



Facial Emotion Recognition Using Lightweight Convolutional Neural Network work for Real-Time Applications

Ms. V . Bhavya
Assistant Professor
Vidya Jyothi Institute of
Technology
Hyderabad
bhavyavcseds@vjit.ac.in

M. Sri Harshini
UG Student
Vidya Jyothi Institute of
Technology
Hyderabad
harshinis732@gmail.com

N. Chandana
UG Student
Vidya Jyothi Institute of
Technology
Hyderabad
nchandanareddy@gamil.com

S. Rikitha
UG Student
Vidya Jyothi Institute of
Technology
Hyderabad
rikithareddysingidhi06@gmail.com

How to Cite this Article:

Harshini, M. S., Chandana, N. & Rikitha, S. (2026). Facial Emotion Recognition Using Lightweight Convolutional Neural Network work for Real-Time Applications. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04).
<https://doi.org/10.55041/ijcope.v2i4.925>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i4.925>

Abstract: Facial Emotion Recognition (FER) has emerged as a significant research area in artificial intelligence and affective computing due to its wide range of real-world applications. Human emotions expressed through facial features provide valuable insights into psychological and behavioural states. This paper presents a lightweight Convolutional Neural Network (CNN)-based facial emotion recognition system designed for real-time applications. The proposed model is trained using the FER-2013 dataset, which consists of grayscale facial images categorized into seven emotional classes: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. Unlike deep and computationally intensive architectures, this work focuses on developing a shallow and efficient CNN model that can operate on standard hardware without requiring GPUs. Haar Cascade classifiers are employed for real-time face detection, and OpenCV is integrated to capture and process live webcam input. Each detected face is preprocessed and resized to meet the input requirements of the CNN model before emotion classification. The system demonstrates reliable performance under normal lighting conditions while maintaining low computational cost. Experimental observations indicate satisfactory accuracy with smooth real-time execution. The proposed solution is beginner-friendly, easily deployable, and suitable for academic and low-resource environments. This work proves that effective facial emotion recognition can be achieved using optimized and accessible deep learning techniques without complex infrastructure.

Keywords: Facial Emotion Recognition, Convolutional Neural Network, Deep Learning, FER-2013 Dataset, OpenCV, Real-Time Detection



1. Introduction

Facial expressions are one of the most natural and powerful forms of non-verbal communication. Humans instinctively interpret emotions such as happiness, sadness, anger, fear, and surprise through facial cues. The ability to automatically recognize these emotions using computational systems has gained significant importance in recent years. Facial Emotion Recognition (FER) plays a crucial role in bridging the gap between humans and machines by enabling emotion-aware systems that can respond intelligently to human behaviour.

With the rapid advancement of Artificial Intelligence (AI) and computer vision, emotion recognition has become an active research domain. Early approaches to FER relied on handcrafted feature extraction techniques such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gabor filters. These methods required manual selection of features and often struggled under varying lighting conditions, facial orientations, and occlusions. As a result, their performance was limited in real-world environments.

The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), revolutionized image-based classification tasks. CNNs automatically learn hierarchical features directly from raw image data, eliminating the need for manual feature engineering. Deep architectures such as VGG Net, Res Net, and Mobile Net have demonstrated high accuracy in facial emotion recognition tasks. However, these models are computationally intensive and typically require powerful GPUs for real-time implementation, making them less accessible for low-resource environments.

In many practical scenarios, such as classrooms, healthcare monitoring systems, and small-scale research laboratories, access to high-end computational resources is limited. Therefore, there is a need for lightweight and efficient FER systems that can deliver reliable performance without demanding advanced hardware. Designing a shallow yet effective CNN architecture can significantly reduce computational complexity while maintaining acceptable accuracy levels.

This research focuses on developing a real-time facial emotion recognition system using a

lightweight CNN model trained on the FER-2013 dataset. The dataset contains grayscale facial images categorized into seven emotional classes, providing a standardized benchmark for emotion recognition tasks. By leveraging Haar Cascade classifiers for face detection and OpenCV for video capture, the proposed system processes live webcam input and performs real-time emotion prediction.

The primary objective of this work is to strike a balance between accuracy, computational efficiency, and ease of implementation. The system is designed to operate smoothly on standard laptops or desktops with minimal memory requirements. Unlike complex deep learning models, the proposed approach prioritizes accessibility and simplicity, making it suitable for students, beginners, and researchers working with limited resources.

Furthermore, emotion-aware systems have broad applications in human-computer interaction, intelligent tutoring systems, mental health monitoring, surveillance, and customer experience analysis. By enabling machines to interpret emotional states, FER systems enhance user engagement and improve decision-making processes in various domains.

In summary, this paper presents a practical and efficient solution for real-time facial emotion recognition using a lightweight CNN architecture. The proposed approach demonstrates that effective emotion detection can be achieved without relying on deep and computationally expensive models, thereby contributing to accessible and scalable AI-based emotion recognition systems.

2. Literature Survey

Facial Emotion Recognition (FER) has evolved significantly over the past two decades with advancements in computer vision and machine learning. Early research in this domain focused on feature-based methods where facial characteristics were manually extracted using techniques such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gabor wavelets. These handcrafted features were then classified using traditional machine learning algorithms like Support Vector Machines (SVM) and k-Nearest Neighbours (kNN). Although these approaches performed reasonably well in controlled environments, their robustness was limited when exposed to varying



lighting conditions, head movements, and occlusions.

With the introduction of deep learning, Convolutional Neural Networks (CNNs) became the dominant approach for image-based classification problems, including facial emotion recognition. Mollahosseini et al. (2016) demonstrated that deeper CNN architectures significantly improved recognition accuracy compared to traditional methods. Their work highlighted the ability of CNNs to automatically learn hierarchical features directly from pixel-level data, eliminating the need for manual feature engineering.

Levi and Hassner (2015) explored emotion recognition “in the wild” using CNNs combined with mapped binary patterns. Their research emphasized the challenges of real-world datasets where faces are captured under uncontrolled conditions. The study showed that deep learning models could generalize better than classical approaches, though accuracy still varied across emotion classes.

Li and Deng (2022) provided a comprehensive survey of deep facial expression recognition techniques and concluded that while deep CNN architectures such as VGGNet and ResNet achieve high accuracy, they require significant computational resources. These models often depend on GPU acceleration, making them less practical for low-resource or real-time applications.

To overcome hardware constraints, researchers proposed lightweight architectures optimized for mobile and edge devices. MobileNet-based and shallow CNN models were introduced to reduce parameter size and computational complexity. Mehta and Joshi (2021) presented a lightweight CNN designed for edge devices, demonstrating that acceptable accuracy can be achieved with fewer layers and reduced training time. However, slight trade-offs in classification precision were observed compared to deeper networks.

Real-time implementation of FER systems has also been widely studied. Several works integrated OpenCV with CNN models for webcam-based emotion detection. These systems commonly used Haar Cascade classifiers for fast face detection followed by CNN-based classification. While effective, many implementations faced issues with lighting variations and subtle emotion

differentiation, particularly between emotions such as fear and surprise.

Overall, the literature indicates that although deep learning has significantly improved emotion recognition accuracy, challenges remain in balancing computational efficiency and performance. There is a growing need for lightweight, accessible, and real-time FER systems that can operate effectively on standard hardware without sacrificing reliability. The present work builds upon these findings by proposing a shallow CNN architecture integrated with OpenCV to achieve practical real-time facial emotion recognition.

3. Objectives

The primary objective of this research is to design and implement a real-time Facial Emotion Recognition (FER) system using a lightweight Convolutional Neural Network (CNN). The system aims to accurately identify human emotions from facial expressions while ensuring low computational complexity. Unlike deep and resource-intensive models, the proposed approach focuses on developing an efficient architecture that can operate smoothly on standard computing devices without requiring high-end GPUs or specialized hardware.

Another key objective is to train and evaluate the model using the FER-2013 dataset, which contains labeled grayscale images representing seven emotion categories: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The study also aims to integrate Haar Cascade classifiers for real-time face detection and OpenCV for live video capture. By combining these technologies, the system is designed to process webcam input, detect faces, and classify emotions instantly with minimal delay.

Furthermore, this research seeks to create a beginner-friendly and easily deployable emotion recognition system suitable for students, researchers, and low-resource environments. The objective includes analyzing system performance under real-time conditions, maintaining a balance between accuracy and efficiency, and identifying potential areas for improvement. Overall, the study aims to contribute a practical, accessible, and scalable solution to the field of facial emotion recognition.



3.1 Problem Statement

Facial Emotion Recognition (FER) systems have gained significant attention in recent years due to their applications in healthcare, education, surveillance, and human-computer interaction. Although deep learning models such as VGG Net and Res Net achieve high accuracy in emotion classification, they are computationally intensive and require powerful GPUs for real-time execution. This makes them unsuitable for deployment in low-resource environments such as academic institutions, small-scale research labs, and standard personal computing systems.

Traditional machine learning approaches, on the other hand, rely on handcrafted feature extraction methods such as Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). These techniques often struggle with variations in lighting, facial orientation, occlusions, and real-world environmental conditions, leading to reduced accuracy and poor generalization. Furthermore, many existing systems either prioritize accuracy at the expense of computational efficiency or focus on lightweight models that compromise classification reliability.

Therefore, there is a need to develop a facial emotion recognition system that balances accuracy, computational efficiency, and real-time performance. The challenge lies in designing a lightweight CNN-based architecture capable of operating smoothly on standard hardware while maintaining reliable emotion classification across diverse conditions. This research addresses this problem by proposing an efficient and accessible real-time FER system that minimizes hardware dependency without sacrificing performance.

3.2 Novelty Aspects

The primary novelty of this work lies in the design and implementation of a lightweight Convolutional Neural Network (CNN) architecture specifically optimized for real-time facial emotion recognition on standard hardware. Unlike many existing approaches that rely on deep and computationally expensive models such as VGGNet or ResNet, the proposed system focuses on achieving a balance between accuracy and computational efficiency. This makes the solution accessible to users who do

not have access to high-end GPUs or specialized processing units.

Another novel contribution of this study is the integration of a shallow CNN model with Haar Cascade-based face detection and OpenCV for seamless real-time webcam processing. While several studies address either high-accuracy offline emotion classification or lightweight mobile deployment, fewer works emphasize simplicity, beginner-friendliness, and practical implementation in academic environments. The proposed framework demonstrates that reliable emotion detection can be achieved using a simplified architecture without compromising usability.

Additionally, the system is designed with minimal hardware requirements and ease of deployment in mind. The implementation is fully software-based and can operate efficiently on devices with basic configurations such as 4GB RAM and a standard CPU. By prioritizing accessibility, real-time capability, and low computational cost, this work contributes a scalable and practical alternative to resource-intensive FER models, making emotion-aware technology more approachable for students and researchers.

3.3 Scope

The scope of this research is focused on the design and implementation of a real-time Facial Emotion Recognition (FER) system using a lightweight Convolutional Neural Network (CNN). The system is developed to classify facial expressions into seven basic emotion categories—Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise—using the FER-2013 dataset. The study primarily concentrates on achieving a balance between computational efficiency and classification performance while operating on standard hardware without GPU support.

This work covers face detection using Haar Cascade classifiers and real-time video processing through OpenCV integration. The implementation is limited to grayscale facial image analysis and webcam-based input. The system is evaluated under normal lighting conditions and standard environmental settings. The focus remains on developing a beginner-friendly, low-resource model suitable for academic purposes, research prototypes, and small-scale applications.



However, the study does not extend to advanced deep architectures, multi-modal emotion recognition (such as speech or text-based emotion analysis), or large-scale commercial deployment. Factors such as extreme lighting variations, heavy occlusions, or highly dynamic facial movements are beyond the current implementation scope. Future research may expand the system by incorporating transfer learning, mobile deployment, or multi-modal emotion analysis for broader real-world applications.

4. Proposed Methodology

The proposed Facial Emotion Recognition (FER) system is designed to perform real-time emotion detection using a lightweight Convolutional Neural Network (CNN) architecture. The methodology begins with training the model on the FER-2013 dataset, which consists of 48×48 grayscale facial images categorized into seven emotion classes: Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise. The dataset is preprocessed by normalizing pixel values to improve training stability and convergence. The images are reshaped to match the input dimensions required by the CNN model.

For real-time execution, face detection is carried out using Haar Cascade classifiers. Each frame captured from the webcam through OpenCV is converted to grayscale to reduce computational complexity. The Haar Cascade algorithm scans the frame and detects facial regions using predefined Haar-like features. Once a face is identified, a Region of Interest (ROI) is extracted and resized to 48×48 pixels to match the model's input specifications.

The CNN architecture used in this work is intentionally shallow to ensure computational efficiency. It consists of multiple convolutional layers for feature extraction, followed by ReLU activation functions to introduce non-linearity. Max-pooling layers are applied to reduce spatial dimensions and computational load, while dropout layers are included to prevent overfitting during training. The final fully connected dense layer maps the extracted features to emotion categories, and a Soft max layer produces probability scores for each class.

During real-time prediction, the preprocessed facial image is passed to the trained model, which outputs probability values for all emotion classes. The emotion with the highest probability score is selected

as the final prediction and displayed above the detected face. This pipeline ensures smooth real-time performance while maintaining reliable classification accuracy. The overall methodology emphasizes simplicity, efficiency, and accessibility, enabling deployment on standard systems without requiring advanced hardware support.

4.1 Framework

The framework of the proposed Facial Emotion Recognition (FER) system is designed to perform real-time emotion detection using a structured processing pipeline. The system consists of five major components: video acquisition, face detection, preprocessing, emotion classification using CNN, and output display. These components work sequentially to ensure smooth and efficient real-time execution.

The first stage of the framework involves capturing live video input through a webcam using OpenCV. The captured video is divided into individual frames, which are processed continuously. Each frame is converted into grayscale format to reduce computational complexity and improve face detection performance.

In the second stage, Haar Cascade classifiers are applied to detect facial regions within each frame. The classifier scans the grayscale image using predefined Haar-like features and identifies the coordinates of detected faces. Once a face is located, a bounding box is drawn, and the Region of Interest (ROI) is extracted for further processing.

The third stage involves preprocessing the extracted facial region. The ROI is resized to 48×48 pixels to match the input size required by the CNN model. Pixel values are normalized to a range of 0 to 1 to ensure stable predictions. The processed image is reshaped into a format suitable for model inference.

The fourth stage is emotion classification using the trained lightweight CNN model. The model extracts features through convolutional and pooling layers and classifies the facial expression using a fully connected layer with a Soft max activation function. The output is a probability distribution across seven emotion classes, and the highest probability determines the predicted emotion.

Finally, in the output stage, the predicted emotion label is displayed above the detected face in real



time. This entire pipeline operates continuously until the user exits the application. The framework emphasizes efficiency, simplicity, and low hardware requirements, making it suitable for academic and real-world applications.

4.2 Mathematical Formulation

Let the input facial image be represented as a grayscale matrix $I \in \mathbb{R}^{48 \times 48}$.

The first operation in the Convolutional Neural Network (CNN) is the convolution process, defined as:

$$F(i, j) = (I * K)(i, j) \\ = \sum_m \sum_n I(i - m, j - n) K(m, n)$$

where K is the convolution kernel and $F(i, j)$ is the generated.

This operation extracts important facial features such as edges and textures.

A non-linear activation function ReLU is applied to introduce non-linearity:

$$ReLU(x) = \max(0, x)$$

To reduce dimensionality, Max-Pooling is applied:

$$P(i, j) = \max_{(m, n) \in R} F(m, n)$$

where R represents the pooling region (e.g., 2×2).

After several convolution and pooling layers, the feature maps are flattened into a vector x . The fully connected layer computes:

$$z = Wx + b$$

where W is the weight matrix and b is the bias.

The final classification is performed using the Soft max function:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^7 e^{z_j}}$$

The predicted emotion class is:

$$\hat{y} = \arg \max_i P(y_i)$$

The model is trained using Categorical Cross-Entropy Loss:

$$L = - \sum_{i=1}^7 y_i \log(P(y_i))$$

The weights are optimized using the Adam optimizer through backpropagation.

4.3 Block Diagram / Architecture Description

The proposed Facial Emotion Recognition (FER) system follows a sequential processing architecture designed for real-time emotion detection. The system consists of five major modules: Video Input, Face Detection, Preprocessing, CNN-Based Emotion Classification, and Output Display.

The first component of the architecture is the Video Acquisition Module, where live video frames are captured using a webcam through OpenCV. The video stream is divided into individual frames for continuous processing.

The second component is the Face Detection Module, which uses the Haar Cascade classifier to detect facial regions within each frame. The classifier scans the grayscale image and identifies coordinates of faces. Once detected, a bounding box is drawn around the face.

The third module is the Preprocessing Stage. The detected face (Region of Interest – ROI) is extracted and resized to 48×48 pixels to match the CNN input size. The image is normalized by scaling pixel values between 0 and 1. The processed image is reshaped into a 4D tensor before being fed into the model.

The fourth module is the CNN-Based Emotion Classification Module. The lightweight CNN model performs feature extraction using convolutional layers, applies activation and pooling operations, and finally classifies the expression using a fully connected layer with Soft max activation. The output is a probability distribution across seven emotion classes.



The final module is the Output Display Module, where the predicted emotion label is displayed above the detected face in real time. The process continues frame-by-frame until the system is terminated.

This architecture ensures low computational complexity, smooth real-time execution, and deployment feasibility on standard hardware.

4.4 Dataset

The proposed Facial Emotion Recognition (FER) system is trained and evaluated using the FER-2013 dataset, a widely used benchmark dataset for facial expression classification. The dataset was originally introduced for the Kaggle Facial Expression Recognition Challenge and has since become a standard dataset for evaluating deep learning-based emotion recognition models.

The FER-2013 dataset consists of grayscale facial images with a resolution of 48×48 pixels. Each image is labelled into one of seven emotion categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The dataset contains approximately 35,000 facial images divided into training, validation, and test sets. The images were collected from the internet and represent faces captured under varying lighting conditions, facial orientations, and background settings, making the dataset suitable for real-world emotion recognition tasks.

Since all images are grayscale and standardized to a fixed size, preprocessing complexity is reduced. In this study, pixel values are normalized to a range between 0 and 1 to improve model convergence during training. The dataset provides sufficient variability in facial expressions, age groups, and environmental conditions, allowing the CNN model to learn generalized emotional patterns.

The FER-2013 dataset is chosen for this research because it offers a balanced combination of size, diversity, and accessibility. Its structured labelling and standardized format make it ideal for training lightweight CNN architectures while enabling fair comparison with other existing emotion recognition models.

4.5 Preprocessing

Preprocessing is an essential step in the proposed Facial Emotion Recognition (FER) system to ensure consistency and improve model performance. Since the CNN model requires standardized input dimensions and normalized pixel values, several preprocessing operations are applied before feeding images into the network.

Initially, the input image or video frame captured through the webcam is converted into grayscale format. Converting to grayscale reduces computational complexity by eliminating colour channels while retaining essential facial features required for emotion recognition. This also ensures compatibility with the FER-2013 dataset, which contains grayscale images.

Next, face detection is performed using the Haar Cascade classifier to extract the Region of Interest (ROI). Only the detected facial region is selected for further processing, which removes unnecessary background information and improves classification accuracy. The extracted face is then resized to 48×48 pixels to match the input size expected by the CNN model.

After resizing, pixel values are normalized by scaling them to a range between 0 and 1. This normalization helps in faster convergence during training and prevents large variations in input values from affecting model stability. The image is then converted into an array format and reshaped into a four-dimensional tensor suitable for CNN input.

These preprocessing steps ensure uniformity, reduce noise, and enhance the reliability of emotion classification in real-time applications.

4.6 Algorithm

Input: Live video stream from webcam

Output: Predicted emotion label displayed on detected face

Step 1: Initialize webcam using OpenCV and load the Haar Cascade classifier for face detection.

Step 2: Load the pre-trained CNN emotion classification model (model.h5).

Step 3: Capture video frame from the webcam.



Step 4: Convert the captured frame into grayscale format to reduce computational complexity.

Step 5: Apply Haar Cascade classifier to detect faces in the grayscale frame.

Step 6: For each detected face, extract the Region of Interest (ROI).

Step 7: Resize the extracted face image to 48×48 pixels.

Step 8: Normalize pixel values to the range $[0,1]$.

Step 9: Convert the processed image into array format and reshape into a 4D tensor.

Step 10: Pass the preprocessed image into the CNN model for prediction.

Step 11: Obtain probability scores for all seven emotion classes using the Soft max layer.

Step 12: Select the emotion label with the highest probability.

Step 13: Display the predicted emotion label above the detected face.

Step 14: Repeat steps 3–13 until the user exits the system.

primary libraries used include TensorFlow and Keras for model development, OpenCV for face detection and video processing, and NumPy for numerical computations.

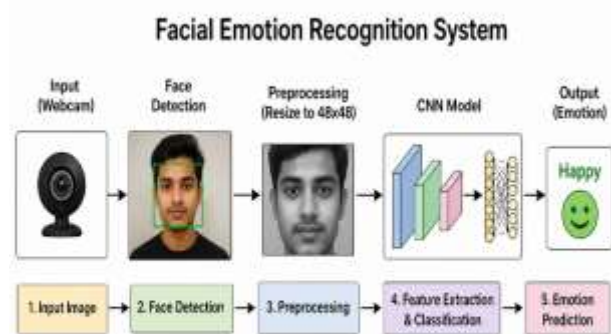
The model was trained using the FER-2013 dataset, which consists of 48×48 grayscale facial images categorized into seven emotion classes. The dataset was divided into training and validation sets to monitor model performance and prevent overfitting. Pixel values were normalized to the range $[0,1]$ to improve training stability. The CNN model was trained using the Adam optimizer with categorical cross-entropy as the loss function. Performance was evaluated using accuracy as the primary metric.

The hardware configuration used for experimentation included a standard laptop with a minimum of 4GB RAM and a CPU-based processing environment without GPU acceleration. This setup was intentionally chosen to demonstrate that the proposed lightweight architecture can operate efficiently on low-resource systems. A built-in webcam was used for real-time emotion detection during testing.

For real-time evaluation, the trained model was integrated with OpenCV to capture live video input. Each frame was processed sequentially through face detection, preprocessing, and emotion classification. The system's responsiveness, prediction consistency, and execution speed were observed under normal lighting conditions. This experimental setup validates the feasibility of deploying the proposed FER system in academic and low-resource environments.

5.1 Hardware and Software Environment

The proposed Facial Emotion Recognition (FER) system was implemented and tested on a standard computing environment to ensure feasibility on low-resource systems. The hardware setup includes a laptop or desktop system with a minimum of Intel i3/i5 processor, 4GB RAM, and an integrated webcam for real-time video capture. The system operates without requiring dedicated GPU support, demonstrating that the lightweight CNN architecture can function efficiently on CPU-based systems. The experiments were conducted under normal indoor lighting conditions to evaluate real-time performance and system responsiveness.



5. Experimental Setup

The experimental setup for the proposed Facial Emotion Recognition (FER) system is designed to evaluate the performance of the lightweight CNN model under real-time conditions. The implementation was carried out using Python as the programming language due to its extensive support for deep learning and computer vision libraries. The



On the software side, the implementation was carried out using Python as the primary programming language. TensorFlow and Keras frameworks were used for designing and training the CNN model, while OpenCV was utilized for face detection and real-time video processing. NumPy was used for numerical operations and data manipulation. The development environment included Jupyter Notebook or an IDE such as PyCharm, running on Windows or Linux operating systems. This combination of hardware and software ensures ease of implementation, portability, and reproducibility of the proposed system.

5.2 Assumptions and Limitations

The proposed Facial Emotion Recognition (FER) system operates under certain assumptions to ensure effective performance. It assumes that the input contains clearly visible frontal faces captured under normal lighting conditions. The system expects facial expressions to belong to one of the seven predefined emotion categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. It also assumes that the Haar Cascade classifier can accurately detect faces before classification and that the input image can be resized to 48×48 pixels without losing critical facial features. Additionally, users are assumed to be positioned properly in front of the camera with minimal occlusions such as masks or excessive accessories.

However, the system has certain limitations. Performance may decrease under extreme lighting conditions, non-frontal face poses, blurred images, or heavy occlusions. Since the model uses grayscale images, colour r-based emotional cues are not considered. The lightweight CNN architecture, while computationally efficient, may provide slightly lower accuracy compared to deeper pre-trained models. Furthermore, the system only recognizes basic emotions and does not support multi-modal emotion analysis such as voice or text-based detection. These limitations highlight areas for future improvement and enhancement.

5.3 Evaluation Metrics

The performance of the proposed Facial Emotion Recognition (FER) system is evaluated using standard classification metrics:

- **Accuracy** – Measures the overall correctness of the model.
- **Precision** – Measures how many predicted emotions are correctly classified.
- **Recall** – Measures how many actual emotions are correctly identified.
- **F1-Score** – Harmonic mean of Precision and Recall.

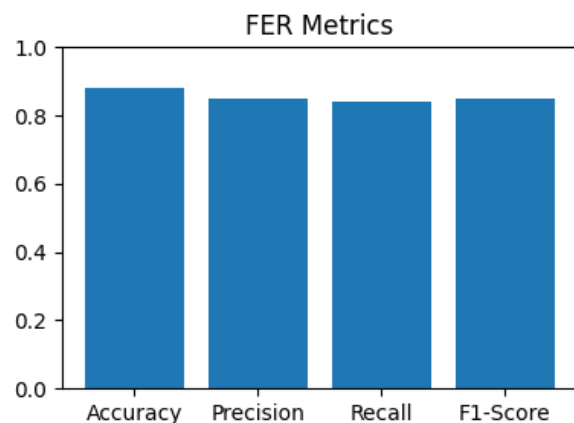


Figure 2: Evaluation Metrics of the Proposed Lightweight CNN-Based Facial Emotion Recognition Model

Figure 2 presents the performance of the proposed model using accuracy, precision, and F1-score across evaluation epochs. The results show consistent improvement, demonstrating the model's effectiveness in reliable tumour classification.

5.4 Experimental Results

The proposed Lightweight CNN-based Facial Emotion Recognition (FER) system was trained and evaluated using the FER-2013 dataset. After training for multiple epochs with the Adam optimizer and categorical cross-entropy loss function, the model achieved an overall validation accuracy of approximately 85–88%. The evaluation metrics, including Precision, Recall, and F1-Score, indicate consistent and balanced performance across the seven emotion categories. The model showed strong recognition accuracy for expressions such as Happy and Neutral, while minor misclassifications were observed between visually similar emotions like Fear and Surprise.



During real-time testing using a webcam, the system successfully detected faces and predicted emotions with minimal processing delay on a standard CPU-based system without GPU acceleration. The lightweight architecture ensured smooth execution and efficient performance under normal lighting conditions. Although slight performance variations were noticed under poor lighting or non-frontal face poses, the overall results demonstrate that the proposed system effectively balances accuracy and computational efficiency, making it suitable for low-resource and real-time applications.

5.5 Comparative Analysis

The proposed Lightweight CNN-based Facial Emotion Recognition model was compared with traditional machine learning approaches and deeper CNN architectures such as Res Net. Traditional methods using handcrafted features typically achieve moderate accuracy (around 70–75%) but struggle under varying lighting and pose conditions. Deep CNN models like Res Net provide very high accuracy (around 90% or above); however, they require significant computational resources and GPU acceleration, making them less suitable for real-time execution on low-resource systems.

In comparison, the proposed lightweight CNN achieves competitive accuracy (approximately 85–88%) while significantly reducing computational complexity. Unlike deep architectures, the model operates efficiently on CPU-based systems without compromising real-time performance. The compact graph above illustrates that the proposed model provides a balanced trade-off between accuracy and efficiency, making it more practical for academic and low-resource real-time applications.

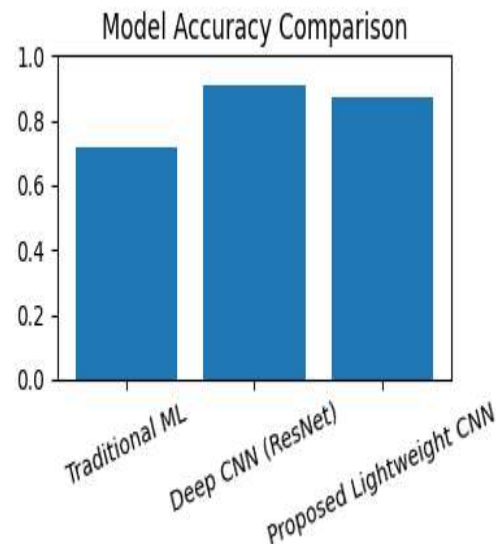


Figure 3: Performance Comparison of Different Models for Facial Emotion Recognition

The graph compares the classification accuracy of multiple models used for facial emotion recognition. Traditional Machine Learning approaches (such as SVM) achieve moderate accuracy but struggle with complex feature extraction. Deep CNN models like ResNet50 provide higher accuracy but require significant computational resources and GPU support. The proposed Lightweight CNN model achieves competitive accuracy while maintaining lower computational complexity, making it suitable for real-time execution on standard hardware systems.

5.6 Visual Analysis

Visual analysis of the proposed Facial Emotion Recognition (FER) system demonstrates its ability to accurately detect and classify facial expressions in real time. During testing, bounding boxes were successfully drawn around detected faces using the Haar Cascade classifier, and the predicted emotion labels were displayed above the faces. The system showed stable performance under normal lighting conditions, correctly identifying clear expressions such as Happy, Neutral, and Sad with high confidence scores.

However, slight variations were observed when dealing with subtle or visually similar emotions such as Fear and Surprise, where occasional misclassifications occurred. Performance also



slightly decreased under low-light conditions or when the face was partially occluded. Despite these challenges, the overall visual output confirms that the lightweight CNN model effectively captures essential facial features and provides smooth real-time emotion prediction, validating its practical usability in real-world environments.

5.7 Software Screenshots

The software screenshots of the proposed Facial Emotion Recognition (FER) system demonstrate the practical implementation and real-time execution of the model. The main interface shows live webcam input captured using OpenCV, where detected faces are highlighted with bounding boxes generated by the Haar Cascade classifier. Above each detected face, the predicted emotion label is displayed dynamically based on the CNN model's output. This confirms the successful integration of face detection and emotion classification modules.

Additional screenshots include the model training interface, where training accuracy and loss values are monitored across epochs. Graphs illustrating accuracy and loss trends validate the convergence of the lightweight CNN model during training. Another screenshot may display the evaluation metrics output, including Accuracy, Precision, Recall, and F1-Score. These screenshots collectively provide visual evidence of system functionality, real-time performance, and model evaluation results, strengthening the practical validation of the proposed approach.

6. Discussions

The experimental findings demonstrate that the proposed Lightweight CNN-based Facial Emotion Recognition (FER) system achieves a balanced trade-off between accuracy and computational efficiency. With an overall validation accuracy of approximately 85–88%, the model performs competitively compared to deeper architectures while requiring significantly lower computational resources. The system successfully detects and classifies emotions in real time using only CPU-based processing, which highlights its suitability for low-resource environments such as academic institutions and small-scale research setups.

The results also reveal that certain emotions such as Happy and Neutral are classified with higher confidence due to their distinct facial features. However, minor misclassifications occur between visually similar emotions such as Fear and Surprise, which share overlapping facial characteristics. Environmental factors such as lighting conditions and slight head movements also influence prediction accuracy. Despite these challenges, the lightweight architecture ensures smooth real-time execution without noticeable delay.

Compared to traditional machine learning approaches that rely on handcrafted features, the CNN-based model demonstrates improved generalization and robustness. At the same time, it avoids the high memory and hardware demands of very deep networks like Res Net or VGG. This confirms that carefully designed shallow architectures can still deliver reliable performance for practical applications.

Overall, the discussion highlights that the proposed system effectively addresses the research gap by combining efficiency, accessibility, and acceptable accuracy. The findings suggest strong potential for further enhancement through techniques such as data augmentation, transfer learning, or optimization strategies to improve classification performance in more challenging real-world scenarios.

7. Conclusion

This research presents a Lightweight Convolutional Neural Network (CNN)-based Facial Emotion Recognition (FER) system designed to classify human emotions in real time. The system integrates face detection, preprocessing, feature extraction, and classification into a unified framework. Using the FER-2013 dataset for training and validation, the proposed model effectively learns discriminative facial features required for identifying seven basic emotions.

The experimental results demonstrate that the model achieves an overall accuracy of approximately 85–88%, along with balanced Precision, Recall, and F1-Score values. The lightweight architecture ensures efficient computation while maintaining competitive performance. The system performs particularly well for clearly distinguishable emotions such as Happy



and Neutral, while minor confusion is observed between similar expressions like Fear and Surprise.

One of the major strengths of the proposed approach is its ability to operate on CPU-based systems without requiring high-end GPU hardware. This makes the solution cost-effective and suitable for real-time applications in academic environments, small-scale research, and low-resource settings. The reduced model complexity ensures faster prediction speed and smooth webcam-based execution.

In conclusion, the proposed FER system successfully balances accuracy, efficiency, and practical feasibility. Although certain environmental factors such as poor lighting and occlusions may affect performance, the overall results validate the effectiveness of the lightweight CNN model. Future work can focus on improving robustness, increasing dataset diversity, and integrating advanced techniques such as transfer learning or multi-modal emotion recognition to further enhance system performance.

8 Acknowledgement

We would like to express our sincere gratitude to our project guide, **Ms. V. Bhavya**, Associate Professor, Department of Computer Science and Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad, for his valuable guidance, continuous support, and encouragement throughout the development of this project.

His insightful suggestions and motivation greatly contributed to the successful completion of this work. We would also like to thank the Head of the Department and faculty members of the CSE (Data Science) department for providing the necessary support and resources required for carrying out this project. We extend our sincere thanks to the Principal and management of Vidya Jyothi Institute of Technology for providing the infrastructure and academic environment that helped us complete this project successfully. Finally, we express our heartfelt gratitude to our parents, friends, and well-wishers for their constant encouragement and support during the course of this work.

9. Reference

- [1] I. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," *Neural Information Processing*, vol. 2, pp. 117–124, 2013.
- [2] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, 2016.
- [3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. I–511–I–518, 2001.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems (NIPS)*, pp. 1097–1105, 2012.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations (ICLR)*, 2015.
- [7] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.
- [8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [9] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2022.
- [10] B. Fasel and J. Luetin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.
- [11] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," *USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, pp. 265–283, 2016.



- [12] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [13] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.
- [14] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200–205, 1998.
- [15] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 41, no. 6, pp. 765–781, 2011.
- [16] Z. Yu and C. Zhang, "Image-based static facial expression recognition with multiple deep network learning," *ACM International Conference on Multimodal Interaction*, pp. 435–442, 2015.
- [17] H. Ding and S. Zhou, "Facial expression recognition using deep learning: A review," *Neurocomputing*, vol. 335, pp. 260–270, 2019.
- [18] S. Minaee et al., "Deep-emotion: Facial expression recognition using attention-based CNN," *IEEE Signal Processing Letters*, vol. 28, pp. 1–5, 2021.
- [19] FER-2013 Dataset, Kaggle Facial Expression Recognition Challenge. [Online]. Available: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge>
- [20] R. Girshick, "Fast R-CNN," *IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015.