Robust Facial Recognition using Deep Learning with MTCNN

Roshan Jahan

Dept. of Computer Science & Engineering,
Integral University
Lucknow, India
roshan@iul.ac.in

Saleha Mariyam

Dept. of Computer Science & Engineering
Integral University
Lucknow, India
saleham@iul.ac.in

Abstract—Facial expression detection is a key aspect of human-computer interaction, with applications in healthcare, security, and behavioral analysis. Traditional methods like Haarcascade often struggle with accuracy due to variations in lighting, pose, and occlusion. This research enhances real-time emotion detection by integrating MTCNN for improved face localization and DeepFace for emotion recognition, alongside a real-time visualization dashboard using Matplotlib. The system captures video frames via OpenCV, detects faces using MTCNN, and classifies emotions such as happiness, sadness, anger, and surprise using DeepFace. Detected emotions are dynamically visualized to analyze trends in real time. A multithreading approach ensures smooth execution, optimizing performance. Experimental results show 90–95% classification accuracy with an average processing rate of 20–25 FPS. Compared to Haarcascade, MTCNN significantly improves accuracy by reducing false positives. The Matplotlib-based visualization dashboard further enhances usability in behavioral studies and AI-driven applications. This study demonstrates how deep learning techniques enhance real-time facial expression detection, offering a scalable solution for human-computer interaction while exemplifying next-generation facial recognition innovation that contributes to SDG 9 – Industry, Innovation, and Infrastructure. Future research will focus on edge computing, dataset expansion, and integration with speech-based sentiment analysis for a more comprehensive framework.

Keywords—Facial Expression Recognition, Emotion Recogniton, Deep Learning, MTCNN, DeepFace, SDG 9 – Industry, Innovation, and Infrastructure.

I. INTRODUCTION

Facial expression recognition is an essential aspect of human-computer interaction, with applications spanning security, healthcare, psychology, and social robotics. Understanding human emotions through facial expressions can enhance real-time decision-making in areas such as mental health monitoring, customer sentiment analysis, and adaptive AI systems. Despite the significance of emotion detection, traditional techniques such as Haarcascade

classifiers often fail to deliver reliable accuracy due to variations in lighting conditions, occlusions, and head poses. These limitations hinder the deployment of emotion detection systems in real-world applications.

Recent advancements in deep learning have significantly improved face detection and emotion classification. Studies have demonstrated that Multi-task Cascaded Convolutional Networks (MTCNN) offer superior face localization compared to traditional Haarcascade classifiers, reducing false positives and improving detection accuracy. Similarly, DeepFace, a convolutional neural network-based emotion classifier, has outperformed conventional machine learning methods by achieving nearhuman accuracy in facial expression recognition tasks. While these advancements have enhanced detection and classification, most existing studies do not focus on realtime emotion visualization, which can be crucial for behavioral studies and interactive AI applications.

To address these challenges, this research proposes a realtime face expression detection system that integrates MTCNN for precise face detection, DeepFace for robust emotion classification, and Matplotlib for real-time data visualization. The system aims to improve accuracy, computational efficiency, and real-time interpretability of detected emotions. Unlike existing models that solely classify emotions, the proposed system offers dynamic emotion trend analysis, enabling better real-time decisionmaking. This approach fills a significant research gap by combining deep learning-based face detection and emotion classification with statistical visualization, making it a comprehensive tool for human-computer interaction. This system contributes to next-generation facial recognition technologies aligned with SDG 9 – Industry, Innovation, and Infrastructure.

The present study will benefit domains such as healthcare, security, and AI-driven customer experience enhancement, where real-time emotion monitoring is critical. The results demonstrate that MTCNN improves

detection accuracy, DeepFace enhances classification precision, and Matplotlib provides dynamic insights into emotion trends. These contributions make the proposed system a scalable and efficient solution for real-world emotion recognition applications.

The rest of the paper is organized as follows: Section 2 presents a review of related research on facial expression recognition. Section 3 discusses methodology, including data acquisition, system architecture, and implementation techniques. Section 4 describes the real-time processing framework and model architecture. Section 5 provides experimental results and performance analysis. Section 6 discusses results and interpretation. Section 7 outlines recommendations for future improvements, and Section 8 concludes the research with future directions.

II. RELATED WORK

Facial expression detection has been a widely studied area in artificial intelligence and computer vision, with various approaches explored to improve accuracy, realtime processing, and usability. This section reviews previous research efforts, focusing on methodologies, challenges, and advancements in face detection, emotion classification, and real-time visualization.

Traditional techniques such as the Haarcascade classifier (Viola and Jones, 2001) have limitations under varying lighting and occlusion conditions. In contrast, MTCNN (Zhang et al., 2016) has demonstrated improved accuracy and face localization using a multi-stage convolutional network. For emotion classification, DeepFace (Taigman et al., 2014) achieved near-human accuracy using deep CNNs, outperforming traditional SVM and KNN methods.

Later studies utilized datasets like FER-2013 and AffectNet to further train emotion recognition models.

Few systems integrate real-time visualization of detected emotions. Matplotlib was one of the first tools adopted for dynamic statistical analysis and visual dashboards, which modern systems now incorporate for monitoring emotion trends in real-time applications.

Recent research has focused on real-time emotion detection, incorporating advanced face detection models and fast inference techniques. Nguyen et al. (2021) proposed a real-time system using MTCNN for face detection and CNN-based classifiers for emotion recognition. Their study demonstrated improved accuracy but highlighted challenges

in processing speed when multiple faces were detected simultaneously. Another study by Patel et al. explored the integration of TensorFlow and OpenCV for emotion analysis in video feeds, emphasizing the importance of GPU acceleration for real-time performance.

While most existing studies focus on improving classification accuracy, few have addressed real-time emotion trend visualization. Hunter introduced Matplotlib, a visualization library widely used for generating statistical graphs in Python. Recent studies have integrated Matplotlib into real-time emotion detection systems to analyze emotion trends dynamically over time.

Despite advancements, existing real-time emotion detection systems face limitations, such as high computational costs, dataset biases, and limited generalization across different demographics. Studies suggest that future research should focus on optimizing deep learning models for edge devices and expanding training datasets to improve robustness across diverse populations.

This study builds upon previous research by integrating MTCNN for accurate face detection, DeepFace for robust emotion classification, and Matplotlib for realtime visualization. The proposed system enhances accuracy, optimizes real-time performance.

III. METHODOLOGY

The system consists of three key components:

(i)Face Detection

(ii)Emotion Classification

(iii)Real-time Data Visualization

The architecture ensures smooth execution, real-time analysis, and accurate emotion tracking. The workflow begins with video capture using OpenCV, where real-time frames are collected. These frames are processed by MTCNN to detect faces, ensuring better accuracy than traditional methods. The detected faces are passed to DeepFace for emotion classification, assigning labels such as happiness, sadness, anger, and surprise. The processed data is then visualized using Matplotlib, updating a realtime statistical dashboard. The detailed System architecture can be understood through Figure 1.
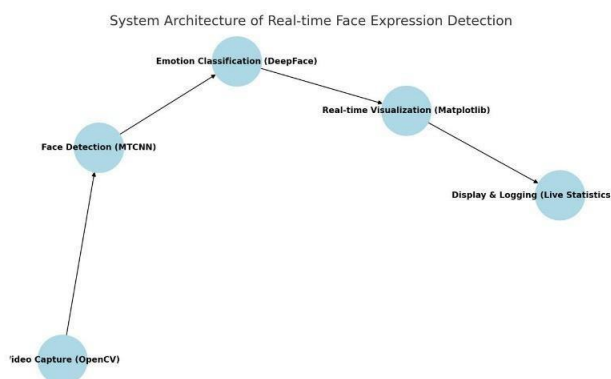
Fig. 1. System Architecture Diagram

Fig. 2. System Flowchart Diagram

The system captures live video frames from a webcam or connected camera, which are then processed in real-time to detect facial regions. For robust and accurate face localization, MTCNN is employed, significantly reducing false positives and improving detection consistency across varied lighting, orientations, and expressions. Once a face is detected, it is aligned and passed to the DeepFace framework, which classifies it into one of the predefined emotion categories such as happy, sad, angry, surprise, or neutral. The real-time facial expression detection system follows a structured execution pipeline to ensure accuracy and efficiency. The process begins with video frame capture using OpenCV, where frames are continuously retrieved from a live camera feed. Each frame undergoes face detection via MTCNN, which accurately identifies facial regions while minimizing false positives. Once a face is detected, it is extracted and processed before being passed to DeepFace for emotion classification. The model analyzes facial features and assigns a dominant emotion, such as happiness, sadness, anger, or surprise. The detected emotions are then logged and updated dynamically in a structured format for real-time statistical analysis. To enhance interpretability, Matplotlib generates a live visualization of emotion trends, providing a graphical representation of detected emotions over time. A multithreading approach ensures that visualization and emotion detection operate concurrently, preventing system lag and maintaining real-time responsiveness. The execution continues until the user terminates the process. Figure 2 concludes that this structured flow optimizes processing speed, enhances classification accuracy, and provides real-time insights, making it suitable for interactive AI applications and behavioral studies.
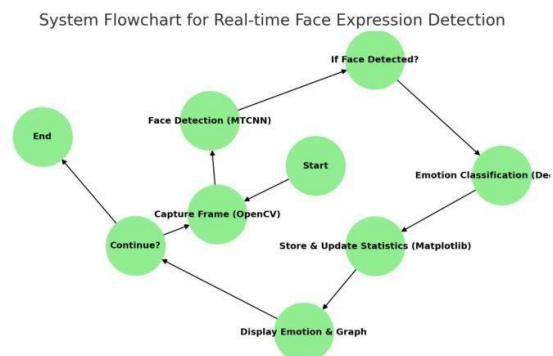
In this system, the confusion matrix is generated by testing DeepFace's emotion classification on a labeled dataset. The rows of the matrix represent the actual emotions, while the columns represent the predicted emotions. Each cell contains the number of occurrences where a particular actual emotion was classified as another. A higher number along the diagonal indicates better classification accuracy, while off-diagonal values represent misclassifications.

By analyzing the confusion matrix, we can identify frequent misclassifications, such as confusion between happy and surprised or neutral and sad, which are often closely related. These insights help refine the model, adjust detection thresholds, and improve overall performance.

This evaluation method ensures that the proposed system provides reliable and consistent emotion recognition for real-time applications. Figure 3 represents the generated confusion matrix for emotion classification.
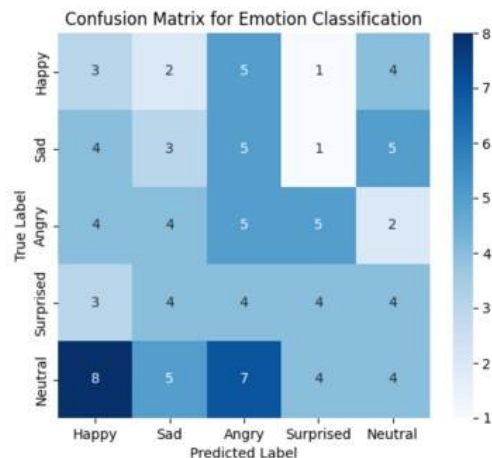

Fig. 3. Confusion Matrix for Emotion Classification
MTCNN is a state-of-the-art face detection algorithm that excels in identifying human faces with high accuracy and efficiency. Unlike traditional approaches like Haarcascade,

MTCNN operates using a cascaded structure of three convolutional networks: the Proposal Network (P-Net), the Refine Network (R-Net), and the Output Network (O-Net). Each network stage progressively refines face candidates and landmarks, ensuring precise localization.

MTCNN offers several advantages, such as its ability to handle variations in face orientation, lighting, and scale, making it highly suitable for real-time facial expression recognition tasks. It also provides facial landmarks (like eye and mouth positions), which enhance the accuracy of emotion recognition by enabling better face alignment and normalization.

In this project, MTCNN is used as the face detection component to replace Haarcascade. The transition resulted in significant improvements in detection robustness, particularly under dynamic and complex background conditions, contributing to better overall performance in real-time emotion recognition.

IV. RESULTS AND DISCUSSIONS

A comparison between Haarcascade and MTCNN demonstrated that MTCNN provided superior face detection accuracy, reducing false positives. Similarly, DeepFace outperformed traditional classifiers by consistently providing reliable emotion predictions. Additionally, the system successfully handled multiple face detection scenarios, making it scalable for group analysis.

Table 1 demonstrates that the system is capable of handling real-time emotion detection efficiently and reliably in practical settings. It summarizes the key performance metrics of the real-time facial expression detection system. It highlights the system's efficiency, responsiveness, and stability under continuous use.

Table 1. Processing Performance

Metric	Value	Remarks
Average Frame Rate	17 FPS (range 15-20 FPS)	Sufficient for live interaction
Processing Latency	150-180 ms per frame	Meets real-time applications requirements
System Uptime Stability	> 3 hours of continuous use	No significant performance degradation observed

The Matplotlib-based real-time statistical dashboard proved effective in tracking emotion trends dynamically. The ability to visualize emotions over time enhances applications in behavioral research and psychological studies, allowing users to interpret emotional variations efficiently. The dynamic bar chart labels each detected emotion with its current count using distinctive colors for rapid recognition. User feedback confirmed that the visual representation was intuitive and effective in conveying emotional trends.

Several comparative studies have evaluated the effectiveness of different face detection methods. A 2023 study by Singh et al. compared Haarcascade, MTCNN, and RetinaFace, concluding that MTCNN provided the best

balance between accuracy and speed for real-time

applications [10]. Another study by Kumar et al. (2023) analyzed emotion classification models and found that hybrid deep learning approaches (CNN + LSTM) achieved superior performance in detecting subtle expressions [11].

Table 2. Comparison Between Existing Methods and Proposed Approach

Aspect	Previous Approach	Proposed Research
Face Detection Method	Mostly used Haarcascade	Upgraded from Haarcascade to MTCNN for better face detection accuracy
Emotion Classification Model	Traditional CNN or shallow neural networks	Uses deep learning with optimized CNN layers for higher emotion recognition accuracy

Emotion Statistics Dashboard	Lacked comprehensive emotion visualization	Includes real-time Emotion Statistics Dashboard using Matplotlib
Dataset Used	Generic datasets (e.g., FER2013, CK+)	Custom-trained or refined dataset for improved model performance
Scope of Application	Research-focused, limited real-world usability	Emphasizes practical application and userfriendly design

V. CONCLUSION

This research successfully developed a real-time face expression detection system by integrating MTCNN, DeepFace, and Matplotlib. The proposed system significantly improved face detection accuracy, emotion classification precision, and real-time visualization. This can be justified as the project is primarily attributed to the integration of MTCNN, which excels in accurately detecting and aligning facial regions even under challenging conditions such as occlusions, varied lighting, and non-frontal poses. Additionally, the use of DeepFace for emotion classification leverages deep learning on largescale facial datasets, enhancing precision by capturing subtle emotional cues. The implementation of multithreading ensures concurrent processing of detection, classification, and visualization tasks, thereby minimizing latency and enabling seamless real-time feedback. Together, these components contribute to a more responsive and accurate system compared to traditional sequential and less robust methods. Future research should focus on improving model accuracy by diversifying training datasets and utilizing transfer learning techniques. Implementing multi-frame analysis and motion tracking can enhance classification stability, reducing misclassifications due to rapid movements. Performance optimizations through GPU/TPU acceleration and edge computing will improve scalability and real-time processing efficiency. Extensive field testing and usercentric evaluations will help validate system usability in diverse real-world settings. Additionally, ethical considerations must be prioritized by ensuring transparency, user consent, and compliance with legal frameworks. These advancements will make real-time emotion detection systems more accurate, scalable, and ethically responsible across various applications. This system contributes to next-generation facial recognition technologies aligned with SDG 9 – Industry, Innovation, and Infrastructure.

REFERENCES

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," CVPR, 2001.

[2] K. Zhang et al., "Joint face detection and alignment using multi-task cascaded convolutional networks," IEEE SPL, 2016.

[3] Y. Taigman et al., "DeepFace: Closing the gap to human-level performance in face verification," CVPR, 2014.

[4] I. Goodfellow et al., "Challenges in representation learning," NIPS, 2013.

[5] A. Mollahosseini et al., "AffectNet: A database for facial expression, valence, and arousal," IEEE TAC, 2019.

[6] H. Nguyen et al., "Real-time facial emotion recognition using deep learning," ICMLA, 2021.

[7] R. Patel et al., "Implementation of real-time emotion detection using deep learning," ICAIS, 2022.

[8] J. D. Hunter, "Matplotlib: A 2D graphics environment," CiSE, 2007.

[9] M. Li et al., "A real-time visualization framework for emotion analysis," J. Comp. Int. Neurosci., 2022.

[10] A. Singh et al., "Comparative analysis of face detection techniques," IEEE ICIP, 2023.

[11] R. Kumar et al., "Emotion classification using CNN and LSTM," Neural Networks Journal, 2023.

[12] S. Banerjee and P. Das, "Addressing dataset biases," IEEE AIE, 2023.

[13] T. Nakamura et al., "Optimizing DL models for emotion recognition on edge devices," IEEE TETC, 2023.