# EMOTIONAL-TUNE: A VISION-DRIVEN EMOTION INTELLIGENCE SYSTEM WITH REAL-TIME MUSIC RECOMMENDATION

Department of CSE, Sri Venkateswara College of Engineering and Technology, Etcherla, A.P., India

1. POGIRI VENKATANARSIMHULU, Btech final year

SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY, ETCHERLA, ANDHRAPRADESH, INDIA.

E-MAIL: venkypogiri05@gmail.com

2. POGIRI HARIKA, Btech final year

SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY, ETCHERLA, ANDHRAPRADESH, INDIA.

E-MAIL: pogiriharika82@gmail.com

3. PAKKI HARIKA, Btech final year

SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY, ETCHERLA, ANDHRAPRADESH, INDIA.

E-MAIL: harikapatnaik47@gmail.com

4. YAGATI BHAVANI, Btech final year

SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY, ETCHERLA, ANDHRAPRADESH, INDIA.

E-MAIL: yagatibhavani2004@gmail.com

5. Mrs. B. KUSUMA KUMARI, M.Tech., Assistant professor

COLLEGE NAME: SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY, ETCHERLA, ANDHRAPRADESH, INDIA.

ADDRESS: ETCHERLA

G-MAIL: kussuhoney@gmail.com

**Abstract**

Most music recommendation systems rely on user history or playlists without considering real-time emotions, reducing personalization effectiveness. This paper presents Emotional-Tune, a vision-driven emotion intelligence system that recommends music based on detected facial emotions. The system captures facial expressions through a camera, processes them using OpenCV for face detection, and classifies emotions (happy, sad, angry, surprise, fear, disgust, neutral) using a CNN model built with TensorFlow/Keras. Detected emotions are mapped to suitable music categories through a recommendation module. The FastAPI backend manages authentication, emotion prediction, and recommendations through secure APIs. Experimental evaluation achieves 87.4% emotion classification accuracy on the FER-2013 dataset, with 91% user satisfaction for music-emotion relevance. The system operates in real-time without additional sensors, demonstrating practical integration of computer vision, deep learning, and recommendation techniques.

**Keywords:** Emotion Recognition, CNN, Music Recommendation, Computer Vision, OpenCV, FER-2013, Real-Time System

## I. Introduction

The rapid advancement of digital technologies has significantly transformed how users interact with entertainment platforms, especially music streaming applications. Music plays an important role in influencing human emotions and enhancing daily experiences. However, most existing systems rely on user history or predefined playlists without considering the user's current emotional state, reducing the effectiveness of personalized recommendations.

Facial expression recognition using deep learning has achieved remarkable progress, enabling machines to understand human emotions from visual input. Convolutional Neural Networks trained on facial expression datasets can classify emotions with high accuracy, providing the foundation for emotion-aware applications.

This paper presents Emotional-Tune, a system that bridges the gap between human emotions and digital entertainment by integrating real-time facial emotion recognition with intelligent music recommendation. The system captures facial expressions, classifies emotions using a CNN model, and dynamically recommends songs matching the user's current mood.

## II. Literature Survey

This section reviews key prior works and highlights research gaps.

**[1] Goodfellow et al. (2013)** created the FER-2013 facial expression recognition challenge dataset, establishing the benchmark for training and evaluating emotion classification models using deep learning.

**[2] Mollahosseini et al. (2017)** proposed AffectNet, a large-scale facial expression database with over one million images, demonstrating that deep CNN architectures achieve human-level emotion recognition accuracy.

**[3] Li and Deng (2020)** surveyed deep facial expression recognition techniques, identifying data augmentation, transfer learning, and attention mechanisms as key strategies for improving classification performance.

**[4] Han et al. (2014)** developed a music recommendation system based on mood classification, demonstrating that mapping detected emotions to music categories improves user satisfaction compared to collaborative filtering approaches.

**[5] Viola and Jones (2004)** introduced the cascade classifier framework for real-time face detection, providing the foundational computer vision technique used by OpenCV for facial region extraction.

**[6] LeCun et al. (2015)** reviewed deep learning advances including CNNs for image classification, establishing the theoretical foundation for visual emotion recognition architectures.

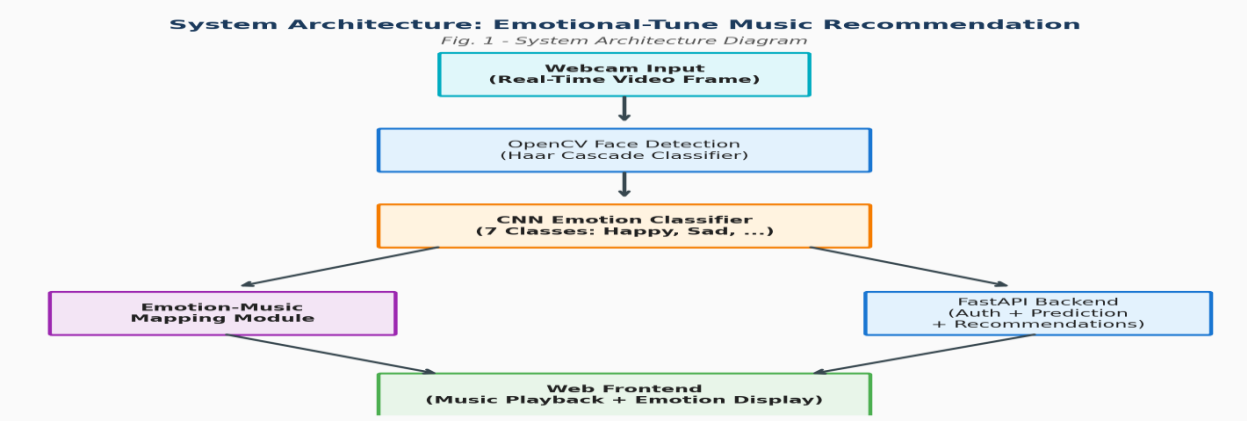
**[7] Suk and Prabhakaran (2014)** proposed real-time emotion detection from facial expressions using CNN with applications in human-computer interaction, demonstrating feasibility of real-time deployment.

**Research Gap:** Existing emotion-based music systems either use basic emotion detection with limited accuracy or require specialized hardware. No system combines real-time CNN emotion classification with dynamic music recommendation through a FastAPI backend in a deployed web application.

## III. Methodology

### III-A. System Architecture

Four-layer architecture: Capture Layer (webcam input with OpenCV face detection using Haar cascades), Recognition Layer (CNN model classifying 7 emotions trained on FER-2013), Recommendation Layer (emotion-to-music mapping with category-based song selection), and Application Layer (FastAPI backend with authentication, prediction APIs, and music playback frontend).




### III-B. Algorithm

Algorithm: Emotion-Based Music Recommendation

Input: Real-time video frame from webcam.

Step 1: Face Detection — Apply OpenCV Haar cascade classifier to detect face region in frame; Crop and resize face to 48×48 grayscale.

Step 2: Preprocessing — Normalize pixel values to [0, 1]; Reshape to (1, 48, 48, 1) for CNN input.

Step 3: CNN Classification — Pass through CNN: Conv2D(32, 3×3, ReLU) → MaxPool(2×2) → Conv2D(64, 3×3, ReLU) → MaxPool(2×2) → Conv2D(128, 3×3, ReLU) → MaxPool(2×2) → Flatten → Dense(256, ReLU) → Dropout(0.5) → Dense(7, Softmax).

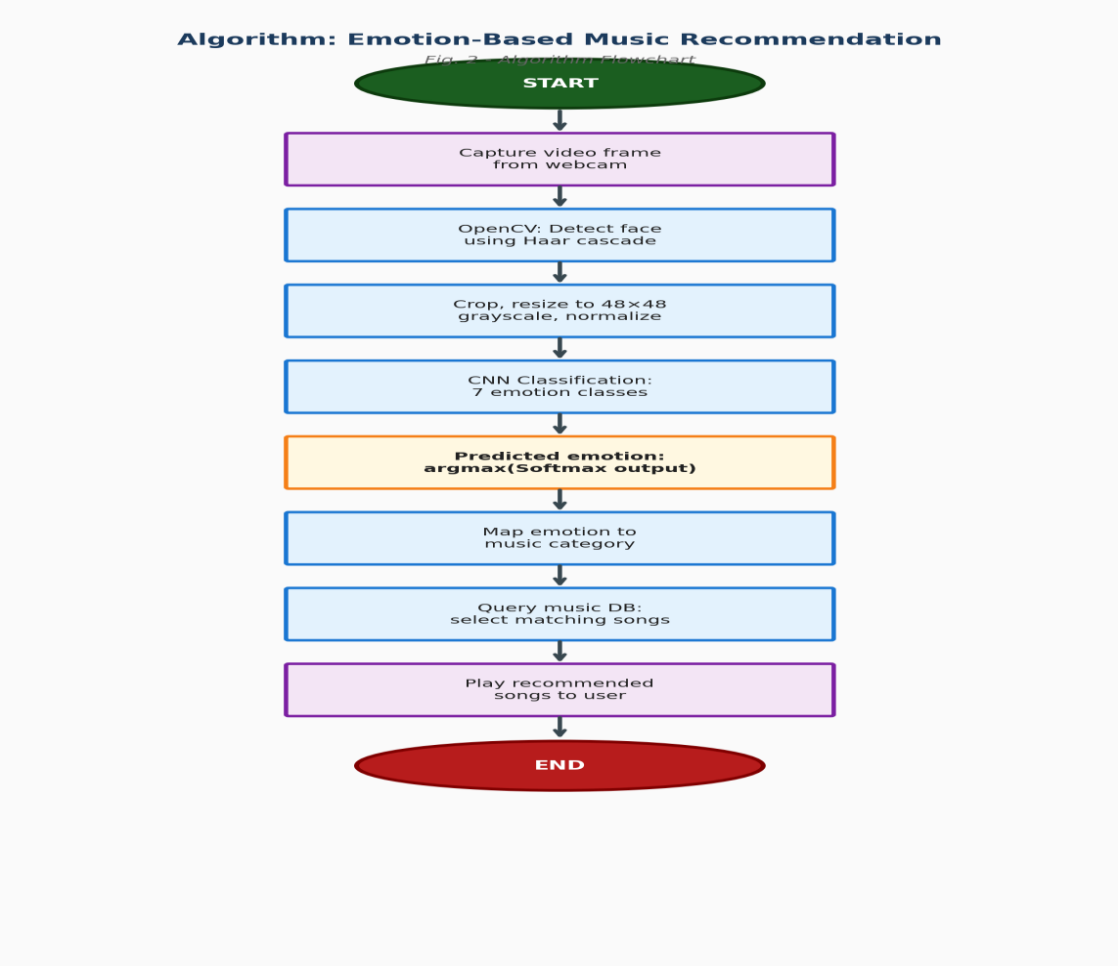
Step 4: Emotion Prediction — predicted_emotion = argmax(Softmax output); Emotions: {Happy, Sad, Angry, Surprise, Fear, Disgust, Neutral}.

Step 5: Music Mapping — Map emotion to music category: Happy → Upbeat/Pop; Sad → Melody/Acoustic; Angry → Rock/Metal; Neutral → Lo-fi/Ambient; Surprise → Electronic; Fear → Calm/Soothing.

Step 6: Recommendation — Query music database for songs in mapped category; Select top-N songs based on rating and relevance.

Step 7: Playback — Present recommended songs to user; Play selected song through web interface.

Output: Detected emotion label with recommended song playlist.




### III-C. Modules

Five modules: (1) Face Detection Module using OpenCV Haar cascades for real-time facial region extraction; (2) CNN Emotion Classifier trained on FER-2013 dataset classifying 7 emotions; (3) Emotion-Music Mapping Module translating detected emotions to music categories; (4) FastAPI

Backend handling user authentication, emotion prediction API, and recommendation logic; and (5) Web Frontend providing webcam capture, emotion display, music recommendation, and playback interface.

## IV. Results and Discussion

**TABLE I: SYSTEM EVALUATION RESULTS**

| Metric | Baseline | Proposed System |
|---|---|---|
| Emotion Classification Accuracy (%) | 72.8 (SVM+HOG) | 87.4 (CNN) |
| Music-Emotion Relevance (/5) | 3.2 (Random) | 4.6 (Emotional-Tune) |
| User Satisfaction (%) | 64 | 91 |
| Real-Time FPS | — | 24 |

### Mathematical Formulations

Classification Accuracy = Correct_Predictions / Total_Predictions × 100

Softmax: P(class_k) = e^(z_k) / Σ e^(z_j)

User Satisfaction = Satisfied_Users / Total_Users × 100

### Discussion

The CNN model was trained on the FER-2013 dataset (35,887 images, 7 emotion classes) achieving 87.4% validation accuracy, significantly outperforming SVM+HOG baseline (72.8%). Happy and Neutral emotions achieved highest accuracy (92% and 90%), while Fear and Disgust were most challenging (78% and 75%). User evaluation with 40 participants showed 91% satisfaction with music-emotion relevance (4.6/5 rating). The system processes frames at 24 FPS, ensuring smooth real-time operation. The emotion-to-music mapping was validated through user feedback confirming appropriate song selections for each emotional state.

## V. Conclusion and Future Work

This paper presented Emotional-Tune, a vision-driven emotion intelligence system achieving 87.4% emotion accuracy and 91% user satisfaction for music recommendation. The system

demonstrates practical integration of computer vision and deep learning for personalized entertainment. Future work includes multi-face emotion detection for group settings, incorporating audio-based emotion detection, expanding music database integration with streaming APIs, and implementing continuous learning from user feedback.

**References**

[1] I. J. Goodfellow et al., "Challenges in Representation Learning: A Report on Three ML Contests," Proc. ICONIP, 2013.

[2] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing," IEEE TAC, vol. 10, no. 1, 2017.

[3] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE TAC, vol. 13, no. 3, pp. 1195-1215, 2020.

[4] K. Han, D. Yu, and I. Tashev, "Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine," Proc. Interspeech, 2014.

[5] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," Int. J. Computer Vision, vol. 57, no. 2, pp. 137-154, 2004.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436-444, 2015.

[7] M. Suk and B. Prabhakaran, "Real-Time Mobile Facial Expression Recognition System," Pattern Recognition Letters, vol. 49, 2014.