.6181
Epoch 6/150
234/234[ ==================] - 30s 112m/step - 1.0111 - accuracy : 0.6318 - val_lossuracy 0.6574
Epoch 7/150
234/234[ ==================] - 29s 116m/step - 0.9118 - accuracy : 0.6574 - val_lossuracy 0.6253
Epoch 8/150
234/234[ ==================] - 32s 119m/step - 0.9189 - accuracy : 0.6874 - val_lossuracy 0.6922
Epoch 9/150
234/234[ ==================] - 32s 119m/step - 0.9165 - accuracy : 0.8490 - val_lossuracy 0.6788
```

Figure 1.
Model Training.

3.6. Web Application Development and CCTV Integration

A web-based application was developed to interface with CCTV cameras and process incoming video streams in real time using the trained TensorFlow model. The application was designed to display emotional states detected from facial expressions and to log the detection results in a database for further analysis. API integration was implemented to ensure seamless and flexible communication between the frontend and backend components of the system.

3.6.1. System Workflow

The operational workflow of the real-time facial emotion recognition system utilizing CCTV cameras can be described sequentially (see Figure 2).

3.6.2. Steps as Follows

3.6.2.1. Real-Time Facial Image Capture by CCTV Camera

CCTV cameras are installed in designated surveillance areas to continuously capture facial images of users in real time. The video feed is transmitted to the frontend system for further processing.

3.6.2.2. Frontend Display and Data Transmission to Backend

The captured images are simultaneously displayed on the frontend interface (user screen) and forwarded to the backend for emotion analysis.

3.6.2.3. Emotion Analysis by Backend Using Deep Learning Models

The backend system employs deep learning techniques to analyze facial expressions and classify them into various emotional states such as sadness, anger, happiness, etc.

3.6.2.4. Transmission of Analysis Results to Frontend

The backend returns the analytical results to the frontend, where the detected emotional trends are visualized in formats such as graphs or emotion percentage distributions on the user interface.

3.6.2.5. Abnormal Emotion Detection and User Notification

If the system detects abnormal emotional states such as intense stress or anger, it triggers an alert mechanism. Notifications are sent to users via messaging platforms or email.

3.6.2.6. Audio Alert Transmission via Speaker

In cases where critical emotional states are identified, the backend system activates an audio alert through a connected speaker to notify the affected individual immediately.

3.6.2.7. User Awareness and Situational Insight

The user receives both the emotional trend data and alerts from the system, enabling effective monitoring and situational analysis for informed decision-making, as shown in Figure 2.

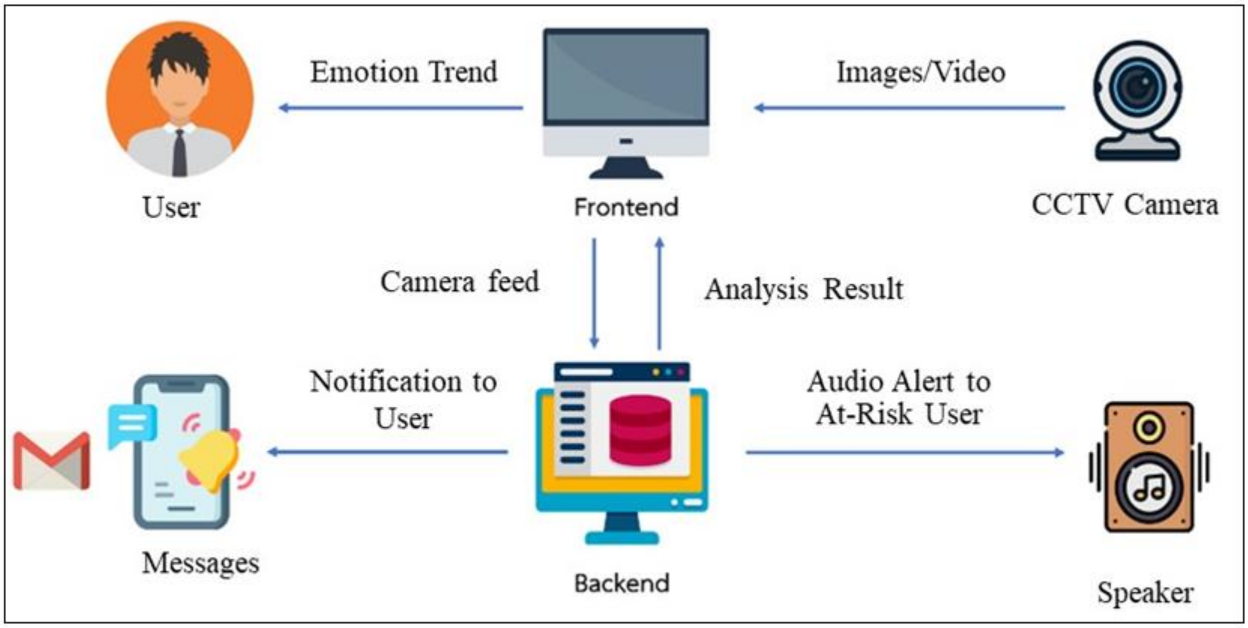

Figure 2.
System Workflow.

3.6.3. System Operation Workflow

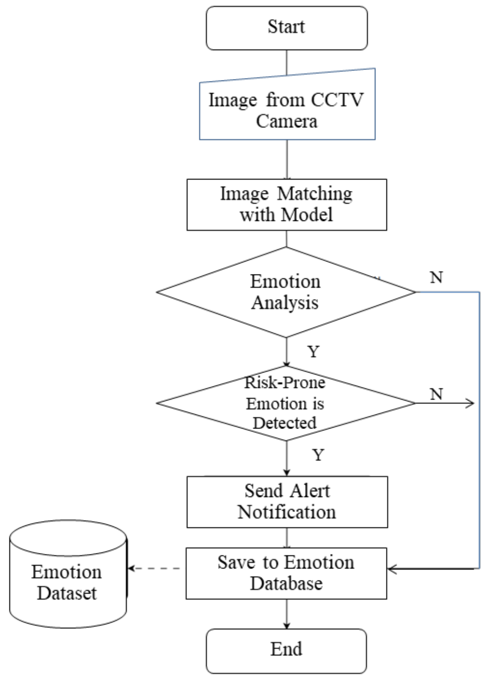

Figure 3.
System Operation Workflow.

The real-time facial emotion recognition system was designed with an emphasis on modularity, responsiveness, and adaptability to real-world CCTV environments. The workflow of the system, as illustrated in Figure X, consists of several key stages that operate in a continuous loop to enable real-time monitoring, detection, and response. The process is described as follows:

3.6.3.1. System Initialization

The system initializes all required components, including camera interfaces, neural network models, and database connections, preparing for live image acquisition.

3.6.3.2. Image Acquisition from CCTV Camera

CCTV cameras serve as the primary input devices, capturing real-time video or still images of individuals within the monitored area. These visual inputs are continuously streamed to the backend processing system.

3.6.3.3. Image Matching with Pre-Trained Model

The incoming image is matched against a pre-trained deep learning model for facial recognition and feature extraction. CNNs, particularly lightweight variants such as MobileNet or EfficientNet, are used to optimize computational efficiency under constrained hardware.

3.6.3.4. Emotion Analysis

The system then performs emotion classification using facial feature vectors. The model, trained on emotion-labeled datasets such as FER 2013 and AffectNet, predicts emotional states such as happiness, anger, sadness, or fear.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 832-853, 2025
DOI: 10.55214/2576-8484.v9i10.10555

3.6.3.5. Detection of Risk-Prone Emotion

The emotion output is analyzed to determine if it represents a risk-prone state (e.g., severe anger or stress). These criteria are defined based on predefined thresholds, which can be adapted through domain-specific configuration.

3.6.3.6. Alert Notification

When a critical emotion is detected, the system issues a real-time alert via connected communication channels such as mobile notifications or control room dashboards. This alerting function is essential in applications such as security monitoring and healthcare settings.

3.6.3.7. Data Logging and Storage

The identified emotion, along with metadata including timestamp and camera location, is logged and stored in a centralized emotion database. This enables retrospective analysis, trend identification, and supports explainability in decision-making.

3.6.4. Loop Continuation

The system returns to the monitoring loop, awaiting the next image input for continuous real-time operation.

4. Results and Discussion

The results are organized into two main parts: (1) the user part and (2) the administrative part. The details are as follows:

4.1. User Part

4.1.1. Web Application Overviews

The user part comprises 11 sections, along with a main navigation menu that provides access to the system's core functionalities:

Home Page: Displays a welcome message, introduces the system's key features, and provides example usage scenarios. It also includes options for user login and registration, as shown in Figure 4.

ISSN: 2576-8484
Vol. 9, No. 10: 832-853, 2025
DOI: 10.55214/2576-8484.v9i10.10555

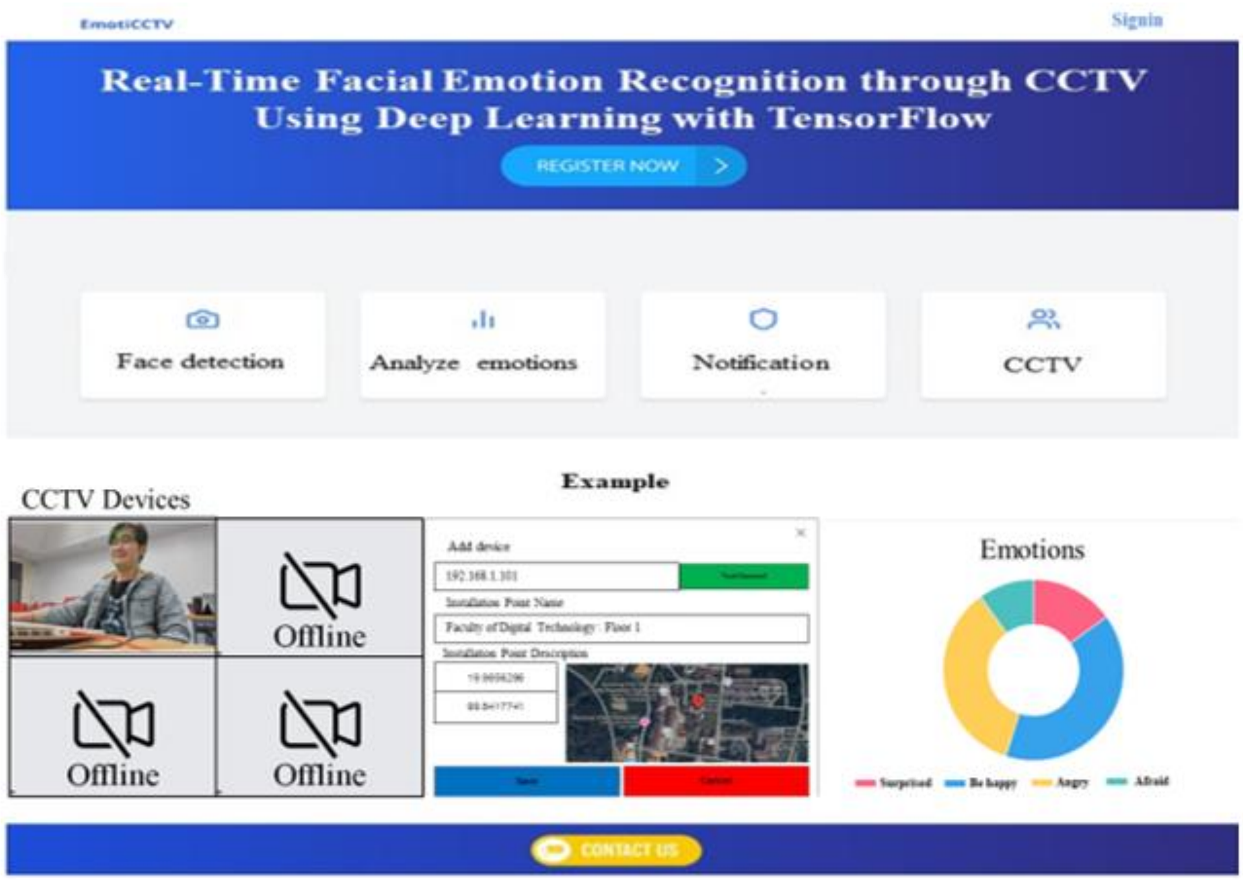

Figure 4.
Main page.

Sign-up Page: Contains a registration form where users can input personal information, including email address, phone number, and password, as shown in Figure 5.

DOI: 10.55214/2576-8484.v9i10.10555

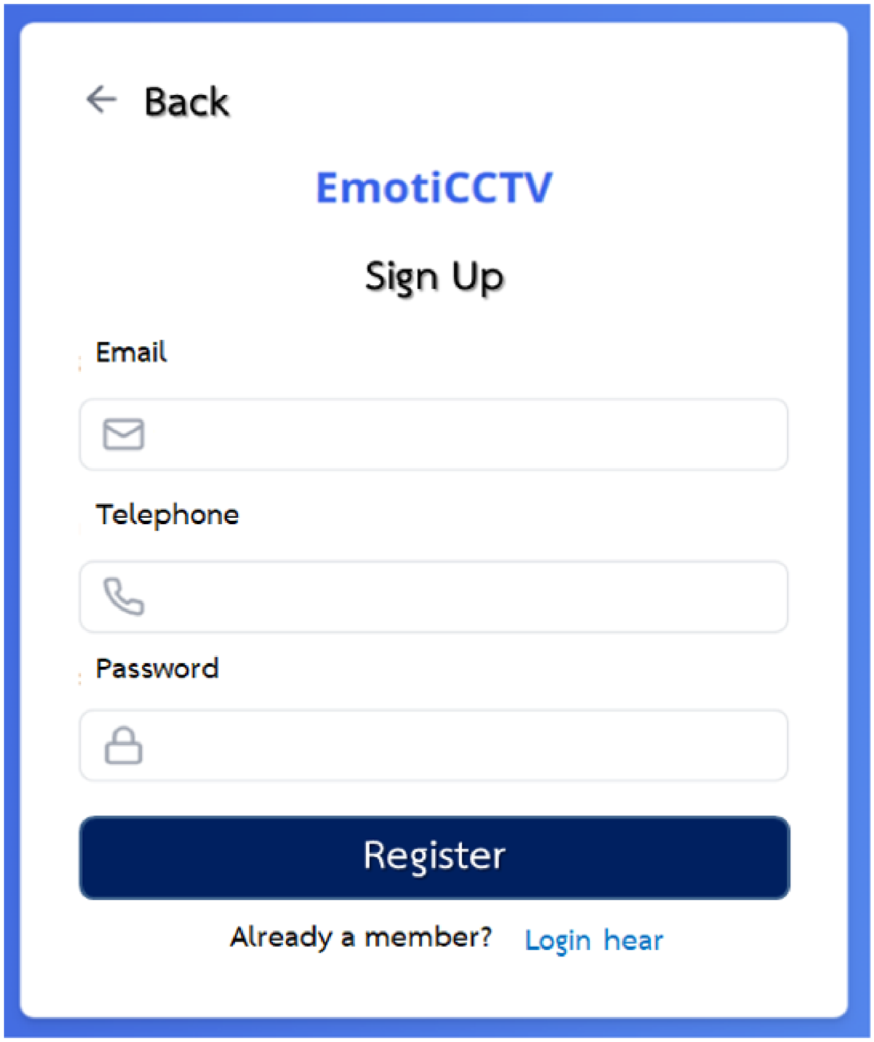

Figure 5.
Sign up page.

Login and Password Recovery Pages: Allow users to log in to the system and reset their passwords via email.

Main Menu: Enables users to navigate to various functional areas, including CCTV monitoring, time scheduling, emotion reports, issue reporting, and user settings, as shown in Figure 6.

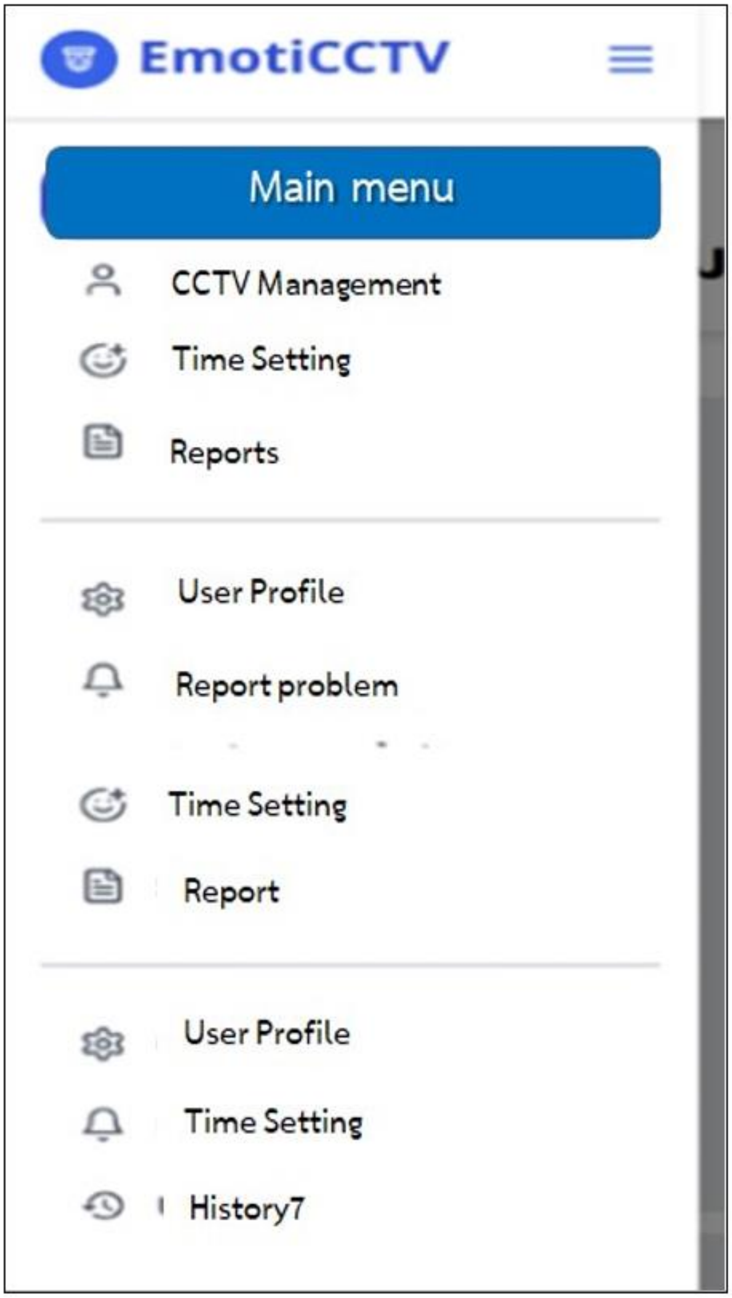

Figure 6.
Main menu.

CCTV Management Page: Supports adding and managing surveillance cameras by specifying the IP address, camera location name, and geographic position, as shown in Figure 7.

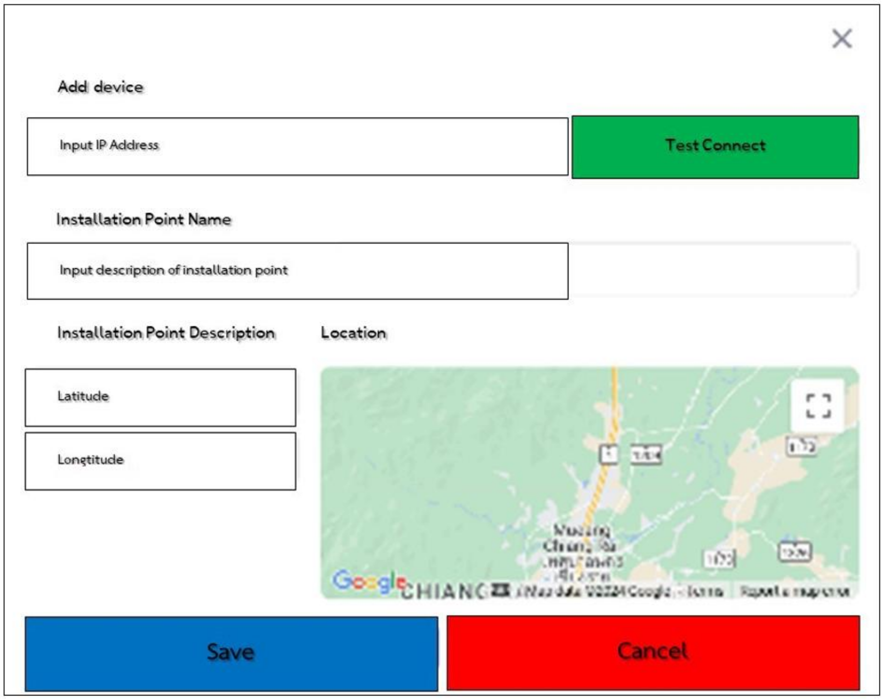

Figure 7.
CCTV Management.

Scheduling Page: Allows users to configure alert timings based on emotional trends and to select custom alert sounds.

Emotion Report Page: Presents emotion analysis results filtered by time range or specific cameras, and includes functionality to export reports in PDF format.

User Settings and Issue Reporting Pages: Enable users to update personal information, report issues with optional image attachments, and track the status of reported issues.

Notification System: Sends email alerts in cases of abnormal emotional detection or potential risk scenarios.

4.1.2. Emotion Detection Results

Figure 8.
Emotion Detection.

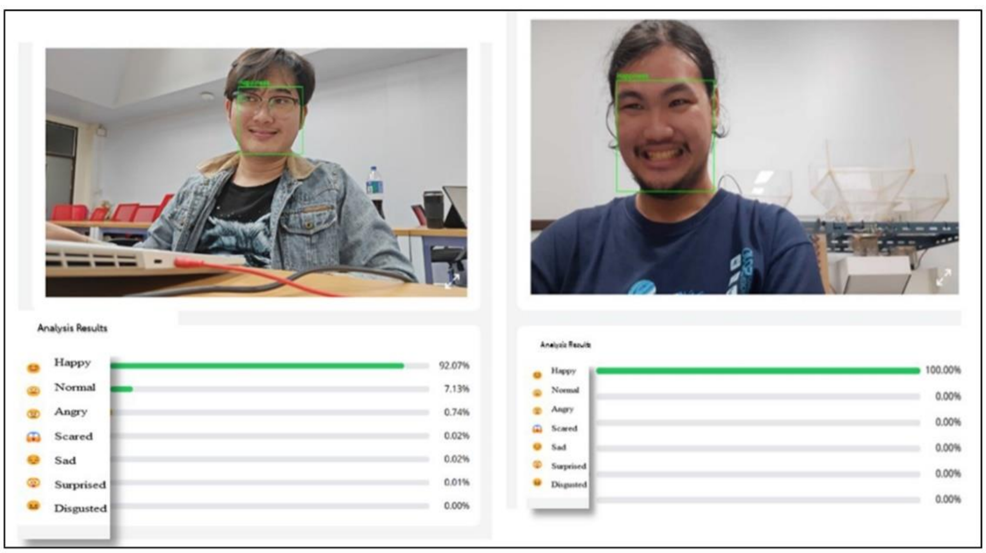

Figure 9.
Emotion Detection Results.

From Figure 8-9, the experimental results from deploying the emotion detection system using four CCTV cameras demonstrate the system's ability to operate in real time and accurately detect human emotions from facial expressions. The system captures video streams from each camera and processes them through a pre-trained artificial intelligence model. The detected emotional states are then visualized as percentage values on a web-based application interface. The results are categorized into seven primary emotions: (1) happiness, (2) neutral, (3) anger, (4) fear, (5) sadness, (6) surprise, and (7) disgust. The emotional percentages are tracked over time to reveal trends, enabling continuous monitoring and assessment of emotional states within the surveillance area.

In cases where high intensity or abnormal emotions are detected, such as elevated levels of anger or fear, the system is programmed to trigger automatic alerts based on predefined thresholds. This functionality highlights the system's practical potential for application in real-world scenarios, particularly in areas related to public safety, spatial management, and behavioral monitoring. The ability to analyze emotional trends from multiple surveillance points enhances situational awareness and supports proactive decision-making in both private and public environments.

4.1.3. Abnormal Emotional Alert via Application

The real-time emotion detection system demonstrated its capability to analyze data from CCTV cameras and issue alerts when high-intensity or abnormal emotional states, such as extreme anger or fear, were detected. These alerts are transmitted directly to the administrator's application in the form of messages that include the camera ID, date, and time of detection, enabling timely monitoring and responsive action. As illustrated in the accompanying figure, the system detected abnormal emotions from CCTV cameras No. 1, 2, and 3 at different times, triggering automatic alerts displayed immediately through the web application interface. This alert mechanism serves as a critical feature that enhances the effectiveness of surveillance and situational management in high-risk environments, particularly in contexts such as educational institutions, government offices, or public spaces where proactive safety measures are essential, as shown in Figure 10.

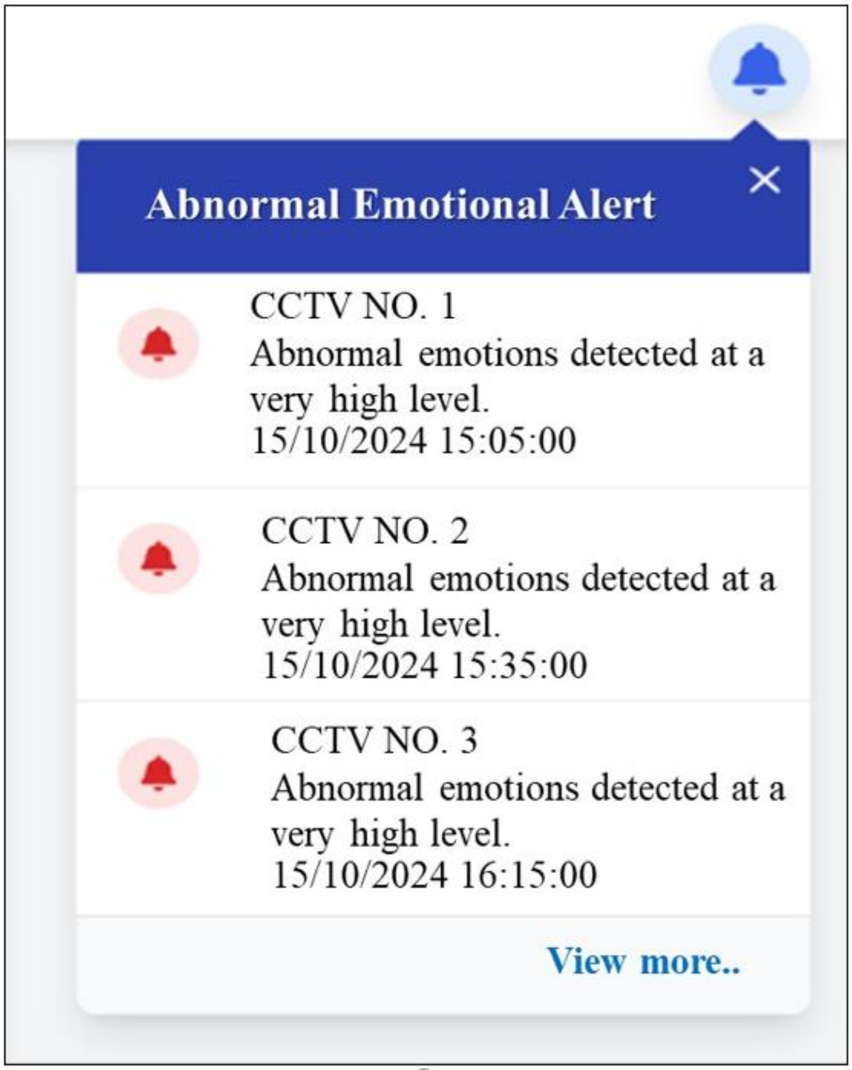

Figure 10.
Emotion Detection Alert.

From Figure 10, the system has detected multiple instances of abnormal emotional states from different CCTV sources. The summarized alert data is as follows: Table 1.

Table 1.
Emotional Notification.

No.	CCTV ID	Description	Date & Time
1	CCTV NO. 1	Abnormal emotions were detected at a very high level.	15/10/2024 15:05
2	CCTV NO. 2	Abnormal emotions were detected at a very high level.	15/10/2024 15:35
3	CCTV NO. 3	Abnormal emotions were detected at a very high level.	15/10/2024 16:15

4.1.4. Abnormal Emotional Notification via Email

In addition to real-time alerts through the web-based application, the system is also equipped with an email notification feature to enhance accessibility and responsiveness for administrators who may not be actively monitoring the dashboard. When the system detects an emotional trend that reaches a concerning level, such as a very high level of sadness, it automatically generates and sends an email alert to designated personnel. The email includes essential details such as the type and severity of the detected emotion, the specific CCTV camera involved, and the location of the incident (e.g., classroom or building area). As illustrated in the sample email, the system identified an elevated level of sadness from CCTV camera number 1, located in classroom 4919. This email-based alert mechanism plays a crucial role in extending the system's surveillance capabilities, ensuring that emotional anomalies can be promptly addressed even when the user is off-site, thereby supporting timely and effective intervention in real-world environments, as shown in Figure 11.

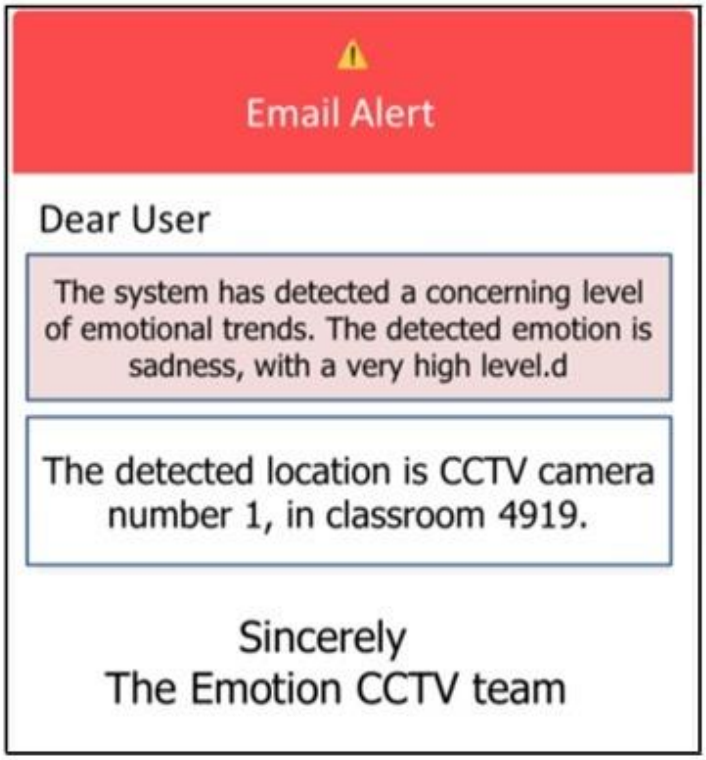

Figure 11.
Emotion Detection Email Alert.

4.1.5. Model Evaluation

The evaluation results of the developed model for facial emotion recognition revealed that the model achieved an accuracy of 0.8024, equivalent to 80.24%, and a loss value of 0.5542. These results indicate a satisfactory level of model performance in terms of both its ability to learn patterns from the data and to make accurate predictions, as shown in Figure 12.

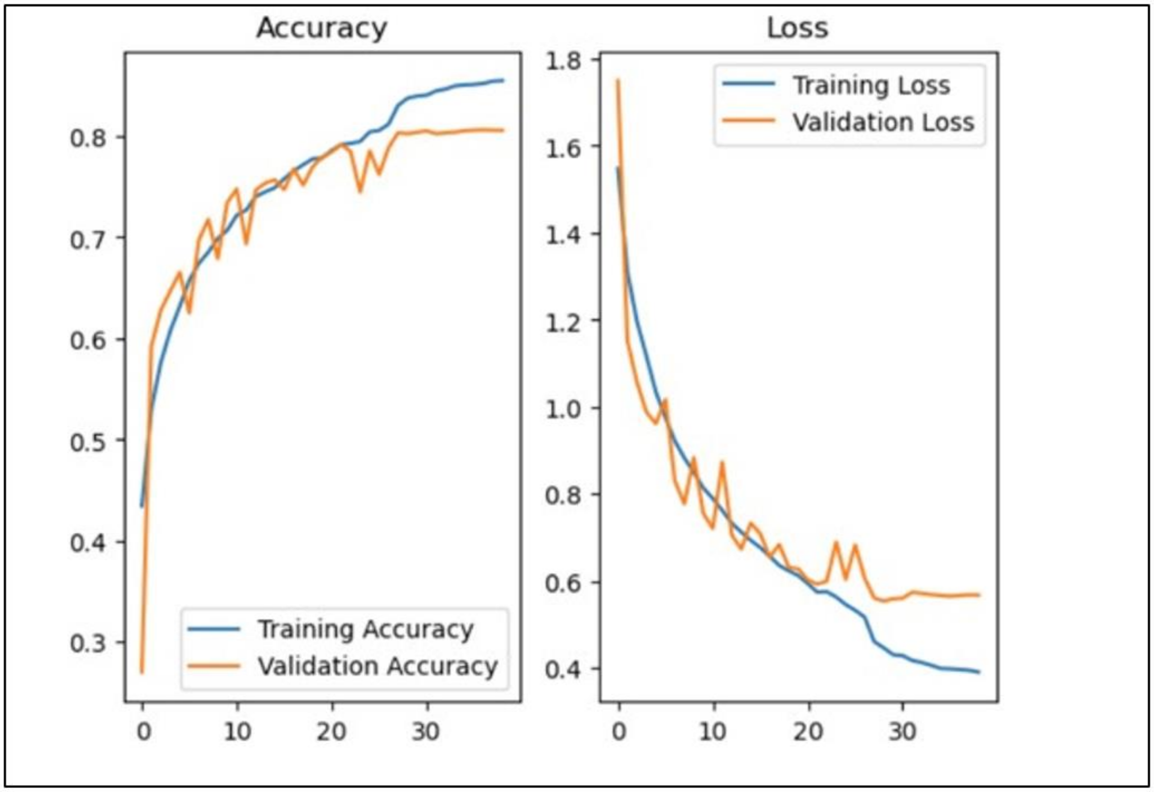

Figure 12.
Model Evaluation.

DOI: 10.55214/2576-8484.v9i10.10555

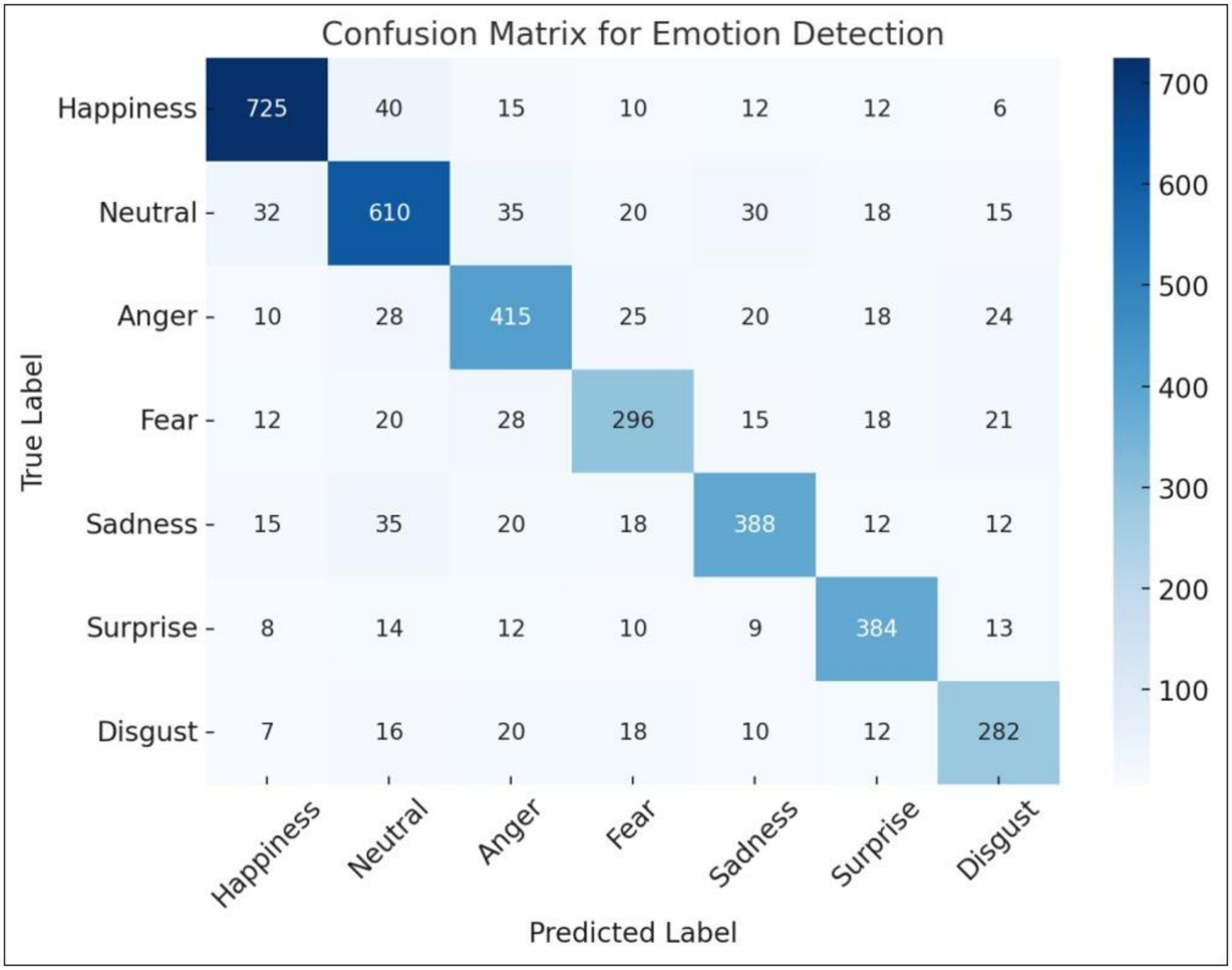

Figure 13.
Confusion Matrix.

Table 2.
Model Performance Evaluation Results.

Emotion Category	Precision (%)	Recall (%)	F1 score (%)	Support (images)
Happiness	85.12	88.45	86.75	820
Neutral	81.34	80.21	80.77	760
Anger	78.56	76.92	77.73	540
Fear	74.21	72.18	73.18	410
Sadness	79.85	77.64	78.73	500
Surprise	87.64	85.32	86.46	450
Disgust	75.42	74.1	74.75	380
Average	80.59	79.54	80.05	—

From Table 2, the model evaluation results indicate that the system can classify facial emotions with a high degree of accuracy. The average Precision, Recall, and F1 scores were 80.59%, 79.54%, and 80.05%, respectively. The emotions "Happiness" and "Surprise" achieved the highest precision values, reflecting the model's strong capability to recognize the distinctive features of these expressions effectively. Conversely, the emotions "Fear" and "Disgust" demonstrated slightly lower performance metrics, which can be attributed to the smaller number of samples in the dataset for these classes and their resemblance to other emotional expressions. These findings suggest that data balancing and the application of data augmentation techniques could further enhance the system's accuracy in future work.

4.2. Administrative Part

The administrative interface consists of five primary screens, designed to facilitate system oversight and user support:

- Admin Login Page: Used for administrator authentication before accessing system functionalities.
- Issue Management Page: Displays issues reported by users, allowing administrators to update issue statuses and notify users via automated email responses.
- User Management Page: Lists registered users and provides functionality to delete user accounts when necessary.
- Emotion Management Page: Enables administrators to add, edit, or remove emotion categories, and to configure system sounds for user selection.
- Admin Settings Page: Allows administrators to update their personal information and change their passwords.

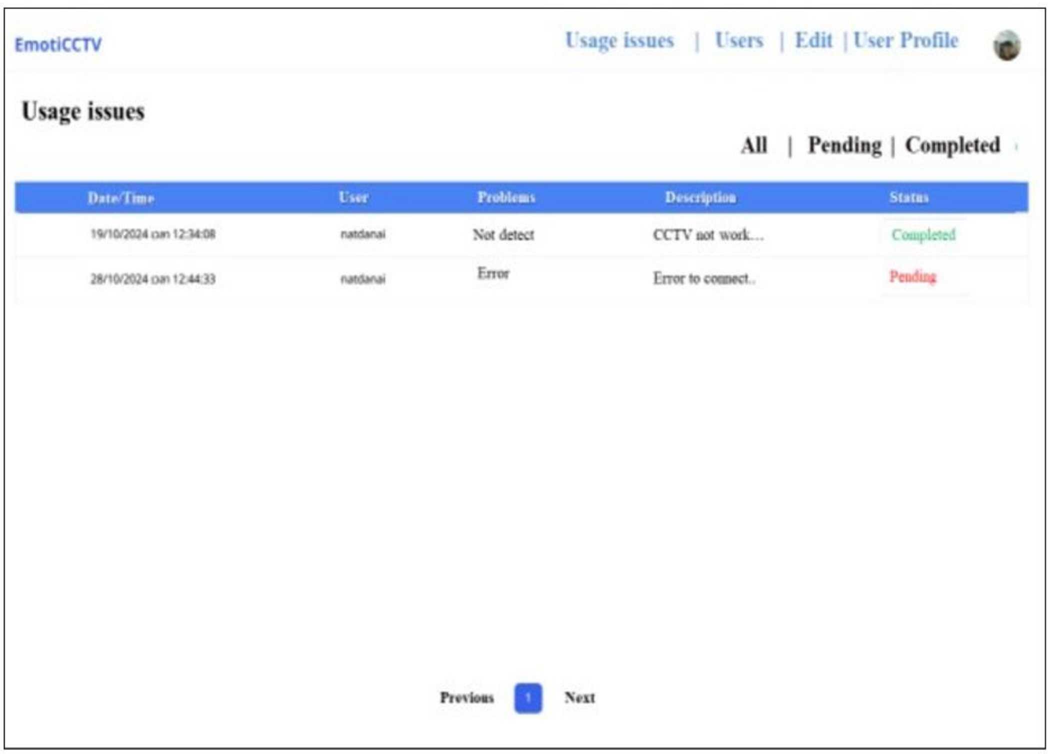

Figure 14.
Problems Page.

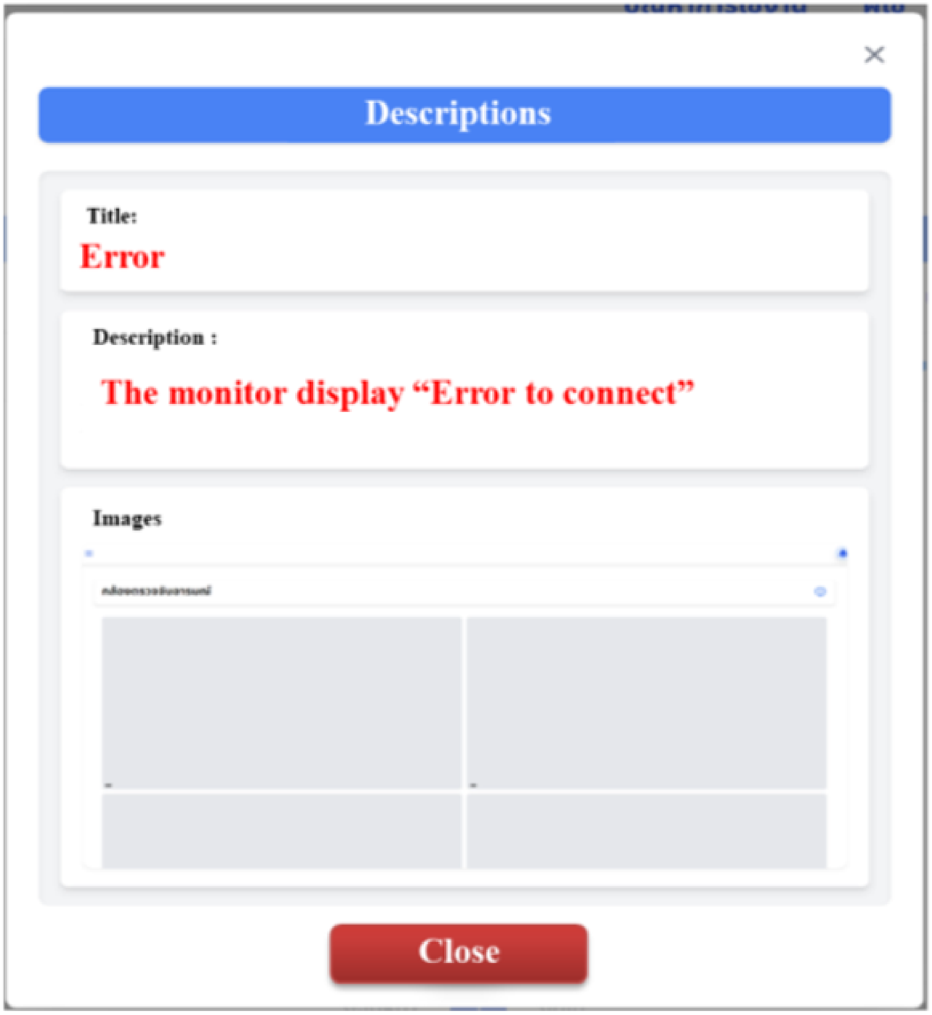

Figure 15.
Problem Description

5. Discussion

Although the system developed in this study demonstrates promising potential for real-time facial emotion recognition using CCTV video streams with satisfactory accuracy, it lacks a systematic comparison with other state-of-the-art models such as YOLOv8, MobileNet, and ViT, all of which are widely adopted in visual recognition and affective computing tasks. Benchmarking the proposed model against these architectures using the same datasets and presenting the results in comparative tables or performance graphs would enhance the interpretability of the model's relative strengths, weaknesses, and suitability for deployment in CCTV-based environments. Moreover, the evaluation of model performance based solely on accuracy and loss does not provide a comprehensive assessment, particularly in scenarios where the dataset suffers from class imbalance among emotion categories. Therefore, additional evaluation metrics such as Precision, Recall, F1 score, and Confusion Matrix should be incorporated to offer deeper insights into the model's classification capability and to inform targeted improvements.

At the same time, the system exhibits certain limitations associated with the quality of CCTV imagery, which may adversely affect performance. These include low image resolution, non-frontal facial angles, and suboptimal lighting conditions. Furthermore, the training datasets employed in this study present imbalanced emotion distributions, such as underrepresented classes like "fear" or "disgust," which can limit the model's learning effectiveness. To address these challenges, it is recommended to develop new datasets that more accurately reflect real-world scenarios and cultural diversity. Techniques such as data augmentation and the use of lightweight models optimized for low-resolution inputs should also be considered to enhance system robustness. In addition, ethical and privacy considerations must be thoroughly addressed, especially in public surveillance contexts. Clear notification regarding emotion detection, the avoidance of unnecessary personal identification, and adherence to privacy regulations such as the General Data Protection Regulation (GDPR) are essential to ensuring the responsible and sustainable deployment of this technology.

6. Conclusion

The experimental results indicate that the developed system can be effectively deployed in real-world contexts. However, to further enhance its capabilities and align with future demands, subsequent research should focus on advancing the system toward multimodal emotion recognition. This involves integrating multiple data sources such as facial expressions, vocal tone, and body movements to produce more accurate and contextually rich emotional assessments that better reflect real human behavior.

In addition, embedding artificial intelligence models into edge computing devices such as Jetson Nano or Raspberry Pi can significantly improve system responsiveness by enabling real-time data processing without relying on cloud-based infrastructure. This not only reduces latency but also enhances data privacy and makes the system more suitable for environments with limited network connectivity. Pursuing these research directions will contribute to the long-term sustainability, efficiency, and ethical deployment of emotion-aware technologies across diverse application domains.

The authors also wish to extend their appreciation to the experts and editors of the Edelweiss Applied Science and Technology Journal for their insightful comments and constructive suggestions, which have greatly contributed to improving the quality and clarity of this manuscript.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgements:

The authors would like to express their sincere gratitude to the Faculty of Digital Technology, Chiang Rai Rajabhat University, Thailand, for providing the research facilities and technical support that made this study possible. This work was partially supported by the Computer Science and Information Technology Program, Faculty of Digital Technology, Chiang Rai Rajabhat University, as part of its ongoing commitment to academic development and innovation.

Copyright:

References

[1] A. Gupta, R. Mehta, and S. Kapoor, "Real-time emotion recognition for surveillance applications: A comprehensive review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4125–4138, 2022.

[2] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: A survey," *Symmetry*, vol. 11, no. 10, p. 1189, 2019. https://doi.org/10.3390/sym11101189

[3] H. Lian, C. Lu, S. Li, Y. Zhao, C. Tang, and Y. Zong, "A survey of deep learning-based multimodal emotion recognition: Speech, text, and face," *Entropy*, vol. 25, no. 10, p. 1440, 2023. https://doi.org/10.3390/e25101440

[4] T. 4Kopalidis, 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, 2024. https://doi.org/10.3390/info15030135
[5] W. J. Han, "Deep learning based detector for real-time facial expression recognition," M.S. Thesis, Nanyang Technological Univ., Singapore, 2019.
[6] W. H. Abdulsalam, R. Al-Hamdani, and M. N. Abdullah, "Facial emotion recognition from videos using deep convolutional neural networks," ResearchGate, 2019. https://www.researchgate.net/publication/371021729_Real_Time_Facial_Emotion_Recognition_using_Deep_Learning_and_CNN
[7] B. Cheng *et al.*, "Robust emotion recognition from low quality and low bit rate video: A deep learning approach," presented at the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), 2017.
[8] S. S. A. Rizvi, S. Kumar, A. Seth, and P. Narang, "AffectSRNet: Facial emotion-aware super-resolution network," *arXiv preprint arXiv:2502.09932*, 2025.
[9] S. Wang and X. Li, "Dynamic resolution guidance for facial expression recognition," presented at the International Conference on Pattern Recognition, 2024.
[10] P. P. Rodrigues, D. M. Silva, and A. P. Silva, "Web-based platform for emotion recognition using deep learning and facial analysis," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 10, pp. 9839–9852, 2021.
[11] H. L. Pham, T. T. Nguyen, and T. H. Nguyen, "A web-based real-time emotion recognition system for user experience evaluation," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Anal. (DSAA), Sydney, Australia*, 2020.
[12] S. Ghosh, M. Ghosh, and A. Konar, "Real-time emotion recognition and trend analysis using web-based applications," *International Journal of Human–Computer Interaction*, vol. 38, no. 14, pp. 1351–1365, 2022.
[13] H. Wang, J. Sun, and T. Wu, "Web-based facial analysis platform: Design and implementation," *Journal of Web Engineering*, vol. 21, no. 3, pp. 455–472, 2022.
[14] X. Zhang, Y. Pan, and L. Zhang, "Deep learning-based facial emotion recognition: A survey and benchmarking," *Information Fusion*, vol. 84, pp. 146–170, 2022.
[15] J. Wang, K. Zhang, and S. Zhao, "Real-time facial expression recognition using FER2013 and transfer learning," *IEEE Access*, vol. 11, pp. 20312–20321, 2023.
[16] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18-31, 2019. https://doi.org/10.1109/TAFFC.2017.2740923
[17] J. Chen, X. Li, and Y. Xu, "A lightweight CNN model for real-time facial expression recognition," *Neurocomputing*, vol. 496, pp. 74–85, 2022.
[18] J. Park, K. Kim, and H. Lee, "Quantized FER for edge devices," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8765–8774, 2021.
[19] Z. Liu, X. Wang, and H. Ma, "Optimizing facial emotion recognition using MobileNet with quantization and pruning," *Pattern Recognition Letters*, vol. 167, pp. 80–88, 2023.
[20] Y. Zhao, L. Wang, and X. Liu, "YOLO-FER: Real-time multi-person emotion detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 1, pp. 45–58, 2023.
[21] C. Li and X. Wang, "EfficientViT: A lightweight transformer for real-time facial emotion recognition," *IEEE Access*, vol. 10, pp. 34567–34578, 2022.
[22] O. C. Oguine, K. J. Oguine, H. I. Bisallah, and D. Ofuani, "Hybrid facial expression recognition (FER2013) model for real-time emotion classification and prediction," *arXiv preprint arXiv:2206.09509*, 2022. https://doi.org/10.48550/arXiv.2206.09509
[23] Y. Zhang, H. Wang, and J. Liu, "Attention-based ResNet for facial emotion recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 512–525, 2022.
[24] R. Kumar, S. Gupta, and A. K. Singh, "Adaptive lighting correction for FER in surveillance," *IEEE Transactions on Image Processing*, vol. 30, pp. 1234–1245, 2021.
[25] E. Martinez, R. Sanchez, and L. Garcia, "Explainable AI for FER in security systems," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 2, pp. 123–134, 2022.
[26] T. Nguyen, V. Patel, and R. Chellappa, "Occlusion-robust FER using GAN-based data augmentation," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7890–7901, 2021.
[27] M. Alshammari and T. Alshammari, "Real-time emotion recognition using YOLOv8 and edge computing for optimized performance," *Computers*, vol. 12, no. 6, p. 117, 2023.
[28] A. Gupta, P. Sharma, and N. Verma, "Web-based real-time FER using TensorFlow.js," *IEEE Transactions on Multimedia*, vol. 24, pp. 321–332, 2022.
[29] C. Li, X. Wang, and S. Zhang, "Lightweight CNN for real-time emotion recognition on edge devices," *IEEE Access*, vol. 9, pp. 12345–12356, 2021.
[30] Y. Liu, Q. Zhang, and M. Sun, "Edge-based real-time emotion recognition: Applications and challenges," *IEEE Internet of Things Journal*, vol. 10, no. 15, pp. 13500–13512, 2023.

[31] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International Conference on Neural Information Processing, pp. 117-124. Berlin, Heidelberg: Springer Berlin Heidelberg*, 2013.

[32] J. Park, H. Choi, M. Song, and K. Kim, "Quantization-aware training for edge-based FER in CCTV systems," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9876–9889, 2023.

[33] K. Zhang, L. Zhou, and F. Wang, "Deep learning approaches for facial emotion recognition in surveillance," *Pattern Recognition Letters*, vol. 145, pp. 1–12, 2021.

[34] S. Lee and H. Kim, "Privacy-preserving FER using federated learning for smart surveillance," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1255–1269, 2023.

[35] S. Lee, J. Park, and H. Kim, "Federated learning for privacy-preserving FER," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2345–2358, 2022.

[36] J. Chen, Y. Wang, and F. Li, "Hybrid transformer-CNN for real-time facial expression recognition," *IEEE Transactions on Affective Computing, Early Access*, 2023.

[37] S. Li, H. Zhou, and J. Chen, "Challenges in deploying facial expression recognition for surveillance systems," in *Proc. ACM Int. Conf. Multimedia (ACM MM)*, 2020.