



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 9.935**

**Volume 15, Issue 5, May 2026**

# Real-Time Emotion Detection using CNN and Song Recommendation System

Ms. T. P. Gurav<sup>1</sup>, Saijya Mulla<sup>2</sup>, Shweta Chougale<sup>3</sup>, Aiman Mulla<sup>4</sup>, Prerana Kamble<sup>5</sup>.

Asst. Prof, Dept. of CSE (DS), D.Y. Patil of Engg. & Tech., Kolhapur, Maharashtra, India<sup>1</sup>

UG Scholar, Dept. of CSE (DS), D.Y. Patil of Engg. & Tech., Kolhapur, Maharashtra, India<sup>2</sup>

UG Scholar, Dept. of CSE (DS), D.Y. Patil of Engg. & Tech., Kolhapur, Maharashtra, India<sup>3</sup>

UG Scholar, Dept. of CSE (DS), D.Y. Patil of Engg. & Tech., Kolhapur, Maharashtra, India<sup>4</sup>

UG Scholar, Dept. of CSE (DS), D.Y. Patil of Engg. & Tech., Kolhapur, Maharashtra, India<sup>5</sup>

**ABSTRACT:** Human emotions are often expressed through facial movements, making facial expressions a key element of non-verbal communication. Detecting these emotions from video streams has become increasingly important in domains such as healthcare, behavioral analysis, and intelligent human-computer interaction. Unlike static images, video input provides continuous emotional transitions, enabling more reliable interpretation. However, practical challenges such as inconsistent lighting, facial occlusions, head movements, and uneven data distribution can impact performance. This work presents a real-time facial emotion recognition system built using Convolutional Neural Networks (CNNs). The system processes live video frames and applies preprocessing techniques, including normalization and dataset balancing, to improve robustness. The CNN model is designed with multiple layers, supported by batch normalization and dropout to ensure stable learning and reduce overfitting. To enhance usability, the system also includes a music recommendation module that suggests songs based on the detected emotional state, offering a more engaging and personalized user experience.

capabilities in Brazil.

**KEYWORDS:** Facial Emotion Recognition, Convolutional Neural Network (CNN), Deep Learning, Emotion Analysis, Real-Time Systems, Music Recommendation.

## I. INTRODUCTION

Facial expressions are one of the most natural ways humans communicate emotions and intentions. In many situations, they convey more information than spoken words. Enabling machines to interpret these expressions allows for more intuitive and responsive systems. Such capabilities are valuable in areas like healthcare monitoring, e-learning platforms, surveillance, and interactive entertainment.

Recent developments in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved the accuracy of emotion recognition systems. These models can automatically learn complex facial features, making them highly effective for real-time applications. Despite these advancements, many systems stop at emotion detection and do not provide any meaningful feedback or response to the user.

To make the system more interactive, this project introduces a song recommendation feature that responds to the detected emotional state. By suggesting music that complements or improves the user's mood, the system creates a more engaging and human-centered experience.

## II. PROBLEM STATEMENT

To develop a real-time system capable of identifying human emotions from live video using CNN and providing relevant song recommendations based on the detected emotional state.

### III. MOTIVATION

As technology becomes more integrated into daily life, the ability of systems to understand human emotions is becoming increasingly important. Many traditional approaches to emotion detection struggle with accuracy, real-time performance, and user interaction.

This project is driven by the idea of building a system that not only detects emotions reliably but also responds in a meaningful way. CNN-based models enable the detection of both clear and subtle facial expressions from live video. By combining this capability with a music recommendation feature, the system goes beyond analysis and actively enhances the user's experience.

#### KEY FEATURES:

- **Real-Time Video Processing Pipeline:** Captures live video frames and processes them efficiently to ensure smooth performance with minimal delay.
- **Facial Region Detection and Alignment:** Uses preprocessing methods (such as MTCNN or BlazeFace) to accurately locate and focus on key facial areas like the eyes, eyebrows, and mouth.
- **CNN-Driven Feature Learning:** Applies a deep CNN architecture to learn both prominent expressions and subtle facial changes, leading to improved emotion recognition accuracy.
- **Enhanced Emotion & Stress Analysis Layer:** Integrates facial features with additional cues to provide a more comprehensive understanding of emotional and stress levels.
- **Emotion-Aware Music Recommendation:** Recommends songs based on the detected emotional state, creating a more personalized and engaging user experience.

Overall, this system connects emotion recognition with practical application, resulting in a solution that is not only intelligent but also interactive and useful in real-world scenarios.

### IV. LITERATURE REVIEW

Facial Expression Recognition (FER) has been widely studied due to its importance in applications such as human-computer interaction, virtual environments, and mental health monitoring. With the rise of deep learning, especially Convolutional Neural Networks (CNNs), there has been a noticeable improvement in the accuracy and speed of emotion recognition systems. Unlike traditional approaches, CNNs can automatically learn and extract meaningful facial features, making them more efficient and reliable.

Recent research has focused on improving the performance of FER systems by using techniques like transfer learning, where pre-trained models are adapted for emotion detection tasks. This helps in capturing complex facial patterns more effectively. However, challenges such as variations within the same emotion (intra-class) and similarities between different emotions (inter-class) still affect performance.

Many studies also highlight issues like overfitting and imbalanced datasets, which can reduce model reliability. To address these problems, preprocessing methods such as normalization, histogram equalization, and data augmentation are commonly applied. These techniques improve data quality and ensure better generalization of the model across different conditions.

Overall, CNN-based approaches have proven to be highly effective in capturing both prominent and subtle facial expressions. When combined with proper preprocessing and real-time processing techniques, FER systems become more accurate and practical for real-world applications.

V. SYSTEM ARCHITECTURE

Facial Emotion Recognition to YouTube Song Recommendation System  
( Using MySQL Database with XAMPP Server )

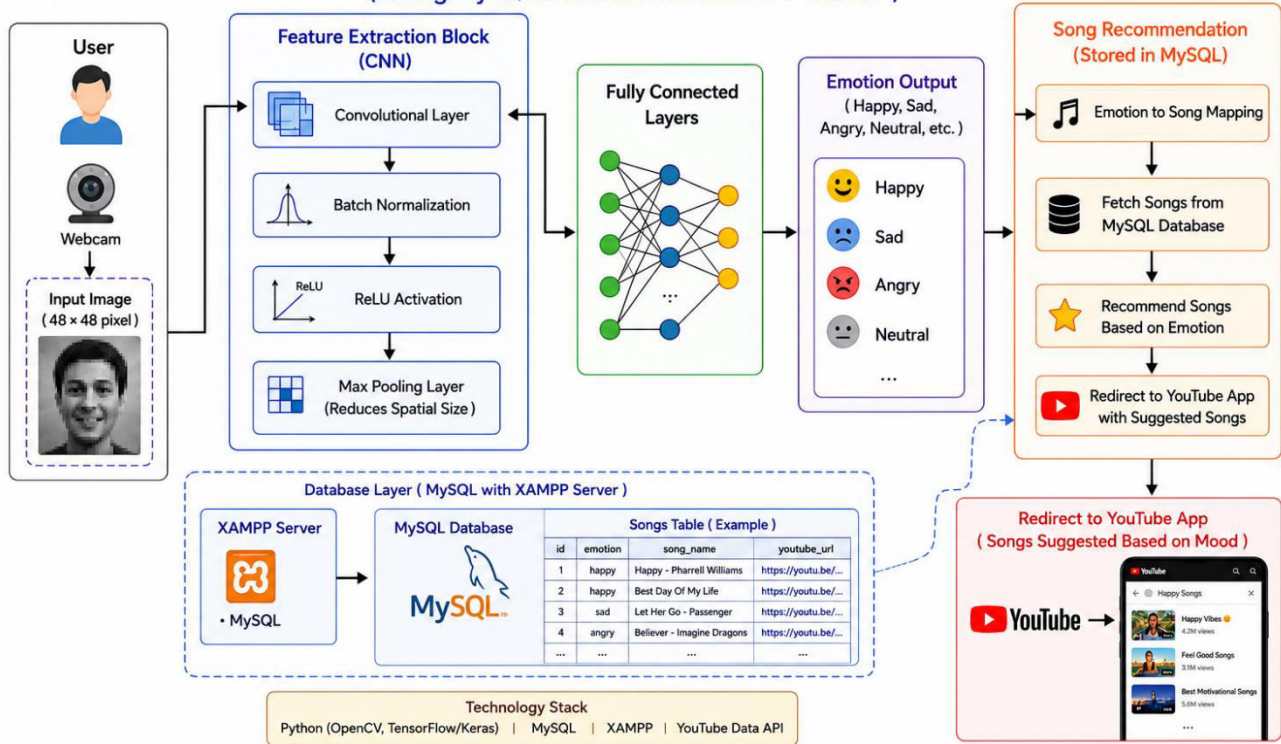


Figure 1: System Architecture

- User Module:** Captures facial expressions using webcam.
- Image Processing Module:** Resizes and preprocesses the input image (48x48).
- Feature Extraction Module (CNN):** Extracts facial features using convolution, ReLU, batch normalization, and pooling layers.
- Emotion Detection Module:** Identifies emotions like happy, sad, angry, and neutral.
- Song Recommendation Module:** Maps emotions with suitable songs.
- Database Module (MySQL + XAMPP):** Stores song details and YouTube links.
- YouTube Redirection Module:** Redirects users to YouTube for mood-based song suggestions

VI. EXISTING SYSTEM

Earlier facial expression recognition methods relied mainly on manually designed features, such as geometric measurements and texture-based descriptors. These approaches were limited in performance because they were highly sensitive to variations in lighting, facial orientation, and occlusions. Additionally, manual feature extraction made them less suitable for real-time applications. Modern systems have shifted towards deep learning techniques, particularly CNN-based models trained on datasets like FER-2013, CK+, and AffectNet. These systems generally follow a structured pipeline:

- Face Detection:** Techniques such as Haar cascades, HOG, MTCNN, Dlib, and YOLO are used to locate faces in real time.
- Preprocessing:** Input images are standardized using methods like grayscale conversion, resizing, and normalization.
- Emotion Recognition:** CNN models classify facial expressions into categories such as happy, sad, angry,

surprised, fearful, disgusted, and neutral.

4. **Stress Analysis:** Some systems estimate stress levels based on facial cues or additional indicators.
5. **Deployment:** Lightweight architectures like MobileNet and EfficientNet are used for real-time applications on edge devices and web platforms.

Although these systems perform well in detection, they often lack personalization and interactive features.

#### **LIMITATIONS OF EXISTING SYSTEM:**

- Limited understanding of contextual factors affecting emotions
- Simplified handling of stress and emotional intensity
- Performance drops under challenging conditions like poor lighting
- Lack of adaptability to individual users
- No interactive response or feedback mechanism
- Absence of integrated recommendation features

#### **VII. PROPOSED SYSTEM**

The proposed system utilizes a CNN model trained on grayscale facial images of size 48×48 pixels. The architecture includes convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. Techniques such as batch normalization and dropout are incorporated to improve performance and prevent overfitting.

The system is implemented using OpenCV for real-time video capture and processing. It continuously detects facial expressions and classifies them into predefined emotion categories. Based on the detected emotion, a song recommendation module suggests suitable music, making the system more engaging and responsive.

#### **VIII. IMPLEMENTATION RESULTS**

The developed system performs real-time emotion recognition using a webcam and a CNN-based model. It begins by capturing live video frames, which are then preprocessed by converting them into grayscale and resizing them to 48×48 pixels. This step ensures consistency and reduces computational complexity, allowing faster processing.

The processed frames are passed through multiple convolutional layers, where important facial features such as eye movements, lip positions, and eyebrow changes are extracted. Techniques like batch normalization and ReLU activation help stabilize the learning process and improve the model's ability to recognize complex expressions.

After feature extraction, the data is flattened and fed into fully connected layers, where the system classifies emotions into categories like happy, sad, angry, surprised, and neutral using a softmax function. To enhance user interaction, the system includes a song recommendation module that suggests music based on the detected emotion. For example, it may recommend energetic songs for positive moods or calming music during stressful situations.

Overall, the system demonstrates reliable performance in real-time conditions and provides an engaging user experience by combining emotion detection with personalized recommendations.

IX. FUTURE SCOPE

The future scope of this project can be extended by integrating music streaming APIs such as the Spotify API or YouTube Music API to provide real-time song playback and personalized playlists instead of using static CSV song datasets. The system can also be enhanced with multilingual song recommendation, voice-based emotion analysis, and user authentication features to improve personalization and user experience.

The project can further be improved by using advanced deep learning models such as ResNet, EfficientNet, or Transformer-based architectures for higher emotion detection accuracy in different lighting conditions and facial angles. Future versions may also support mobile applications, cloud deployment, and wearable device integration for real-time emotion-aware entertainment and mental wellness support systems.

X. RESULTS & ANALYSIS

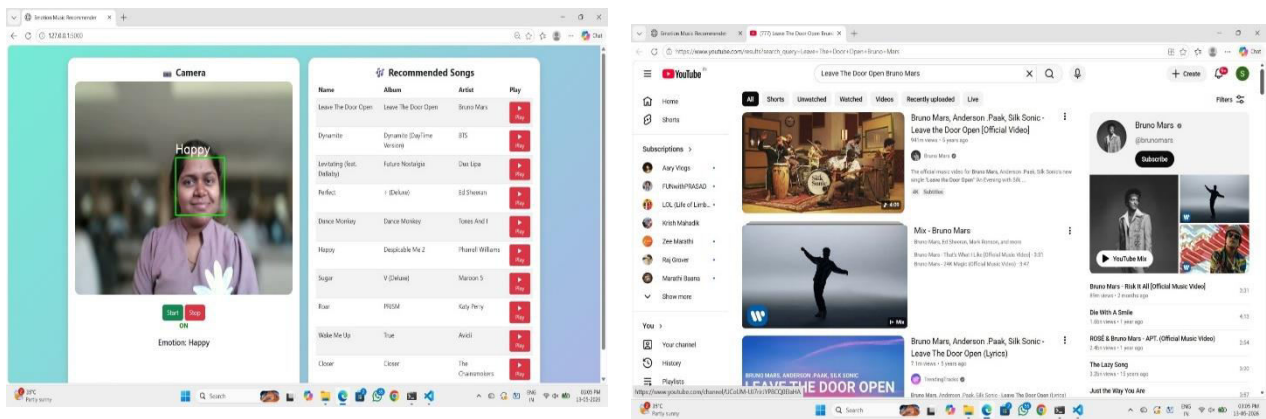


Fig.6. Emotion (Happy)

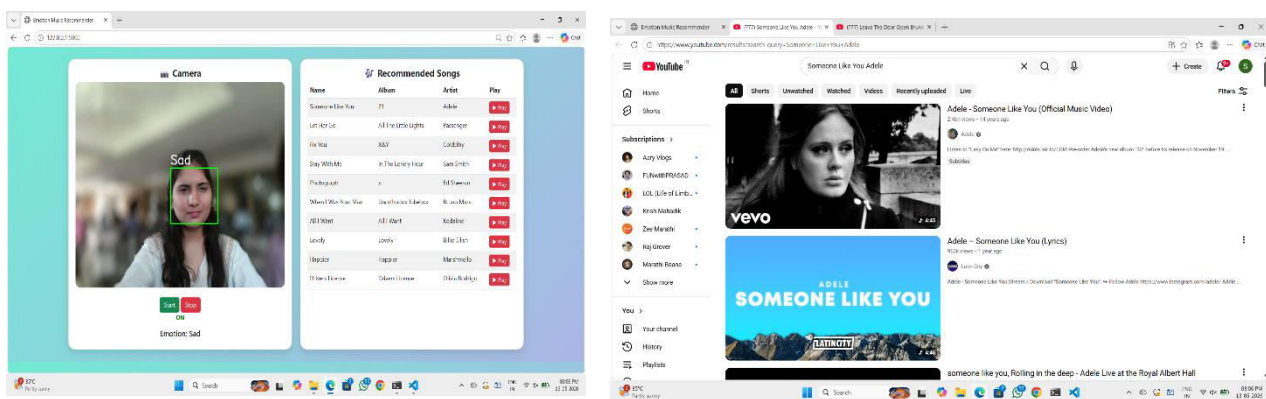


Fig.7. Emotion (Sad)

XI. CONCLUSION

This project demonstrates a practical approach to real-time facial emotion recognition using a CNN-based model. The system effectively processes live video input and accurately classifies emotional states. The addition of a song recommendation module enhances user interaction by providing meaningful responses based on detected emotions. The proposed system not only improves user experience but also highlights the potential of combining emotion recognition with intelligent applications in areas such as healthcare, entertainment, and human-computer interaction.

#### REFERENCES

- [1] P. Singh, R. Kumar, and S. Yadav, “Facial emotion recognition using CNN and transfer learning techniques,” *Procedia Computer Science*, vol. 226, pp. 1089–1098, 2024.
- [2] S. Zhao, H. Li, and Y. Wang, “CNN-based recognition of facial expressions with consideration of inter-class and intra-class variability,” *Sensors*, vol. 23, no. 24, pp. 9658–9672, 2023.
- [3] A. Li and W. Deng, “A comprehensive survey on deep learning for facial expression recognition,” *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2023.
- [4] A. Li and W. Deng, “A comprehensive survey on deep learning for facial expression recognition,” *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2023.
- [5] M. Lee and J. Kim, “Convolutional neural network-based analysis of facial expressions,” *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 567–578, 2020.
- [6] R. Gupta and S. Verma, “Neural network framework for detecting emotions in video sequences,” *Journal of Multimedia Processing*, vol. 15, no. 3, pp. 201–210, 2019.
- [7] S. Yang and T. Chen, “Improved Haar cascade techniques for face detection,” *Computer Vision and Image Understanding*, vol. 150, no. 6, pp. 78–85, 2018.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details