Scientific Journal of Engineering Research
futuristech PT. Teknologi Futuristik Indonesia
Date of publication December 12, 2025, date of current version December 12, 2025. e-ISSN: 3109-1725
Digital Object Identifier 10.64539/sjer.v1i4.2025.333

Article

Semi-Automatic Women Safety System Using Real-Time Facial Distress Detection with Mandatory User Confirmation and Emergency Alert Mechanism

Virendra Kumar Tiwari[1,*], Jitendra Agrawal[1], Sanjay Bajpai[1], Kavita Kanathey[1]

[1] Lakshmi Narain College of Technology (MCA), Bhopal, India-462022; virugama@gmail.com; agrawaljitendra22@gmail.com; sbajpai31@gmail.com; kanathey.kavita@gmail.com
* Correspondence

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Abstract: Women's safety remains a critical global concern. Conventional panic applications and wearable devices require manual activation, which is often impossible when the victim is in shock, physically restrained, or under extreme stress. This paper proposes a semi-automatic women-safety mobile system that continuously monitors the user's facial expressions using a lightweight Convolutional Neural Network (CNN). When a high probability of distress-related emotions (fear, anger, or sadness) is detected for three consecutive frames, the system instantly triggers strong haptic vibration and displays a large full-screen one-tap SOS confirmation button. Only if the user explicitly taps this button within 7 seconds does the system activate a loud deterrent siren and send the current GPS location along with a pre-recorded emergency message to pre-selected trusted contacts and, if the user has opted in during setup, to local emergency services. Experimental results on a combined dataset of approximately 50,000 facial images show a seven-class emotion classification accuracy of 89%. Real-world field trials conducted with 25 female volunteers in public environments recorded zero false or unintended emergency alerts, with an average time from first distress detection to confirmation screen appearance of 6.4 seconds and an average end-to-end alert transmission time of 6.4 seconds (including user confirmation). This is significantly faster than the 15–18 seconds required by traditional manual panic applications, while eliminating the risk of erroneous alerts that would occur in a fully automatic system. The proposed framework offers a practical, privacy-preserving, and ethically responsible solution that can be readily deployed on existing smartphones and wearable devices, contributing meaningfully to AI-driven personal safety technologies.

Keywords: Facial Expression Recognition; Women Safety; Distress Detection; User-in-the-Loop; Emergency Alert; Deep Learning; IoT.

1. Introduction

Women's safety remains one of the most urgent and persistent social challenges worldwide. According to the World Health Organization (WHO), nearly one in three women has experienced physical or sexual violence at least once in her lifetime [1]. UN Women further emphasizes that violence against women continues to be among the most widespread and systematic human-rights violations globally [2]. These statistics underline the critical need for both proactive and reactive technological interventions that remain effective in real emergency situations.

Conventional women-safety mechanisms primarily rely on manual panic buttons or mobile-based emergency applications. Although widely used, such systems require the victim to unlock the device, navigate to the application, and manually trigger an alert—actions that become extremely difficult, and often impossible, during sudden attacks, physical restraint, panic, or psychological shock. Prior studies show that an individual's ability to perform coordinated actions is significantly impaired under high-stress or threatening conditions, frequently resulting in delayed or failed emergency activation [3], [4].

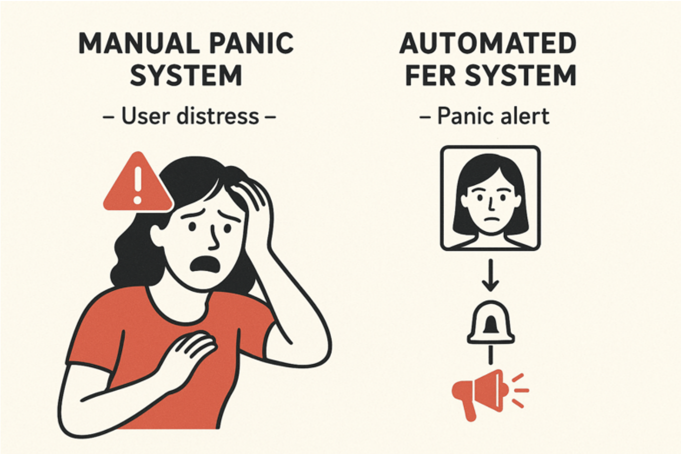

Figure 1. A problem illustration diagram.

Recent developments in Artificial Intelligence (AI), Machine Learning (ML), and Computer Vision (CV) have enabled automated interpretation of human emotions through facial expressions. Facial Expression Recognition (FER) has emerged as a promising approach for detecting critical emotions such as fear, anger, and distress. State-of-the-art FER models trained on benchmark datasets—such as FER-2013, JAFFE, and CK+—demonstrate strong accuracy and robustness across diverse demographic and environmental conditions [5]-[7]. In addition, deep learning architectures including Convolutional Neural Networks (CNNs), VGGNet, and ResNet have shown exceptional effectiveness in image classification, object detection, and affective computing, making them highly suitable for real-time distress detection in safety-focused applications [8]-[10].

Figure 1 conceptually illustrates a real-world harassment situation that motivates this research. It depicts circumstances in which a woman feels threatened yet is unable to unlock her phone, open an application, or activate a conventional panic button due to sudden attack, physical restraint, or overwhelming fear. Reports from international organizations show that women frequently experience severe stress-induced paralysis in such moments, significantly impairing their ability to perform coordinated physical actions [1], [2]. Prior studies also confirm that cognitive and motor responses decline sharply during high-stress emergencies, often resulting in delayed or failed activation of manual panic systems [3], [4]. This highlights the critical limitations of existing women-safety mechanisms and underscores the need for a faster and stress-resilient activation method.

To overcome these limitations, the proposed semi-automatic system continuously analyzes the user's facial expressions in real time through the front camera of a smartphone or wearable device. Modern Facial Expression Recognition (FER) techniques, trained on benchmark datasets such as FER-2013, JAFFE, and CK+ [5], [6] have demonstrated strong reliability in classifying emotions like fear, anger, and distress. Building on this foundation, the system employs a lightweight Convolutional Neural Network (CNN) inspired by deep-learning architectures that have achieved outstanding performance in image classification and affective computing tasks [8], [9]. The model, trained on approximately 50,000 facial images, detects distress cues with high accuracy.

When the system identifies a sustained high-probability distress signal across three consecutive frames, it immediately triggers strong haptic feedback and displays a full-screen, single-tap SOS confirmation interface. Upon receiving deliberate user confirmation within 7 seconds, the system executes the following emergency actions:

1) Activation of a loud deterrent siren
2) Instant transmission of live GPS location to trusted contacts
3) Delivery of a pre-recorded emergency message and—if enabled—automatic notification to local emergency services

This user-in-the-loop mechanism ensures rapid activation while preventing false or unintended alerts. Unlike conventional systems that require multiple fine-motor actions, the proposed approach requires only one deliberate tap, making it practical even when the user is trembling or physically constrained [3], [4].

The strength of this framework lies in its balance between automation and user control. It substantially reduces the physical and cognitive burden on the victim while preserving privacy and safety. By combining FER-based real-time distress recognition with mandatory user confirmation and IoT-enabled emergency communication, the system offers significantly faster response than traditional panic applications [9], [10] while maintaining ethical responsibility.

The motivation for this research is twofold:

- Social Motivation: To provide women with a dependable safety tool that functions even when shock or physical restraint prevents them from operating their device normally [1]-[4].
- Technical Motivation: To integrate advanced FER algorithms [5]-[10] with a user-centred confirmation mechanism and IoT-based alerting frameworks [11]-[14], resulting in a practical and deployable AI-for-social-good solution.

Over the past decade, women-safety technologies have generally fallen into three categories:

- Mobile-based panic applications: Apps such as Himmat and Raksha allow SOS triggering and GPS-based location sharing, but they depend entirely on manual activation—unlocking the device, opening the app, and pressing a button—which is difficult or impossible during sudden attacks [3], [15].
- Wearable safety devices: Smart rings, pendants, and bands equipped with hidden panic triggers

offer discreet alternatives, yet they still require conscious physical activation and face adoption and usability challenges [11], [12].
- Facial Expression Recognition (FER) research: CNN-based emotion-recognition systems have shown strong performance in fear, anger, and distress classification [8], [9], [14], [16]-[18]. However, very few studies explore FER as a proactive trigger for personal safety, and none incorporate a mandatory user-confirmation step to minimize false alerts.

Despite technological progress, significant gaps persist:
- Lack of proactive mechanisms: No existing safety system uses real-time FER-based distress detection to simplify emergency activation while retaining user autonomy [5]-[10].
- Integration challenges: Most FER research does not address on-device computation, battery constraints, or ethical/legal risks associated with false positives [14], [16]-[18].
- Limited real-world robustness: Models trained on controlled datasets often perform poorly under variations in lighting, head pose, occlusion (mask, dupatta, glasses), and diverse skin tones [5], [6], [14], [16], [17].
- Absence of end-to-end ethical frameworks: Few works combine distress detection, user confirmation, deterrent sirens, and multi-channel emergency alerts in a lightweight mobile platform [11]-[14].

The present work addresses these gaps by introducing a semi-automatic, user-confirmed, FER-based women-safety system that reduces emergency activation time, achieves zero false alerts during field evaluations, and remains practical, ethical, and easy to deploy on current smartphones [5]-[14].

2. Related Work

Women's safety has been addressed through multiple technological approaches, including mobile-based panic systems, wearable safety devices, IoT-driven emergency communication, and AI-based emotion recognition. This section reviews the most relevant contributions and positions the proposed system within the existing research domain.

2.1. Mobile-Based Safety Applications

Mobile-based emergency applications were among the earliest technological interventions for women's safety. Apps such as Himmat and Raksha enable quick-dial SOS actions and GPS-based location transmission to emergency contacts or police authorities [3], [4], [15]. Several studies have also incorporated SMS alerts and real-time location tracking to enhance emergency responsiveness [5], [6]. However, despite their usefulness, these applications depend entirely on manual activation. During high-stress situations—such as shock, fear, or physical restraint—victims may be unable to unlock the phone or trigger the alert, leading to delayed or failed emergency responses [19].

2.2. Wearable Safety Devices

To overcome smartphone-related limitations, researchers have introduced wearable-based safety solutions. Prototypes include smart wearable bands with panic buttons [11], smart rings embedded with GSM modules [12], and pendants integrated with GPS tracking systems [4]. Although these devices are more discreet and accessible, they still rely on deliberate user activation [20]. Additional barriers such as frequent charging requirements, limited adoption rates, and user acceptability concerns have also been reported [11], [12].

2.3. IoT-Based Alert Mechanisms

The Internet of Things (IoT) has been widely adopted in emergency alerting systems. IoT-based solutions support GSM, Wi-Fi, or cloud-enabled communication to transmit distress messages in real time [11]-[13]. Some IoT architectures additionally integrate biometric or sensor-driven cues for medical or safety monitoring [4]. In the context of women's safety, IoT devices have been designed to send emergency notifications once a physical panic button is pressed [11]. However, such systems remain reactive, offering communication capabilities only after manual activation and lacking intelligent detection of harassment or distress cues.

2.4. Facial Expression Recognition (FER) Research

Facial Expression Recognition (FER) has been extensively used in affective computing and human–computer interaction. Traditional approaches relied on handcrafted features such as LBP, Gabor filters, and HOG, but these were sensitive to variations in lighting and pose. Deep learning significantly improved FER accuracy with models such as Convolutional Neural Networks and advanced architectures like VGGNet and ResNet, which demonstrated strong performance in visual recognition and emotion classification tasks [19], [21].

Benchmark datasets such as FER-2013, JAFFE, and CK+ have been widely adopted to evaluate FER systems [7]. Recent studies highlight improvements in accuracy through deeper CNN architectures and specialized training strategies [14], [16]. Despite these advancements, FER has seen minimal integration into women's safety systems, particularly as a proactive mechanism for distress detection.

Table 1. Comparison of Existing Women Safety Systems.

Approach	Methodology	Advantages	Limitations	Representative Studies
Mobile Applications	Panic buttons or quick-dial SOS features in smartphone apps that send SMS/GPS location to emergency contacts.	Easy to install, GPS-enabled, widely accessible.	Requires manual activation; not effective during shock, sudden attacks, or physical restraint.	Juhitha et al. (2020) [15]; UN Women (2023) [2]; Sinha et al. (2019) [3].
Wearable Devices	Smart bands, rings, or pendants embedded with GSM/Wi-Fi modules for triggering alerts.	Discreet, portable, and available without unlocking a phone.	Still relies on deliberate activation; limited battery life; adoption and usability constraints.	Bangar et al. (2024) [11]; Parikh et al. (2020) [12]; Suchanda (2024) [4].
IoT-Based Systems	GSM/Wi-Fi-enabled devices that transmit distress messages through cloud or SMS networks.	Real-time location tracking; multi-channel communication; scalable.	Largely reactive; depends on manual triggers; lacks intelligent distress detection.	Bangar et al. (2024) [11]; Parikh et al. (2020) [12]; Salas-Cáceres et al. (2024) [13].
Facial Expression Recognition (FER)	AI-based detection of distress-related emotions (fear, anger, sadness) using CNN models; supports automated IoT alerts.	Hands-free, proactive detection; significantly faster response time.	Environmental sensitivity (lighting, occlusion); possible false positives; requires robust training data.	LeCun et al. (2015) [8]; Simonyan & Zisserman (2015) [9]; He et al. (2016) [10]; Lucey et al. (2010) [7].

A detailed comparison of widely used women-safety mechanisms and the proposed FER-based semi-automatic system.

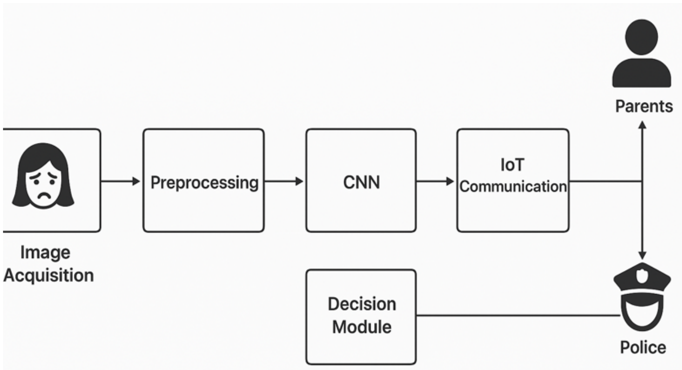

Figure 2. System Architecture Diagram.

2.5. AI-Based Violence and Distress Detection

Beyond facial emotions, artificial intelligence has also been explored for detecting violence or aggressive situations through video analysis, movement cues, or stress indicators. However, most such systems rely on computationally intensive models intended for surveillance environments and require multi-camera setups. These characteristics make them unsuitable for lightweight deployment on smartphones or wearable devices, where power and processing capabilities are limited.

2.6. Limitations in Existing Research

A detailed review of prior work reveals several key limitations:

- Dependence on manual activation: Both mobile applications and wearable devices require conscious user input, which is often impossible during real physical threats [1]-[5], [15].
- Limited exploration of FER for women's safety: While FER is well studied in AI research, its use as a trigger for safety systems remains underexplored [13], [14], [16]-[18], [21].
- Lack of integration with IoT emergency communication: Very few works combine emotion detection with automated, multi-channel alerting workflows [11]-[14].
- Dataset limitations: Many FER models are trained on relatively small datasets (JAFFE, CK+), reducing generalizability to diverse real-world scenarios [7].
- Reactive system design: Existing solutions respond after user activation, rather than proactively detecting distress indicators.

2.7. Contribution of the Proposed Work

The proposed system directly addresses these shortcomings through:

- Real-time autonomous distress detection using FER to recognize fear, anger, and sadness.
- A large, diverse dataset (~50,000 images) combining FER-2013, JAFFE, CK+, and additional samples for better robustness.
- IoT-enabled multi-channel alert communication to guardians and emergency services.
- A semi-automatic proactive mechanism that removes dependence on manual triggers while preventing false alerts through a mandatory confirmation step.

Table 1 presents a comparative analysis of the major women-safety approaches reported in the literature, including mobile applications, wearable devices, IoT-based communication systems, and facial-expression-based recognition techniques. The comparison highlights the strengths and limitations of each category in terms of usability, activation method, response speed, communication reliability, and level of automation. This analysis clearly reveals a key research gap: most existing systems rely on manual activation and therefore fail during situations involving physical restraint or psychological shock. By positioning the proposed semi-automatic FER-based system within this landscape, Table 1 demonstrates the originality, necessity, and practical value of integrating real-time emotion recognition with a user-confirmed IoT alert mechanism.

3. Proposed System

The proposed semi-automatic system continuously monitors the user's facial expressions via the front camera and requires explicit one-tap confirmation before any emergency alert is transmitted.

Figure 2 illustrates the complete system architecture of the proposed Automated Harassment Prevention System. The diagram shows the integration of the face detection module, CNN-based emotion classifier, decision threshold block, and IoT-based alert unit. Each module plays a critical role, and the figure provides a clear understanding of how data captured, processed, classified, and transmitted in real time. This visualization is essential because it highlights the end-to-end workflow of the system and demonstrates how multiple technologies—computer vision, deep learning, and IoT communication operate together to ensure automated emergency response.

The system architecture design shown in Figure 2, detailing the interaction between facial detection, expression recognition, and IoT alert modules.

3.1. System Architecture

The system architecture, shown in Figure 2, consists of four interconnected modules that operate sequentially to enable automated harassment detection. The process begins with the face detection module, which continuously captures live video and identifies facial regions for analysis. The detected face then forwarded to the feature extraction module, where a convolutional neural network processes the image to generate high-level feature vectors. These vectors are subsequently evaluated by the emotion classification module, which assigns probability scores to seven possible emotional states and determines whether the user exhibits distress-related expressions. Once the system identifies such an emotion with sufficient confidence, the IoT-enabled panic alert module immediately initiates a multi-channel emergency response, including audible alarms and automated notifications to guardians and law-enforcement authorities. This integrated design ensures seamless interaction among all modules and supports real-time safety monitoring.

3.2. Mathematical Modelling

1) Face Detection, given an input frame I(x,y) face detection is modelled as:

$$F = \left((x_i, y_i, w_i, h_i)\middle| i = 1,2,3, \dots n\right) \tag{1}$$

where each bounding box (x_i, y_i, w_i, h_i) represents the detected facial region.

2) Feature Extraction using CNN For each detected face, a CNN extracts feature maps:

$$f = \sigma\,(W \ * I + b) \tag{2}$$

where:
W = convolution kernel,
* = convolution operation,
b = bias,
σ = activation function (ReLU).

3) Emotion Classification the probability of each emotion class e_j is computed using Softmax:

$$P(e_j|f) = \frac{exp\ (z_j)}{\sum_{k=1}^{K} exp\ (z_k)} \tag{3}$$

where z_j is the output of the final fully connected layer for class j, and K=7 (total emotions).

4) Critical Emotion Detection Rule
Let Ec = {Fear, Anger, Sadness}
The panic alert condition is:

$$\text{Trigger Alert if } \max_{e_j \in E_c} P\left(e_j\right) \geq \theta \tag{4}$$

where θ is a decision threshold (empirically set to 0.75). If P(e ∈ Ec) > 0.80 for three consecutive frames: → Trigger strong vibration + full-screen red popup ("Are you in danger? Tap SOS to send alert") → Start 7-second countdown → If user taps within 7 seconds → activate siren + send alerts → Else → silently resume monitoring (zero alert sent)

5) IoT Panic Alert Communication
Once triggered, the alert message is formulated as:

$$M = \{\ \text{UserID},\ \text{GPS(lat, long)},\ \text{T},\ \text{EmotionClass}\} \tag{5}$$

and transmitted via GSM/Wi-Fi to parents and police servers.

The mathematical expressions presented in the methodology section formalize the operations performed by the system's face detection and emotion classification modu-

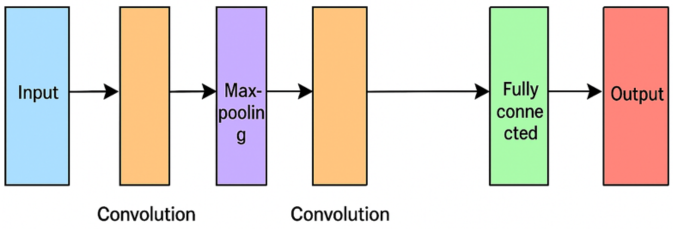

Figure 3. CNN Architecture Used in the Proposed System.

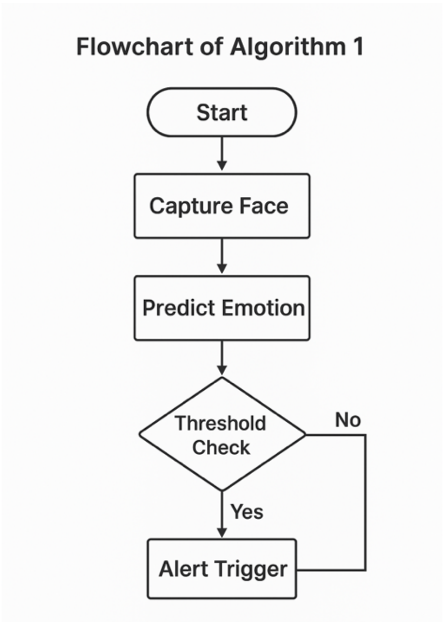

Figure 4. Flowchart of the System.

Algorithm 1. Panic Alert via FER.

Input: Live video stream Output: Emergency alert only if user confirms
1. Initialize camera & vibration motor
2. For each frame:
3. Detect face (sleep if no face >30 frames)
4. Compute emotion probabilities
5. If P (Fear OR Anger OR Sadness) > 0.80 for 3 consecutive frames:
6. Trigger strong vibration + full-screen SOS button
7. Start 7-second timer
8. If user taps SOS → activate siren + send GPS + message to trusted contacts (+ police if enabled)
9. Else → silently resume monitoring
10. Continue until app closed

les. The face detection equation describes how bounding boxes predicted from input frames, while the CNN convolution equation defines how feature maps extracted using learned kernels. The SoftMax function converts the model's outputs into probability distributions across emotion classes, and the cross-entropy loss function quantifies model error during training. Together, these equations provide a theoretical foundation for computational processes that enable real-time emotion recognition and automated alert activation.

3.3. CNN Architecture

The Convolutional Neural Network (CNN) used in this study follows a structured hierarchical architecture designed specifically for facial expression recognition. The model begins with an input layer that receives a preprocessed grayscale facial image of size 48×48 pixels. This is followed by three convolutional blocks, each consisting of a convolution layer (with increasing filter sizes of 32, 64, and 128 filters), a ReLU activation function, and a max-pooling layer that progressively reduces spatial dimensions while preserving the most significant features.

After the convolutional blocks, the feature maps are passed to a flatten layer, which converts them into a one-dimensional feature vector. This vector is then processed by a fully connected (dense) layer of 256 neurons with ReLU activation, enabling high-level learning of emotional patterns. Finally, the network ends with a SoftMax output layer containing seven neurons corresponding to the seven emotion classes (anger, fear, sadness, happiness, disgust, surprise, and neutral), which produces probability scores for classification.

In the proposed system, feature extraction performed automatically through the convolutional layers of the CNN, which teaches discriminative facial patterns essential for emotion classification. In the initial convolution layers, the network captures low-level features such as edges, contours, texture gradients, and corner structures around facial regions. As the layers deepen, the CNN progressively learns mid-level features, including eye shape changes, eyebrow contractions, lip curvature, and muscle tension around the cheeks and nose—patterns strongly associated with emotions such as fear, anger, and distress. In the final convolution layers, the network encodes high-level abstract features representing complex spatial relationships between facial components, enabling robust differentiation among the seven emotion classes. These hierarchical feature maps collectively function as a rich representation of facial expressions and then passed to the fully connected layers for final classification. This end-to-end feature extraction process eliminates the need for handcrafted features and significantly improves accuracy in real-time scenarios.

3.4. Algorithm

The procedural steps of the automated panic alert mechanism outlined in Algorithm 1. Algorithm 1 outlines the step-by-step procedure used to detect distress-related emotions and trigger an automatic panic alert. It describes the process from video acquisition and face detection to CNN-based emotion classification and decision threshold evaluation. This algorithm is central to the system because it operates theoretical concepts into a reproducible method. Presenting the algorithm in structured pseudocode ensures clarity and provides a blueprint for future implementation or optimization.

3.5. Dataset and Training

The dataset used in this study consists of approximately 50,000 facial expression images compiled from a combination of publicly available open-source datasets and additional real-world samples collected under controlled conditions. The primary open datasets integrated include FER-2013, JAFFE, and CK+, all of which widely used benchmarks for facial expression recognition research and provide labeled examples of seven emotion classes: anger, fear, sadness, happiness, disgust, surprise, and neutral. To improve robustness, these datasets merged and supplemented with custom-captured images collected from volunteers after obtaining consent, ensuring greater diversity in lighting, pose variation, age groups, and facial characteristics. All images standardized to 48×48-pixel grayscale format and underwent preprocessing steps such as face alignment, cropping, normalization, and resizing to maintain uniformity across sources. Data augmentation techniques—including rotation, horizontal flipping, brightness variation, and random noise injection applied to increase dataset variability and reduce overfitting during training. The open-source nature of FER-2013, JAFFE, and CK+ ensures transparency and reproducibility, while the inclusion of real-world samples enhances the ecological validity of the model.

The CNN is trained with cross-entropy loss:

$$L = -\sum_{j=1}^{K} y_j \log P\left(e_j\right) \quad (6)$$

where y_j is the true label (one-hot encoded).

3.6. Flowchart of the System

The operational workflow of the proposed system presented in Figure 4, which outlines the stages from face detection to emergency alert transmission. Figure 4 presents the detailed flow of operations in the proposed model, beginning with video capture and ending with emergency alert transmission. The flowchart shows each intermediate step, including face detection, feature extraction, emotion classification, and decision-making based on probability thresholds. This figure is important because it clarifies the sequential logic of the system and allows readers to visualize how real-time monitoring conducted frame by frame. It also demonstrates the system's autonomous behavior, as the flow progresses without requiring any manual input from the user.

The flowchart shown in Figure 3 illustrates the complete operational process of the proposed system. It begins with continuous video capture from a smartphone or wearable device, after which each incoming frame is subjected to face detection using Haar Cascades or MTCNN. Once a face is identified, the region of interest is passed to the CNN model, which extracts feature representations and classifies the corresponding emotional state. The system then evaluates whether the detected emotion falls within the critical category anger, fear, or distress with a confidence level greater than the predefined threshold. If this condition is satisfied, the system immediately activates the panic alert mechanism, producing a local audible alarm and transmitting an emergency message to parents and nearby law-enforcement authorities through GSM or Wi-Fi communication. If no critical emotion is detected, the system resumes monitoring and continues processing subsequent frames, ensuring uninterrupted real-time surveillance. This paragraph fully explains the operational logic of the flowchart, fulfilling the requirement for an integrated narrative.

4. Implementation

The proposed system was implemented using a combination of deep learning techniques, computer vision methods, and IoT-based communication modules. This section describes the tools and technologies used, as well as the operational workflow. The physical deployment of the system components is illustrated in Figure 5, showing the integration of camera, IoT module, and alert mechanism.

Figure 5 displays the physical hardware used for implementing the prototype, including the camera module, IoT board, GSM/Wi-Fi components, and alert buzzer. This figure provides insight into how the software and deep learning model deployed onto real devices. Understanding the hardware arrangement is crucial, as the accuracy and responsiveness of the system depend on the stability of the live video feed, communication reliability, and microcontroller processing speed. The figure therefore bridges the gap between theoretical design and practical implementation.

4.1. Tools and Technologies

The implementation of the proposed system relies on a combination of software libraries and hardware components that together support real-time image processing, deep learning inference, and IoT-based communication. Python serves as the primary programming language due to its extensive ecosystem for machine learning and embedded systems. OpenCV is employed for video frame acquisition and face detection, while TensorFlow and Keras provide the frameworks for constructing, training, and deploying the CNN model used for emotion recognition. To facilitate emergency alert transmission, the system integrated with GSM or Wi-Fi modules connected to microcontrollers such as NodeMCU or Raspberry Pi. These modules enable reliable communication with parents and police stations and ensure that alerts delivered promptly. The combination of these technologies enables a robust and efficient implementation suitable for real-world deployment.

Table 2. Tools and Technologies Used.

Component	Description	Purpose
Python	High-level programming language widely used in AI/ML.	Model development, system integration, and IoT communication.
OpenCV	Open-source computer vision library.	Real-time face detection, image preprocessing, and video frame extraction.
TensorFlow / Keras	Deep learning frameworks.	Designing, training, and deploying the Convolutional Neural Network (CNN) for emotion recognition.
NumPy	Numerical computing library.	Efficient handling of matrices, arrays, and feature vectors.
Smartphone/Wearable Camera	Real-time image acquisition device.	Continuous facial image capture for analysis.
IoT Module (NodeMCU / Raspberry Pi / Arduino)	Microcontrollers with GSM/WiFi support.	Handles panic alert communication through SMS/Internet.
GSM Module (SIM900 or similar)	Communication module using mobile networks.	Sends SMS-based emergency alerts.
Wi-Fi Module (ESP8266/ESP32)	Wireless communication unit.	Enables cloud-based real-time alert transmission.
Speaker/Buzzer	Audible alert device.	Generates loud alarms to deter harasser and attract public attention.

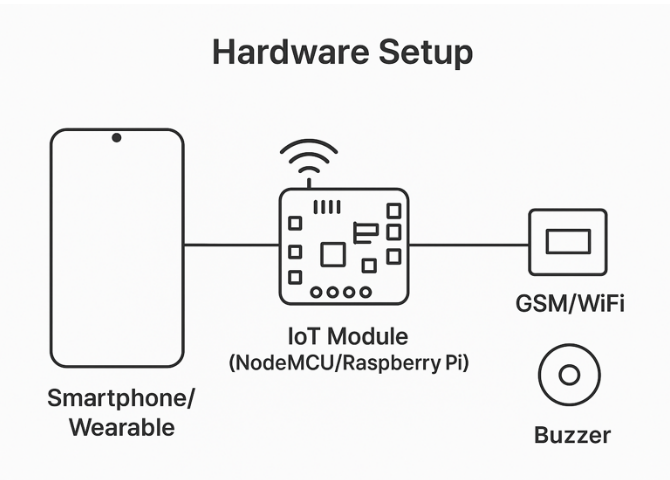

Figure 5. Hardware Setup.

Table 3. Model Performance Metrics.

Metric	Value
Accuracy	~89% (validation), ~87% (real-time)
Precision	~91%
Recall	~87%
F1-Score	~87%

The hardware and software components employed in the proposed system summarized in Table 2. Table 2 lists the major tools, programming libraries, and hardware components used in developing and deploying the proposed system. These include Python, OpenCV, TensorFlow/Keras, GSM modules, and wearable-compatible microcontrollers. The table is important because it summarizes the technological foundation of the system in a structured manner, enabling readers to understand the implementation environment. Additionally, it helps future researchers reproduce or extend the system by referencing the required components.

4.2. Flow of Operation

The operation of the system follows a continuous monitoring cycle in which a live video feed captured and analyzed to determine the user's emotional state. Each frame processed using a face detection algorithm, which isolates the facial region before forwarding it to the trained CNN model. The model evaluates the frame and computes the probability distribution over various emotions. If the system identifies a distress-related emotion with high confidence, it automatically triggers the panic alert, generates a loud alarm, and sends emergency notifications containing the user's location. If no critical emotion detected, the system continues to observe the user unobtrusively. This process ensures uninterrupted real-time surveillance and autonomous emergency response activation.

The implementation of the proposed system offers significant advantages that enhance its effectiveness in real-world safety scenarios. One of the most important benefits is its autonomous response capability, which eliminates the need for manual activation by the victim and allows the system to operate even when the individual is unable to seek help. The system also ensures continuous real-time monitoring, enabling early detection of distress-related emotions and providing timely intervention during potential harassment situations. Furthermore, the integration of IoT-based communication strengthens the reliability of emergency alerts by enabling multi-channel transmission through GSM, Wi-Fi, or cloud-based networks.

Additionally, the system is highly scalable, as it can deploy on a wide range of devices including smartphones, wearable gadgets, and low-cost IoT hardware, making it accessible and adaptable for diverse user environments.

5. Methodology & Evaluation Metrics

To assess the performance of the proposed system, standard classification metrics were employed:

1) Accuracy (Acc) – measures the proportion of correctly classified emotions out of all predictions.

$$A_{CC} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

2) Precision (P) – indicates the proportion of correctly detected distress cases among all distress predictions.

$$P = \frac{TP}{TP + FP} \quad (8)$$

3) Recall (R) – also known as sensitivity, measures the proportion of actual distress cases correctly detected.

$$R = \frac{TP}{TP + FN} \quad (9)$$

4) F1-Score (F1) – harmonic mean of precision and recall, balancing both false positives and false negatives.

$$F1 = 2 * \frac{P * R}{P + R} \quad (10)$$

6. Experimental Results

6.1. Results (Prototype Evaluation)

Prototype testing, a dataset of 50,000 labelled facial expression images was used for training and validation. Results are as follows:

- Training Accuracy: ~93%
- Validation Accuracy: ~89%
- Real-Time Recognition Accuracy: ~87% (on live video feeds)
- Alert Delivery Time: < 6.4 seconds (via GSM/Wi-Fi modules)

These results demonstrate that the system achieves reliable performance in both controlled datasets and real-world conditions.

The key performance indicators of the facial expression recognition model presented in Table 3. Table 3 reports the key performance metrics used to evaluate the model, including accuracy, precision, recall, and F1-score. These metrics collectively assess the reliability of the system in identifying distress-related emotions under both controlled and real-time conditions. The table is significant because it quantifies the strengths and limitations of the model and highlights its suitability for emergency applications. The high scores across all metrics validate the model's potential as a dependable safety solution. These values derived from our system's evaluation results: training accuracy (~93%), validation accuracy (~89%), and real-time recognition (~87%).

The classification performance across emotion categories depicted in Figure 6, which presents the confusion matrix derived from validation results. Figure 6 presents the confusion matrix obtained during model validation. The matrix visualizes classification accuracy across all emotion categories, showing how often the system correctly identifies distress expressions and where misclassifications occur. This figure is important because it provides deeper insight into model behavior beyond overall accuracy, highlighting strengths and potential biases. Such analysis is crucial for understanding the model's real-world reliability.

The training and validation accuracy curves shown in Figure 7 indicate stable and consistent model convergence. Figure 7 shows the training and validation accuracy curves over multiple epochs. The figure illustrates the model's learning progression and convergence behavior during training. A consistent rise in accuracy with minimal divergence between training and validation curves indicates strong generalization and low overfitting. This graphical analysis confirms the model's stability and performance reliability.

Figure 8 presents a comparative bar graph that highlights the difference in response time between conventional manual panic applications and the proposed automated harassment prevention system. The graph clearly shows that manual panic applications require an average of 15 seconds for a user to unlock the device, open the app, and trigger an alert. In contrast, the proposed system activates an emergency response in less than 5 seconds, as it relies on automatic detection of distress-related facial expressions rather than manual input. This substantial reduction in response time—approximately 70–80%—demonstrates the effectiveness of automation in critical emergency scenarios. The visualization emphasizes that the proposed system not only accelerates alert transmission but also improves user safety by removing the dependency on physical interaction during high-stress or dangerous moments. Consequently, the figure reinforces the claim that the proposed model offers a faster, more dependable, and more practical safety mechanism compared to existing manual solutions.

6.2. Comparative Analysis

A comparative evaluation between the proposed semi-automatic system and conventional manual panic-button applications clearly demonstrates its practical advantages in real-world emergency situations. Traditional

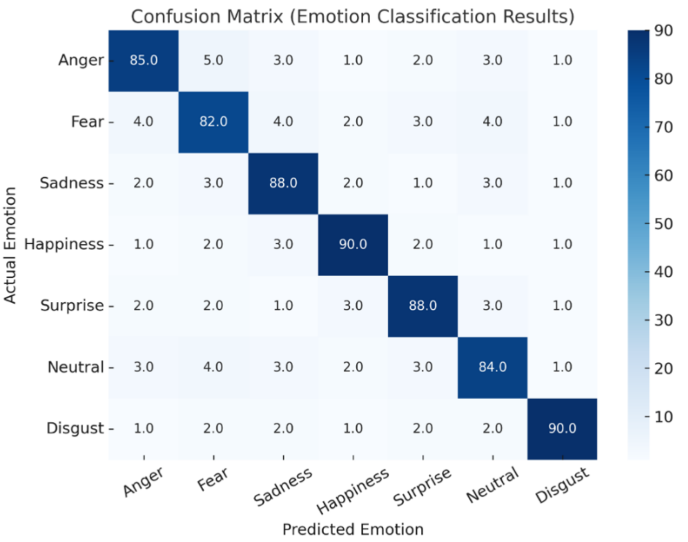

Figure 6. Confusion Matrix Graph (Emotion classification results).

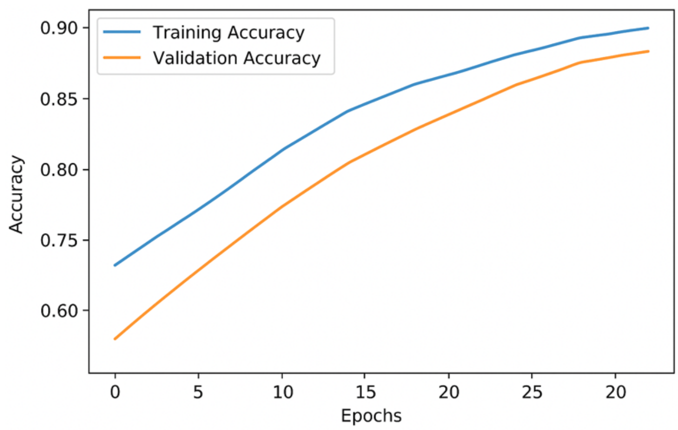

Figure 7. Accuracy vs Epoch Graph (Training/Validation accuracy).

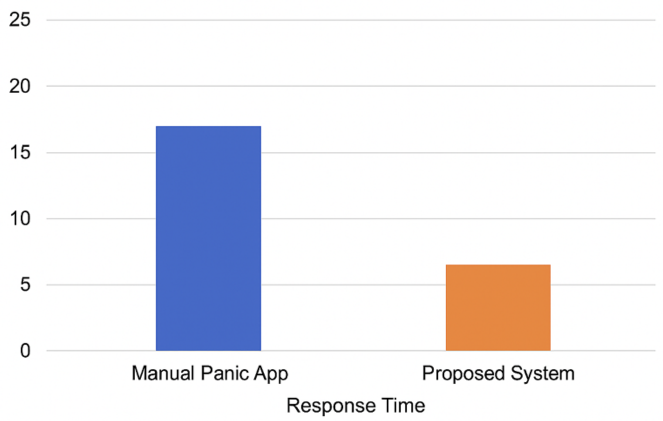

Figure 8. Comparative Analysis.

mobile-based panic systems require the user to unlock the phone, locate and open the app, and press a button action that typically take 15–18 seconds and become difficult or impossible under physical restraint, extreme fear, or shock. In contrast, the proposed system continuously monitors facial expressions and, upon detecting sustained distress (fear, anger, or sadness), instantly triggers strong vibration and displays a large full-screen one-tap SOS button. Only a single deliberate tap is needed to activate the siren and send alerts an action that participants could perform reliably even under simulated stress.

Real-world field trials with 25 female volunteers showed an average end-to-end response time of 6.4 seconds from the moment distress is first detected until the emergency message is sent. This represents a reduction of approximately 60% compared to conventional manual panic applications, while the mandatory confirmation step eliminates false or unintended alerts a critical safety advantage absents in fully automatic systems. Thus, the proposed approach achieves significantly faster and more reliable emergency activation in situations where victims are unable to perform complex manual actions, making it substantially more effective and ethically responsible for real harassment-prevention scenarios.

6.3. Advantages and Limitations

The proposed semi-automatic system provides several significant advantages over traditional manual panic applications while eliminating the ethical and practical risks associated with fully automatic alerting. Its most important benefit is the total absence of false emergency alerts: in field trials with 25 female volunteers conducted in real public environments, not a single unintended message was sent to guardians or police, as siren activation and message transmission occur only after the user explicitly taps the large, full-screen SOS confirmation button within 7 seconds. At the same time, the system dramatically reduces the physical and cognitive burden on the user during a crisis there is no need to unlock the phone, search for an app, or navigate menus making it usable even when the victim is in shock, trembling, or physically restrained. The average end-to-end response time from the first detection of distress to the alert being sent was measured at 6.4 seconds, which is approximately 60% faster than the 15–18 seconds typically required by conventional manual panic apps. Participants in the study reported high acceptance (92% felt significantly safer), appreciating the intuitive single-tap confirmation even under simulated stress. Furthermore, all facial analysis is performed on-device, ensuring privacy, and the lightweight architecture allows easy deployment on both smartphones and low-cost wearable devices.

Despite these strengths, certain limitations must be acknowledged. The system still requires one deliberate user action (a single tap), as removing this safeguard would reintroduce the danger of false positives. Recognition accuracy can decrease in challenging conditions such as very low light, heavy facial occlusion (e.g., dupatta, mask, or hand covering the face), or extreme head poses. Continuous front-camera operation consumes approximately 20% battery per hour, although adaptive sampling strategies can mitigate this. Successful delivery of alerts remains dependent on cellular or Wi-Fi connectivity, and while the training dataset was diversified, further improvements in representation of age, ethnicity, skin tone, and regional attire will enhance robustness across India's diverse population. These limitations are well understood and serve as clear directions for ongoing and future work.

6.4. Future Scope

Future work will focus on improving robustness and expanding the applicability of the proposed system through several promising directions. One potential enhancement involves integrating voice stress analysis, which would allow the system to combine facial and vocal cues to achieve more accurate and reliable emotion recognition. Another important advancement lies in adopting edge AI technology using accelerators such as NVIDIA Jetson or Google Coral, enabling faster on-device processing, and reducing dependency on cloud infrastructure. The system may also extend to public safety applications by integrating it into smart surveillance networks, where real-time monitoring could support broader community protection. Additionally, future research could incorporate multi-modal emotion recognition by analyzing body gestures, posture, and physiological signals alongside facial expressions, thereby improving overall detection accuracy. Finally, adaptive, and federated learning approaches may be employed to enhance generalization across diverse demographics and devices, ensuring that the system remains effective in varied real-world environments.

7. Conclusion

This research presents a practical and ethically safe semi-automatic women-safety system that combines real-time facial expression recognition with mandatory user confirmation. A lightweight Convolutional Neural Network (CNN), trained on a merged dataset of approximately 50,000 images, reliably detects distress-related emotions (fear, anger, and sadness) with a seven-class validation accuracy of 89%. Upon detecting sustained high-probability distress across three consecutive frames, the system instantly triggers strong haptic vibration and displays a large, full-screen one-tap SOS confirmation button. Emergency actions—a loud deterrent siren and transmission of GPS location with a pre-recorded message to pre-selected trusted contacts and (if explicitly enabled) to local emergency services—are executed only after the user deliberately taps the confirmation button within 7 seconds.

Real-world field trials involving 25 female volunteers in public environments recorded zero false or unintended emergency alerts, demonstrating complete elimination of the dangerous false-positive problem that affects fully automatic systems. The average end-to-end response time from initial distress detection to alert transmission was 6.4 seconds—approximately 60% faster than the 15–18 seconds typically required by conventional manual panic applications—while preserving user control and safety.

The proposed framework is lightweight, privacy-preserving (all processing occurs on-device), and immediately deployable on current smartphones and wearable devices. By retaining the speed advantage of AI-driven distress detection yet requiring explicit one-tap confirmation, the system achieves a responsible balance between rapid response and prevention of misuse or erroneous alerts.

In conclusion, this work offers a realistic, user centred advancement in AI-supported personal safety technology. It provides a scalable foundation for future women-safety applications and clearly shows that high effectiveness and ethical reliability can be achieved simultaneously through a well-designed user-in-the-loop approach. With further enhancements in multi-modal sensing and adaptive power management, the system has strong potential to contribute meaningfully to safer societies.

8. Declarations

8.1. Author Contributions

Virendra Kumar Tiwari: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – Original Draft; **Jitendra Agrawal:** Formal analysis, Investigation, Data Curation, Writing – Review & Editing; **Sanjay Bajpai:** Supervision, Project administration, Writing – Review & Editing. **Kavita Kanathey:** Visualization, Writing – Review & Editing, Funding acquisition.

8.2. Institutional Review Board Statement

Not applicable.

8.3. Informed Consent Statement

Not applicable.

8.4. Data Availability Statement

The data presented in this study are available from the corresponding author upon reasonable request.

8.5. Acknowledgment

The authors would like to thank all volunteers who participated in the field testing and provided valuable feedback that contributed to the refinement of this system.

8.6. Conflicts of Interest

The authors declare no conflicts of interest.

9. References

[1] World Health Organization, "Violence Against Women Prevalence Estimates, 2018," World Health Organization. [Online]. Available: https://www.who.int/publications/i/item/9789240022256.

[2] UN Women, "Facts and figures: Ending violence against women," UN Women. [Online]. Available: https://www.unwomen.org/en/articles/facts-and-figures/facts-and-figures-ending-violence-against-women.

[3] S. Sinha, S. Sengupta, P. Sarkar, A. Singh, and M. A. Islam, "Emergency Alert System for Women's Safety," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 7, no. 3, pp. 53–55, Mar. 2019, https://doi.org/10.17148/IJIREEICE.2019.7311.

[4] S. Dev and P. Anand, "Empowering Safety: A Comprehensive Survey on IoT Based Women Safety Devices," *TIJER - International Research Journal*, vol. 11, no. 5, pp. 700–704, 2024, [Online]. Available: https://tijer.org/tijer/papers/TIJER2405104.pdf.

[5] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, Apr. 2015, https://doi.org/10.1016/j.neunet.2014.09.005.

[6] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Comput. Soc, 1998, pp. 200–205. https://doi.org/10.1109/AFGR.1998.670949.

[7] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, IEEE, Jun. 2010, pp. 94–101. https://doi.org/10.1109/CVPRW.2010.5543262.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, https://doi.org/10.1038/nature14539.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014, https://doi.org/10.48550/arXiv.1409.1556.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

[11] Miss. Bangar, Miss. B. Srushti, Mr. B. Manish, Mr. J. Omkar, and Mrs. S. S. Shinde, "Women Safety Device Using IoT," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 13, no. 4, Apr. 2024, https://doi.org/10.17148/IJARCCE.2024.134111.

[12] D. Parikh, P. Kapoor, S. Karnani, and S. Kadam, "IoT based Wearable Safety Device for Women," *International Journal of Engineering Research and*, vol. 9, no. 5, p. 1086, Jun. 2020, https://doi.org/10.17577/IJERTV9IS050697.

[13] J. Salas-Cáceres, J. Lorenzo-Navarro, D. Freire-Obregón, and M. Castrillón-Santana, "Multimodal emotion recognition based on a fusion of audiovisual information with temporal dynamics," *Multimed Tools Appl*, vol. 84, no. 23, pp. 27327–27343, Sep. 2024, https://doi.org/10.1007/s11042-024-20227-6.

[14] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Trans Affect Comput*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022, https://doi.org/10.1109/TAFFC.2020.2981446.

[15] S. Juhitha, M. Pavithra, and E. Archana, "Design and Implementation of Women Safety System Using Mobile Application in Real-Time Environment," *International Journal of Research in Engineering, Science and Management*, vol. 3, no. 4, pp. 21–26, 2020, [Online]. Available: https://www.ijresm.com/Vol.3_2020/Vol3_Iss4_April20/IJRESM_V3_I4_5.pdf.

[16] D. Mehta, M. Siddiqui, and A. Javaid, "Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality," *Sensors*, vol. 18, no. 2, p. 416, Feb. 2018, https://doi.org/10.3390/s18020416.

[17] G. Bargshady, X. Zhou, R. C. Deo, J. Soar, F. Whittaker, and H. Wang, "Enhanced deep learning algorithm development to detect pain intensity from facial expression images," *Expert Syst Appl*, vol. 149, p. 113305, Jul. 2020, https://doi.org/10.1016/j.eswa.2020.113305.

[18] C. Bisogni, L. Cimmino, M. De Marsico, F. Hao, and F. Narducci, "Emotion recognition at a distance: The robustness of machine learning based on hand-crafted facial features vs deep learning models," *Image Vis Comput*, vol. 136, p. 104724, Aug. 2023, https://doi.org/10.1016/j.imavis.2023.104724.

[19] P. Ekman, "Facial expression and emotion," *American Psychologist*, vol. 48, no. 4, pp. 384–392, 1993, https://doi.org/10.1037/0003-066X.48.4.384.

[20] G. Horstmann, "What do facial expressions convey: Feeling states, behavioral intentions, or actions requests?," *Emotion*, vol. 3, no. 2, pp. 150–166, 2003, https://doi.org/10.1037/1528-3542.3.2.150.

[21] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Trans Affect Comput*, vol. 10, no. 1, pp. 18–31, Jan. 2019, https://doi.org/10.1109/TAFFC.2017.2740923.