Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Integrating Emotion Recognition in IoT-based Cameras: A Deep CNN Approach for Real-time Image Capture

Jermine G. Valen-Dacanay

Doctor in Information Technology Student, Department of Information Technology, College of Information Technology and Computer Science, University of the Cordilleras, Baguio City, 2600, Philippines, Email: jgv7470@students.uc-bcf.edu.ph

Real-time image capture using Internet of Things (IoT) cameras plays an important part in data collection for security and public safety applications. However, relying primarily on visual data limits the opportunity to gain a deeper understanding of human behavior, which may be an additional component in detecting possible crises. The integration of emotion recognition modules in a streaming video for real-time image capture is the main objective of this research. The goal is to detect and interpret human emotions in real-time to provide proactive warnings and information to improve situational awareness. The system analyzes facial expressions through captured images from streaming monitoring video to detect emotions in real time, thereby enhancing emergency sensing with an AI-driven system. The methodology uses machine learning for image classification and emotion detection. This is to detect emotional states based on facial expressions. Convolutional Neural Network (CNNs) architecture and structure was used to recognize and understand human emotions in real time. The model was trained and validated on publicly available facial emotion datasets, such as FER2013. Its performance was assessed using parameters such as accuracy, precision, and recall. The OpenCV tool was used to conduct real-time image processing and computer vision tasks while streaming protocols such as RTSP permitted network-based real-time video access. The results demonstrate that the system effectively recognizes a range of emotions when integrated into an IoT-based camera. The CNN algorithm integrated in image and emotion detection provides an acceptable performance in terms of human emotion detection capability. How do its potential to improve situational awareness in high-risk areas by providing timely alerts and notifications. Integrating the emotion recognition model with IoT-based cameras shows significant promise in enhancing situational awareness. The system's ability to recognize emotions in real time can potentially improve security and public safety by triggering timely alerts and notifications.

Keywords: Deep learning optimization, Convolutional Neural Network (CNN), Internet-of-Things systems, Facial expression recognition (FER), real-time image detection.

1. Introduction

The use of emotion detection technology in monitoring cameras within a university environment presents a unique opportunity to enhance campus safety and student well-being. Universities, such as Urdaneta City University in Pangasinan, Philippines, are dynamic places with a wide range of activities, making them ideal for testing advanced monitoring technologies. Generally, the common emergency crisis in a university setting includes health and medical issues such as injuries, sudden illnesses, and mental health crises in some cases. Safety concerns such as riots, assaults, and active shooter situations are also prevalent, especially in high-risk areas. Additionally, universities face environmental hazards like fires, extreme weather, and earthquakes. Infrastructure failures, including power outages, gas leaks, and technical data breaches, pose serious risks, while social emergencies and incidents involving hazardous substances in science labs further complicate campus security.

Early identification of emergencies is essential to ensure public safety and minimize damage. Rapid response enables emergency responders to act faster, potentially saving lives and reducing injuries. The immediate response of authorities related to this is crucial in saving lives. The ability to respond quickly and effectively is what makes emergency disaster recognition so important. The sooner an issue is detected, the sooner authorities can take steps to mitigate its impact. Issuing like issuing alerts when a suspect is seen loitering in a restricted area or triggering alarms when a suspicious person is roaming around. Furthermore, early detection enhances preparedness by allowing authorities to allocate resources first and initiate a response, ensuring rapid and effective response in the event of a crisis [1]. Timely public warnings also empower individuals to take necessary precautions and ensure their safety [2].

Traditional methods of Emergency Crisis Detection (ECD), such as human surveillance and manual alerts, have significant limitations in today's complex environment. Human boundaries, like fatigue and gradual response times, can result in ignored or behind-schedule responses. These methods additionally face scalability problems, as manually tracking large amounts of information is not possible [3]. Furthermore, human analysis of raw data can be slow and error-prone, hindering early detection of warning signals [4]. Modern ECD technologies leverage sensor networks, social media analytics, and satellite imagery to overcome these challenges [5]. Sensor networks monitor environmental conditions and detect anomalies, while social media analytics use AI to analyze real-time data for trends and anomalies indicating potential threats [6]. Satellite imagery provides a broad view of large-scale events.

Artificial intelligence (AI) has revolutionized early disaster detection by enabling faster, more accurate, and larger-scale threat detection. AI algorithms deal with large volumes of facts quickly and correctly, addressing scalability problems. Techniques that include sentiment evaluation and natural language processing (NLP) can be applied to social media facts to identify sentiment signs related to potential emergencies [7]. Environmental sensors collect data on water levels, seismic activity, air quality, and other topics, which AI algorithms analyze to detect irregular changes [8]. Satellites offer a bird's-eye view of large-scale events like wildfires, floods, or massive human displacement, with AI identifying patterns and changes indicating emergencies [9]. By integrating data from various sources, AI systems provide a comprehensive and accurate picture of potential threats, enhancing the reliability of early warnings [10].

Emotion Recognition (ER) technology can further enhance AI-powered ECD systems by identifying emotional cues such as fear, anxiety, and anger from social media posts or faces [15]. This additional information aids in earlier and more accurate risk identification. Despite advancements in identifying basic emotions [11], challenges remain in emotion recognition technology, such as difficulty in identifying subtle emotions, cultural differences in expression [12], and the impact of age-related changes [13] or accessories like eyeglasses [14] on accuracy. Current systems perform best under controlled conditions with well-defined datasets, but future studies may improve accuracy by integrating text-based data from social media posts and audio analysis with facial expressions.

The main objective of this study is to incorporate the emotion recognition resulting from visuals of IoT-based cameras. This study aims to investigate whether integrating emotion recognition with real-time video streams in IoT-based cameras provides acceptable emotion accuracy in uncontrolled environments. Further, this study aims to investigate whether emotion recognition combined with real-time video output from IoT-based cameras results in acceptable emotional detection accuracy in unregulated environments. It explores implementation strategies for emotion recognition modules in IoT-based cameras within a university setting, focusing on enhancing campus safety and student well-being. This approach aims to improve traditional monitoring systems by providing deeper insights into human behavior through indirect analysis of facial expressions. This research uses machine learning techniques, specifically convolutional neural networks (CNNs), to recognize and interpret human emotions. The system uses this algorithm to detect urgency based on facial emotion cues. In addition, the study uses publicly available datasets such as FER2013 to train and test the emotion detection algorithms. The CNN model performance is evaluated based on metrics such as accuracy, precision, and recall. The study will use RTSP and other related streaming protocols to capture real-time video over the Internet with OpenCV for real-time image processing tasks. This solution aims to improve the overall performance of the university security monitoring system to aid authorities in case of emergencies.

2. Methods

Presented in this part are the systems development methodology, and the experimental setup of the study.

Systems Development Methodology

The development of an AI-driven system for emotion recognition involved the combination of artificial intelligence modules and IoT-based cameras. The objective was to enable these cameras to detect and analyze human emotions in real-time video. This collaboration required the installation of hardware and software necessary to capture, process, and analyze video data in real-time. IoT-based cameras equipped with high-resolution sensors and internet connectivity have been deployed in strategic locations within a specific area inside the university campus. Fig. 1 presents the project methodology.

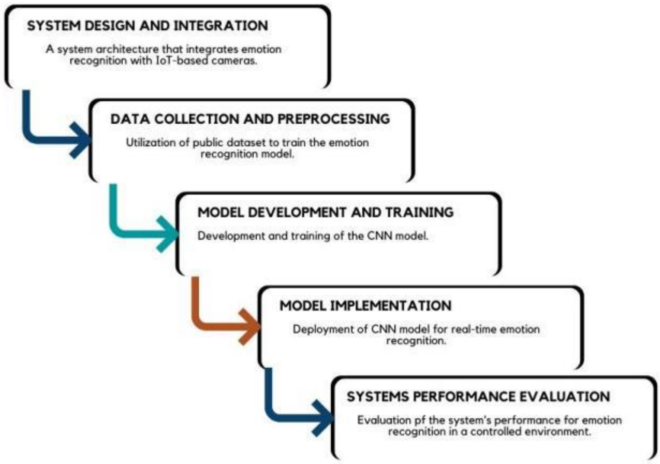

Fig. 1. Project Methodology

Design and Integration. The first phase involves developing a system architecture that integrates emotion sensing and IoT-based cameras. This includes selecting appropriate IoT cameras equipped with high-resolution sensors and internet connectivity to enable real-time video streaming. Cameras are placed at strategic locations in the university to capture relevant data for analysis.

Data Collection and Preprocessing. In this phase, the study will utilize publicly available data such as FER2013 to train the emotion recognition model. The collected images are pre-processed, which includes scaling, normalization, and augmentation, to ensure that the model can recognize as many cases as possible. This phase is important to increase the robustness and validity of the model when deployed in a monitoring system.

Model Development and Training. This phase focuses on the development and training of the CNN model. From the FER2013 dataset, images are preprocessed for training and then trained for emotion classification. To ensure the learning process of the model, an assessment of its performance is performed using measures such as accuracy(1), precision (2), and recall (3).

Model Implementation. Implementation is done by integrating the CNN model into the monitoring system. This is to test the model for real-time image capture and emotion classification. OpenCV is utilized for processing of the captured images. The system uses streaming protocols such as Real-Time Streaming Protocol (RTSP) to provide seamless video transmission and real-time access to the Internet.

Systems Performance Evaluation. This phase focuses on the evaluation of the system in detecting emotions from a real-time scenario. Evaluation focuses on the system's ability to accurately detect faces and classify emotions. Evaluating system performance in a controlled environment is the final step. The response time and the number of captured faces or frames are the focus of evaluation.

Experimental Setup

Model development and training consists of identifying first the architecture of the CNN model. Then, the model is trained using the FER2013 dataset. Enhancement techniques on the images are used like rotation, flipping, and scaling increase the variety of training data. Emotion recognition is done using a convolutional Neural Network (CNN) technique in VGG-16 architecture. Model accuracy, recall, and training accuracy are all evaluated using the TensorFlow framework. Hyperparameters such as the number of batches, and number of epochs are tuned to improve the performance of the models.

The system architecture consists of deploying the Internet of Things (IoT) cameras in controlled locations to capture and transmit video. A central server, connected to a reliable network, collects data from the IoT device and transfers it to an application for preprocessing. An IoT camera collects and processes data in advance by continuously capturing video feeds. Faces are extracted from videos using OpenCV's facial recognition algorithms. The captured facial photos are then processed normally, scaled to 48 by 48 pixels to comply with the image structure where the model was trained like the FER2013 dataset and turned to grayscale.

3. Results

This part presents the proposed emotion recognition model in VGG-16 architecture, the emotion recognition network structure, the model evaluation after the integration to real-time emotion recognition, and the proposed application of the developed ER module.

The emotion recognition module

This part discusses the structure of the emotion recognition module. This module comprises a VGG16 architecture trained in FER2013 dataset.

Data Used. This study made use of the FER2013 dataset. A total of 28,709 grayscale images, each measuring 48 by 48 pixels, were utilized. The labels on the images represented seven different emotions: surprise, anger, disgust, fear, happiness, sadness, and neutrality. These data sets are valuable due to their vastness and diversity, which makes them perfect for building complex models applicable to real-world situations. The dataset composition used in this study is shown in Table I.

Table 1. Dataset Composition For Training And Testing

CLASSES	TRAINING	TESTING	Total per Class
Angry	3196	799	3995
Disgust	349	87	436
Fear	3278	819	4097
Happy	5772	1443	7215
Neutral	3972	993	4965
Sad	3864	966	4830
Surprise	2537	634	3171

The 28,709 images were split for training and model testing. This dataset is split into training and testing sets, providing a robust foundation for training and evaluating deep learning models. Dataset splitting follows eighty percent for training and the remaining 20% for testing or 22,967 images for training and 5,742 images for testing.

Model Structure. The VGG16 Convolutional Neural Network (CNN) architecture is renowned for its ease of use and excellent picture classification results. The University of Oxford's Visual Geometry Group first presented the deep convolutional neural network architecture, or VGGNet, in a 2014 study [16]. It is noted for its simplicity and depth. It consists of 16 weight layers, including 13 convolutional layers and 3 fully connected layers. The convolutional layers utilize small 3x3 receptive fields, enabling the network to learn intricate patterns while maintaining computational efficiency [17]. The emotion recognition module is designed in VGG16 architecture. Fig. 2 presents the model architecture. This CNN model with all 33,624,775 parameters successfully handled in the learning process, with a total memory footprint of 128.27 MB. The deployment of CNNs in real-world applications is often constrained by model size and run-time memory requirements [26]. Model size is a challenge as the millions of trainable parameters and network structure information need to be stored on disk and loaded into memory during inference, which can be a significant resource burden for embedded devices [26].

Layer (type)	Output Shape	Param #
conv2d_39 (Conv2D)	(None, 48, 48, 64)	640
conv2d_40 (Conv2D)	(None, 48, 48, 64)	36,928
max_pooling2d_15 (MaxPooling2D)	(None, 24, 24, 64)	0
conv2d_41 (Conv2D)	(None, 24, 24, 128)	73,856
conv2d_42 (Conv2D)	(None, 24, 24, 128)	147,584
max_pooling2d_16 (MaxPooling2D)	(None, 12, 12, 128)	0
conv2d_43 (Conv2D)	(None, 12, 12, 256)	295,168
conv2d_44 (Conv2D)	(None, 12, 12, 256)	590,080
conv2d_45 (Conv2D)	(None, 12, 12, 256)	590,080
max_pooling2d_17 (MaxPooling2D)	(None, 6, 6, 256)	0
conv2d_46 (Conv2D)	(None, 6, 6, 512)	1,180,160
conv2d_47 (Conv2D)	(None, 6, 6, 512)	2,359,808
conv2d_48 (Conv2D)	(None, 6, 6, 512)	2,359,808
max_pooling2d_18 (MaxPooling2D)	(None, 3, 3, 512)	0
conv2d_49 (Conv2D)	(None, 3, 3, 512)	2,359,808
conv2d_50 (Conv2D)	(None, 3, 3, 512)	2,359,808
conv2d_51 (Conv2D)	(None, 3, 3, 512)	2,359,808
max_pooling2d_19 (MaxPooling2D)	(None, 1, 1, 512)	0
flatten_3 (Flatten)	(None, 512)	0
dense_9 (Dense)	(None, 4096)	2,101,248
dropout_6 (Dropout)	(None, 4096)	0
dense_10 (Dense)	(None, 4096)	16,781,312
dropout_7 (Dropout)	(None, 4096)	0
dense_11 (Dense)	(None, 7)	28,679

Total params: 33,624,775 (128.27 MB)

Trainable params: 33,624,775 (128.27 MB)

Non-trainable params: 0 (0.00 B)

Fig. 2. The model architecture summary

Model performance in training and validation process. The VGG16 model for emotion detection was trained with the following specifications: categorical cross-entropy loss function, Adam optimizer, and a batch size of 64 with early quitting and learning rate reduction callbacks to block overfitting. The generalization capability of the model was enhanced by using data enhancement techniques such as rotation, zooming, and horizontal flipping. These techniques helped to identify the invariance of the model and made significant adjustments to the changes in the input images. Fig. 3 presents the model performance over training in 100 epochs.

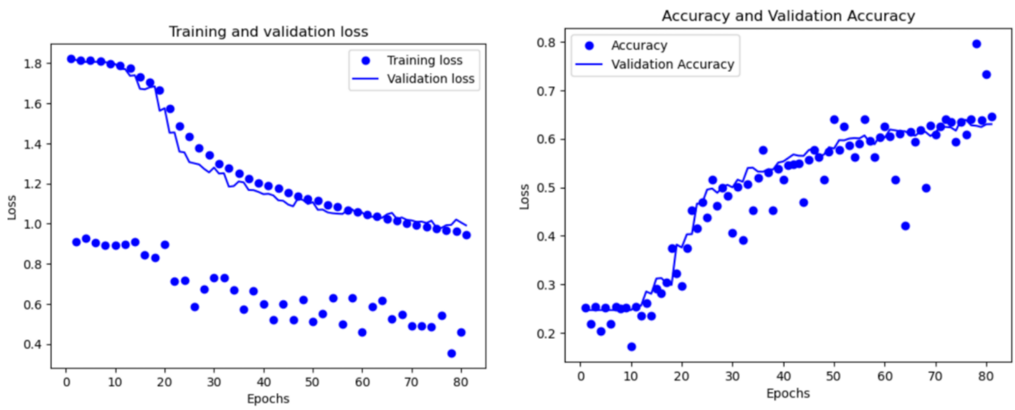

Fig. 3. Model performance during the training and testing process.

The model achieved a training accuracy of 0.6533 or 65.33 percent and a validation accuracy of 0.6414 or 64.14 percent. These metrics demonstrate how well the model can classify emotions in the FER2013 dataset. The close values of training and validation accuracy indicate that they are a refined model that is not overfitting. The accuracy of 65.33% on the training set and 64.14% on the validation set is a performance that the CNN model was able to accurately identify emotions of images from the dataset and has an average performance in detecting facial images and classifying emotions. During training, some accuracy results reached 93% and 88% respectively, but with a validation result of 65%. It depicts overfitting during model training.

The confusion matrix was another measure used in this study to assess how well the VGG model identified emotions. It is a useful tool for evaluating the performance of a classification model which provides a general overview of the prediction accuracy of the model across multiple classes [19, 20, 21]. Fig. 4 shows the performance of the model during the training and validation process in FER2013.

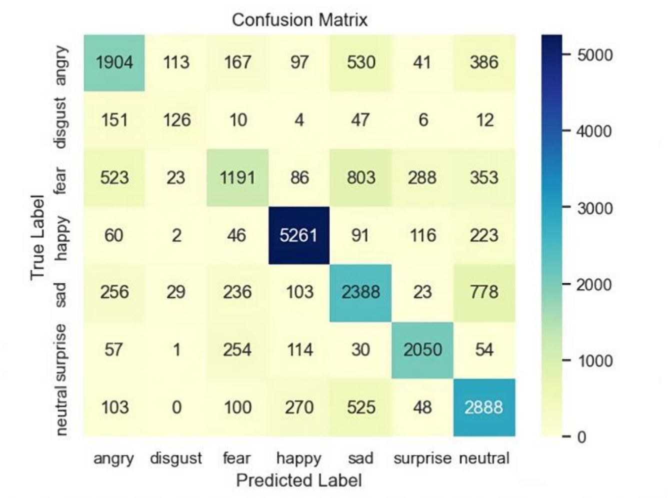

Fig. 4. VGG model prediction across seven (7)emotion classes

The confusion matrix was used to calculate performance indicators including accuracy, precision, recall, F1- score, which indicates the number of true positives, true negatives, false positives, and false negatives for each side [20, 22]. This graph displays how the model performs in determining emotions from the dataset. As shown in Figure 4, the model's performance in identifying and classifying seven (7) emotion categories in the FER2013 dataset is illustrated. Table 2 presents the model performance in classifying emotions using accuracy, recall, and precision metrics.

IoT-based emotion recognition network structure.

Monitoring using IoT-based cameras with emotion recognition modules is divided into layers, each performing separate but complementary functions that ensure efficient and accurate monitoring. Fig. 5 presents the IoT-based emotion recognition network structure.

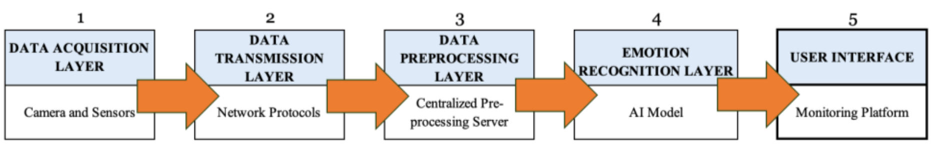

Fig. 5. IoT-based emotion recognition network structure.

The data acquisition phase is the foundation, with IoT-enabled cameras being strategically deployed in the monitored area. These cameras capture continuous video, providing the real-time visual information needed for later use. The captured video data then passes through the data transmission layer, where it is securely and reliably transferred to a central server. This layer uses communication protocols such as HTTP and RTSP. Once at the server, the data enters the data preprocessing layer. Modern computer vision algorithms such as OpenCV are used to perform facial recognition and extraction from videos. Subsequently, the preprocessed photos are loaded into a specific machine-learning model trained with CNN VGG-16 on the FER2013 dataset for emotion recognition. This model analyzes the facial features in the images and classifies the emotions, providing a probability distribution for each emotion class. The final step is to display the forecast in the user interface. A web browser built with Streamlit framework will serve as an interface for real-time visualization of observed emotions.

System Evaluation

The CNN model utilized in this study to recognize emotions had a moderate level of detection ability. F1-score, accuracy, precision, and recall, are examples of performance measures for machine learning models. The confusion matrix [21, 23, 24] serves as the foundation for the methodology for these metrics. The computations below are based on the values generated from the performance of the CNN model shown in Figure 4. The important metrics, such as accuracy, precision, and recall, were used in this study to assess the convolutional neural network (CNN) model's performance in emotion identification.

Accuracy. The model's accuracy of 0.65 in this instance indicates that it accurately predicts emotions in 65% of the datasets. By comparison, considering the size of the dataset, this suggests that there is still improvement above random chance. Accuracy is computed as the sum of true positives and true negatives over the total number of predictions [21, 24]. Accuracy is computed as:

$$\text{Accuracy} = (TP + TN) / \text{SSamples} \qquad (1)$$

Here, (1) TP is the true positive, TN is the true negative, and SSamples as the total number of samples or summation of samples used in either training or testing [21, 24, 25].

Precision. The precision is 0.64, meaning that 64% of the time the model accurately predicts an emotion. In situations where false positives are a significant or troublesome feature, accuracy is crucial. The proportion of true positives among the predicted positives resulted in precision [24]. Precision is computed as:

$$\text{Precision} = TP / (TP + FP) \qquad (2)$$

Here, (2) TP is the true positive, and FP is the false positive [21, 24, 25].

Recall. Recall measures the accuracy of the model to identify any issues associated with a particular emotion. To determine the model’s ability to identify all positive instances, a recall metric is used [21,24]. The CNN model utilized in this study has a recall value of 0.65 showing that the model can detect 65% of the real data. A high recall value to detect real emotions efficiently. This is an important factor when it is applied to detecting urgency or emergency crises. Recall is computed as:

$$\text{Recall} = TP / (TP + FN) \qquad (3)$$

Here, (3) TP is the true positive, and FN is the false negative [21, 24, 25]. Table II presents the computation of the model performance across the seven (7) classes of emotion in the FER2013 dataset.

Table 2.Model Performance Across Seven (7) Classes Of The Fer2013 Dataset

Class	Accuracy	Precision	Recall
Angry	0.4766	0.6234	0.5880
Disgust	0.2890	0.4286	0.3539
Fear	0.2907	0.5943	0.3646
Happy	0.7292	0.8864	0.9072
Sad	0.4810	0.5410	0.6263
Surprise	0.4244	0.7970	0.8008
Neutral	0.9108	0.6153	0.7341

Real-time image capture performance. The real-time performance of the emotion recognition system is demonstrated through performance testing with an Internet of Things-based camera. This section discusses the CNN model's performance when integrated with real-time image capturing. The image was captured using a streaming platform specifically a web application where an IoT-based camera is accessed. TABLE III presents the specifications of the camera utilized in this study.

Table 3. Camera Specifications

Camera Specifications	
Resolution:	720p (1280 x 720 pixels)
Aperture:	f/2.0 (approximate)
Video Recording:	Supports 720p video recording at 30 fps
Integrated Microphone:	Three-mic array with directional beamforming for improved voice capture

The system captured and processed 157 frames in 30 seconds, at an average capture rate of 5.07 frames per second (fps). Using this frame rate, the system can quickly extract and evaluate individual frames in the video input, allowing it to quickly detect emotion. Fig. 6 illustrates the real-time performance of the system.

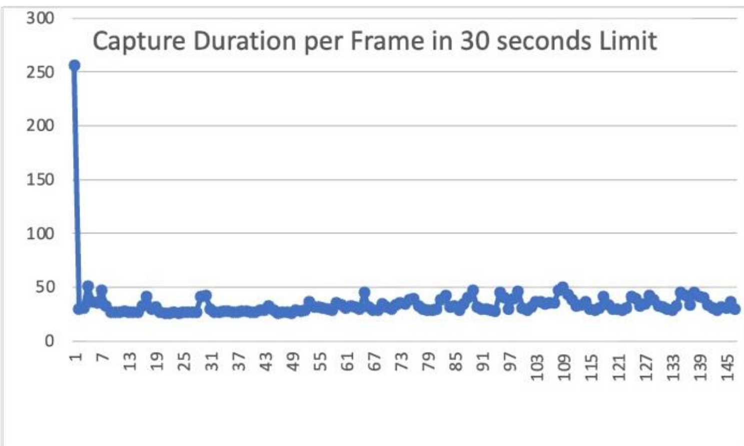

Fig. 6. Real-time capture performance of the system.

The ability of the system to capture and process 157 frames in thirty (30) seconds, at an average frame rate of 5.07 or approximately five (50 frames per second (fps), highlights its effectiveness in processing video input for emotion recognition in a real-time manner. The frame rate of a system is a crucial factor in determining its effectiveness in real-time applications that require timely analysis and response. In some cases, these applications have strict end-to-end quality-of-service requirements, and the usage of real-time techniques can be beneficial. Recent research has highlighted the significance of frame rate in the context of augmented reality (AR) head-mounted displays (HMDs), where the rate at which AR objects are rendered and displayed is a key determinant of device energy consumption and user experience[27].

The system can quickly recognize emotions from video input by processing each frame at that speed, which improves response and makes it more suitable for situations that may arise. For example, in the context of security, the technology enables real-time detection of stress or suspicious behavior and allows for constant monitoring of individual moods.

Proposed Application of ER Module in Emergency Monitoring System

This study proposes integrating crisis detection with an emotion recognition system. Fig. 7 shows the procedure for detecting a crisis in the form of a flow diagram. This process flowchart depicts the integration of an emotion recognition module into an IoT-based camera network to detect potential emergency crises.

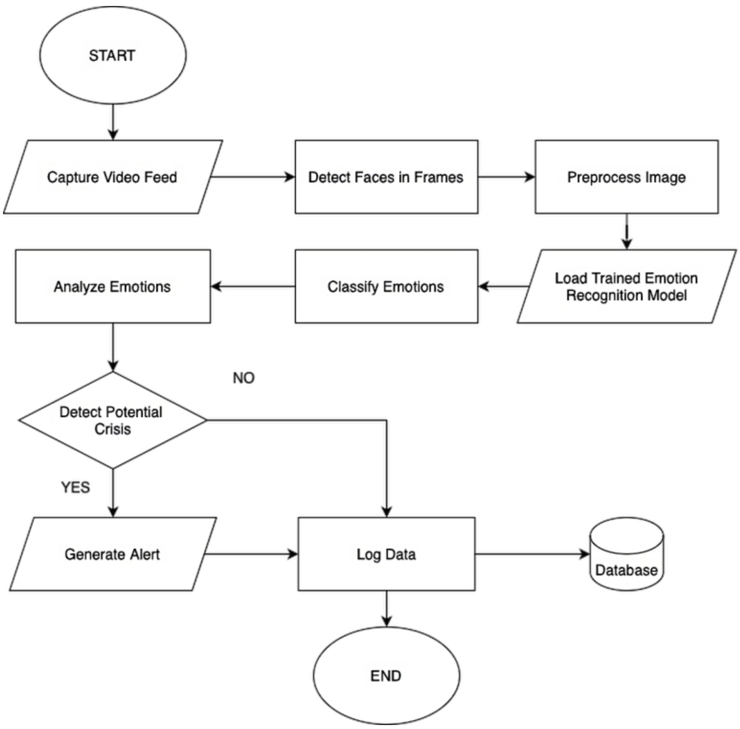

Fig. 7. Crisis detection process with integrated emotion recognition module

At the beginning of the process, an IoT-based camera continuously captures video footage from its surroundings. Multiple faces in the frame are then detected in real time by processing these feeds with a face recognition system. To prepare images for analysis, the system uses an emotion recognition model trained as previously discussed. For this purpose, the emotion recognition module was based on the VGG-16 architecture, which was trained on the FER2013 data set. Then, pre-processed images of this image are provided to classify emotions. The identified emotions are submitted for further analysis to look for patterns that may indicate anxiety, fear, or other emotions that often precede emergencies.

The most important decision in the flowchart is to detect potential crises. Here, the system examines whether the analyzed emotions indicate a possible emergency. If no threat is detected, the system falls back to begin collecting a new video feed and processes it by recording data for logging and performance monitoring but, if a potential threat is detected, the system defaults channel Generate Alert is used. Finally, regardless of whether an alarm is triggered or not, the system records information about emotions, decisions, and known behaviors for future research and development. This provides information on the long-term performance of the model. The process is then completed, ready to be repeated from successive video frames to maintain stable attention and feedback.

4. Discussion

With a capability of 5.07 frames per second (fps), the device is ideal for situations that require real-time monitoring, such as public safety or checkpoints. This rate ensures that the system captures enough frames per second to provide a comprehensive and continuous understanding of the monitored environment while not missing important times that could signal an emergency.

The average emotion detection time per cycle is 34.55 milliseconds (ms). This rapid identification rate demonstrates the efficacy of the convolutional neural network (CNN) model employed for emotion recognition. This speed performance assures timely emotion classification providing a guarantee that the system can detect and respond to possible threats quickly. The quick adaptation of the CNN model for image and emotion detection is useful for applications that require an immediate response, such as distress detection in crowded areas or security monitoring in complicated surroundings.

As demonstrated by the findings, real-time emotion recognition with an IoT-based camera and a Convolutional Neural Network (CNN) model VGG-16 works successfully. The system's ability to sustain a high capture rate and fast processing time guarantees continuous and comprehensive environmental monitoring. This means it's ideal for uses like public safety, healthcare, and workplace safety where prompt detection and response to sensory signals are required. Balanced performance indicators including 0.65 recall, 0.64 accuracy, and 0.65 precision are used to demonstrate the model's validity and dependability in identifying emotions.

Considering the complexity of the task and the inherent complexities of the dataset (e.g., changes in facial expressions, occlusions, and low resolution), these results indicate that the model effectively learns and generalizes features relevant to emotion classification. The slight decrease in validation accuracy (0.6533 to 0.6414) indicates that the model maintains good generalization.

5. Conclusions

This research demonstrates that real-time emotion recognition with an IoT-based camera and a Convolutional Neural Network (CNN) model VGG-16 works successfully. A high capture rate and quick processing time are maintained by the system, ensuring ongoing and thorough monitoring.

This study recommends improving the system performance through further optimization. This can be achieved by changing hyperparameters, testing to another set of datasets aside from FER2013, and applying other preprocessing algorithms to improve recognition accuracy while minimizing false positives. Furthermore, since a balanced dataset affects model performance, it is important to consider a balanced dataset before training and testing. Considering data imbalances by training CNNs and estimating hidden features can be a useful tool to understand and manage the vulnerabilities of models [18]. The limited set of features of the other emotion classes of the FER2013 dataset as reflected in the dataset information somehow makes the CNN model more susceptible to errors on minority class data. This study also recommends the use of IoT-based cameras with higher specifications that perform effectively in an open setting. To enhance its applicability to real-world applications, the system can be tested in other critical areas where there is a higher possibility of an emergency crisis.

Acknowledgment

I would like to express my gratitude to the Commission on Higher Education (CHED) for their support, which was extremely helpful in conducting this research through the SIKAP program. My sincere gratitude is extended to the University of the Cordilleras and the Urdaneta City University for their steadfast assistance and resources. I am deeply thankful to my mentors, Dr. Nancy M. Flores, my adviser, and Dr. Thelma D. Palaoag, my professor, for their guidance and encouragement. My sincere gratitude is also extended to my family, friends and classmates, Richard Myrick T. Arellaga, Jann Vincent Paul C. Lagmay, Mc-Ornoc T. Lamao, and UC-DIT SIKAP Batch 2025 for their wise counsel and unwavering support.

References

1. L. Cao, “AI and data science for smart emergency, crisis and disaster resilience,” International Journal of Data Science and Analytics, vol. 15, no. 3, pp. 231–246, Mar. 2023, doi: 10.1007/s41060-023-00393-w. Available: https://doi.org/10.1007/s41060-023-00393-w
2. F. C. Moreno et al., “Fatigue as a key human factor in complex sociotechnical systems: Vessel Traffic Services,” Frontiers in Public Health, vol. 11, Apr. 2023, doi: 10.3389/fpubh.2023.1160971. Available: https://doi.org/10.3389/fpubh.2023.1160971
3. L. Li, Z. Ma, M. Bensi, and G. B. Baecher, “Social Media Crowdsourcing for Earthquake Damage Assessment,” Geo-Risk 2023, Jul. 2023, doi: 10.1061/9780784484975.018. Available: https://doi.org/10.1061/9780784484975.018
4. A. Wilke and R. Mata, “Cognitive bias,” in Elsevier eBooks, 2012, pp. 531–535. doi: 10.1016/b978-0-12-375000-6.00094-x. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/cognitive-bias
5. Q. Zhang, L. T. Yang, Z. Chen, and P. Li, “A survey on deep learning for big data,” Information Fusion, vol. 42, pp. 146–157, Jul. 2018, doi: 10.1016/j.inffus.2017.10.006. Available: https://doi.org/10.1016/j.inffus.2017.10.006
6. M. S. M. Suhaimin, M. H. A. Hijazi, E. G. Moung, P. nor E. Nohuddin, S. Chua, and F.

Coenen, “Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions,” Journal of King Saud University. Computer and Information Sciences/Mağalaẗ Ğam’aẗ Al-malīk Saud : Ùlm Al-ḥasib Wa Al-ma’lumat, vol. 35, no. 9, p. 101776, Oct. 2023, doi: 10.1016/j.jksuci.2023.101776. Available: https://doi.org/10.1016/j.jksuci.2023.101776

7. M. Avvenuti, S. Bellomo, S. Cresci, L. Nizzoli, and M. Tesconi, “Towards better social crisis data with HERMES: Hybrid sensing for EmeRgency ManagEment System,” Pervasive and Mobile Computing, vol. 67, p. 101225, Sep. 2020, doi: 10.1016/j.pmcj.2020.101225. Available: https://doi.org/10.1016/j.pmcj.2020.101225
8. V. Gude, S. Corns, and S. Long, “Flood prediction and uncertainty estimation using deep learning,” Water, vol. 12, no. 3, p. 884, Mar. 2020, doi: 10.3390/w12030884. Available: https://doi.org/10.3390/w12030884
9. Y. Li, H. Zhang, X. Xue, Y. Jiang, and Q. Shen, “Deep learning for remote sensing image classification: A survey,” Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery/Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, vol. 8, no. 6, May 2018, doi: 10.1002/widm.1264. Available: https://doi.org/10.1002/widm.1264
10. C. Liang and J. Dong, “A survey of deep learning-based facial expression recognition research,” Frontiers in Computing and Intelligent Systems, vol. 5, no. 2, pp. 56–60, Sep. 2023, doi: 10.54097/fcis.v5i2.12445. Available: https://doi.org/10.54097/fcis.v5i2.12445
11. H. A. Elfenbein and N. Ambady, “On the universality and cultural specificity of emotion recognition: A meta-analysis.,” Psychological Bulletin, vol. 128, no. 2, pp. 203–235, Jan. 2002, doi: 10.1037/0033-2909.128.2.203. Available: https://doi.org/10.1037/0033-2909.128.2.203
12. A. Adel and A. Farid, “Performance evaluation of machine learning for recognizing human facial emotions,” International Journal of Intelligent Information Technologies, vol. 17, no. 3, pp. 63–79, Jul. 2021, doi: 10.4018/ijiit.2021070105. Available: https://doi.org/10.4018/ijiit.2021070105
13. M. Franca, N. Bolognini, and M. Brysbaert, “Seeing emotions in the eyes: a validated test to study individual differences in the perception of basic emotions,” Cognitive Research, vol. 8, no. 1, Nov. 2023, doi: 10.1186/s41235-023-00521-x. Available: https://doi.org/10.1186/s41235-023-00521-x
14. A. I. Middya, B. Nag, and S. Roy, “Deep learning based multimodal emotion recognition using model-level fusion of audio–visual modalities,” Knowledge-based Systems, vol. 244, p. 108580, May 2022, doi: 10.1016/j.knosys.2022.108580. Available: https://doi.org/10.1016/j.knosys.2022.108580
15. OpenCV, “OpenCV - Open Computer Vision Library,” OpenCV, Jun. 01, 2024. Available: https://opencv.org/
16. K. Simonyan and A. Zisserman, “Very deep convolutional networks for Large-Scale image recognition,” arXiv (Cornell University), Jan. 2014, doi: 10.48550/arxiv.1409.1556. Available: https://arxiv.org/abs/1409.1556
17. A. Y. Nawaf and W. M. Jasim, “A pre-trained model vs dedicated convolution neural networks for emotion recognition,” International Journal of Power Electronics and Drive Systems/International Journal of Electrical and Computer Engineering, vol. 13, no. 1, p. 1123, Feb. 2023, doi: 10.11591/ijece.v13i1.pp1123-1133. Available: https://doi.org/10.11591/ijece.v13i1.pp1123-1133
18. D. Dablain, K. N. Jacobson, C. Bellinger, M. Roberts, and N. V. Chawla, “Understanding CNN fragility when learning with imbalanced data,” Machine Learning, Apr. 2023, doi: 10.1007/s10994-023-06326-9. Available: https://doi.org/10.1007/s10994-023-06326-9
19. S. Haghighi, M. Jasemi, S. Hessabi, and A. Zolanvari, “PyCM: Multiclass confusion matrix library in Python,” Journal of Open Source Software, vol. 3, no. 25, p. 729, May 2018, doi:

10.21105/joss.00729. Available: https://doi.org/10.21105/joss.00729
20. S. Raschka, “An Overview of general performance Metrics of Binary classifier Systems,” arXiv (Cornell University), Jan. 2014, doi: 10.48550/arxiv.1410.5330. Available: https://arxiv.org/abs/1410.5330
21. L. Yu and N. Zhou, “Survey of Imbalanced Data Methodologies,” arXiv (Cornell University), Jan. 2021, doi: 10.48550/arxiv.2104.02240. Available: https://arxiv.org/abs/2104.02240
22. K. M. Ting, “Confusion Matrix,” in Springer eBooks, 2011, p. 209. doi: 10.1007/978-0-387-30164-8_157. Available: https://doi.org/10.1007/978-0-387-30164-8_157
23. A. Luque, A. Carrasco, A. Martín, and A. De Las Heras, “The impact of class imbalance in classification performance metrics based on the binary confusion matrix,” Pattern Recognition, vol. 91, pp. 216–231, Jul. 2019, doi: 10.1016/j.patcog.2019.02.023. Available: https://doi.org/10.1016/j.patcog.2019.02.023
24. A. Tharwat, “Classification assessment methods,” Applied Computing and Informatics, vol. 17, no. 1, pp. 168–192, Jul. 2020, doi: 10.1016/j.aci.2018.08.003. Available: https://doi.org/10.1016/j.aci.2018.08.003
25. ShivamPandeyMSFT, “Results of machine learning models - Finance | Dynamics 365,” Microsoft Learn, Jun. 03, 2022. Available: https://learn.microsoft.com/en-us/dynamics365/finance/finance-insights/confusion-matrix
26. Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, “Learning Efficient Convolutional Networks through Network Slimming,” 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, doi: 10.1109/iccv.2017.298. Available: https://doi.org/10.1109/iccv.2017.298
27. J. Choi, S. Park, and J. Ko, “Analyzing Head-Mounted AR Device Energy Consumption on a Frame Rate Perspective,” https://ieeexplore.ieee.org/xpl/conhome/7963990/proceeding, Jun. 2017, doi: 10.1109/sahcn.2017.7964948. Available: https://doi.org/10.1109/sahcn.2017.7964948