Available online at
www.mdl.mazedan.com

MCET

www.mazedan.com
©2024 Mazedan International Research Academy
www.mazedan.com/mcet

MAZEDAN COMPUTER ENGINEERING TRANSACTIONS
e-ISSN: 2583-0414
Article id: MCET0601004
Vol.-6, Issue-1
Received:14 MAY 2025
Revised: 30 MAY 2025
Accepted:15 JUNE 2025

MULTIMODAL EMOTION DETECTION

PRAJAKTA. KURHE*, ASHWINI SANAP, GAYATRI GADE, ANKITA BATWAL, & VARSHA DEOKAR

DOI: https://doi.org/10.5281/zenodo.15915660

Citation: Kurhe, P., Sanap, A., Gade, G., Batwal, A. & Deokar, V. (2025). Multimodal Emotion Detection. *Mazedan Computer Engineering Transactions*, 6(1), 13–15.

Abstract

As we know, many people suffer from unhappiness in today's world, so there is a requirement for a system that recommends music based on human emotions such as anger, sadness, happiness, etc. In this paper, we propose a system named Multimodal Emotion Detection which helps people by suggesting music and movies based on their emotions. We implemented this system using the FER2013 dataset, which contains 35,887 images of different emotions such as sadness, happiness, neutral, surprise, disgusted, fearful, and anger. We used Python's TensorFlow framework and Haar Cascade algorithms to recommend movies and music based on human facial emotions. The main concern in existing recommendation systems is manual sorting. To avoid this, we propose this model, which automatically plays music and movies without requiring much browsing time.

Keywords— Image processing, Facial recognition, Movie Recommendation, Song Recommendation.

1. INTRODUCTION

This project is designed to recognize a person's emotions just by looking at their face. It works by capturing an image of the user, then using image processing techniques to analyze facial features and understand the emotion being shown. Once the emotion is detected, the system responds by playing songs that suit the user's mood and recommending movies they might enjoy all without the user needing to search manually. The main goal is simple: to lift the user's mood and make their day a little better, just by recognizing how they feel.

Problem Statement

Develop an intelligent system that can recognize facial expressions from real-time image input and, based on the detected emotion, play appropriate music tracks and recommend suitable movies. The system should be able to

- Accurately Detect Facial Expressions.
- Map Emotions to Media Content.
- Fetch and Play Media Content.
- Provide a User-Friendly Interface

2. PROPOSED SYSTEM

The automatic analysis and understanding of music by the computer is the new possibility in the field of music information retrieve First we login in the system. In capture phase we capture the face or image in the camera. Here we use a camera for face detection. The main objective of face detection technique is to identify the face which is happy or angry. In the data storage phase, previously stored images representing different emotions (happy, sad, angry) are compared with new images from the capture phase. Based on this comparison, the system automatically plays songs that match the detected emotion, and the results are displayed in the display phase.

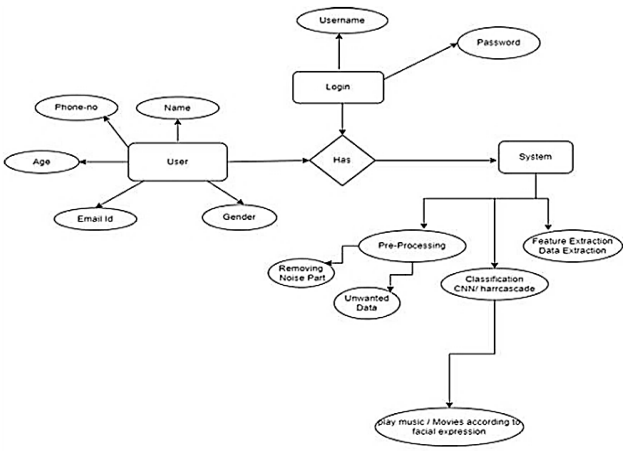

Figure 1 Working Flow

To store and manage user information effectively, an SQLite database was designed and implemented using DB Browser for SQLite. The database includes a table named registration, which captures essential details of each user interacting with the system. The fields in the registration table are as follows

- Full name - The full name of the user
- Address - User's residential location

SND College of Engineering and Research Centre, Yeola, Maharashtra, India,
*Corresponding author email-kurheprajakta.coe@gmail.com

- Username - A unique name chosen by the user for login identification
- Email - Email address for user communication
- Phone No - Mobile number of the user
- Gender - Gender represented numerically (e.g., 1 for male, 2 for female)
- Age - User's age in years
- Password - A secure password used for user authentication

This table structure allows efficient storage, retrieval, and authentication of user data within the system. By linking user information with detected emotions and recommendations, the database serves as the backbone of the system's personalized media suggestion logic. The use of SQLite ensures a lightweight, standalone, and portable solution that is easy to integrate with the Python backend and Tkinter GUI frontend.

Recognition of faces

Face recognition is a key component of the emotion-based recommendation system, using OpenCV for real-time face detection. The process involves capturing an image via the device's camera, detecting the face using techniques like the Haar Cascade Classifier, and preprocessing the facial region (grayscale conversion, resizing, normalization). The processed image is then fed into a CNN, which classifies the expression into emotions like happy, sad, angry, or surprised. This emotional insight is used to deliver personalized movie or music recommendations.

Feature Extraction

Feature extraction in emotion detection involves identifying key facial features like eyes, eyebrows, mouth, and contours after face detection using OpenCV. A Convolutional Neural Network (CNN) automatically learns and extracts these features as numerical patterns, reducing image complexity and enhancing classification accuracy and speed. Deep learning techniques make the system robust to changes in lighting, facial orientation, and expressions.

Emotion Detection

Emotion detection is the system's core function, using a CNN trained on datasets like FER2013 or Affect Net to classify facial expressions into emotions such as happy, sad, angry, and more. After feature extraction, the model analyzes facial muscle changes to detect emotions in real time, enabling personalized recommendations. Accuracy is improved through ongoing training, dataset preprocessing, and reliable image handling with OpenCV.

Recommendations

After detecting the user's emotion, the system provides personalized music and movie recommendations based on their emotional state. It uses content-based and collaborative filtering to match media to the user's mood, such as soothing content for sadness or upbeat options for happiness. Recommendations are managed using an SQLite database and displayed via a Tkinter GUI, enhancing the user experience with real-time, emotion-aware suggestions.

3. WORKING PRINCIPLES AND ALGORITHMS

Input

User's facial image captured through webcam.

Output

Emotion is detected, and personalized movie and music recommendations are displayed.

Problem Description: Identifying the emotional state of the user through facial expressions and providing personalized entertainment (movie and music) recommendations based on the detected emotion.

The system operates by capturing the user's face through a webcam and analyzing their facial expressions using a deep learning model. Based on the detected emotion, relevant movies and music are recommended to enhance the user's mood or experience. All data is managed and stored using an SQLite database, and the GUI is handled via Tkinter.

- Start
- Capture the user's facial image through the Camera
- Detect the face using OpenCV's Haar Cascade face detector.
- Preprocess the face image by converting it to grayscale and resizing it.
- Pass the preprocessed image to the trained CNN model to classify the emotion (e.g., happy, sad, angry, neutral, etc.).
- Store the detected emotion along with a timestamp and user data into the SQLite database.
- Fetch personalized movie and music recommendations from the database based on the detected emotion.
- Display the emotion and recommended content using the Tkinter interface.
- Repeat the process for the next user or end the session.

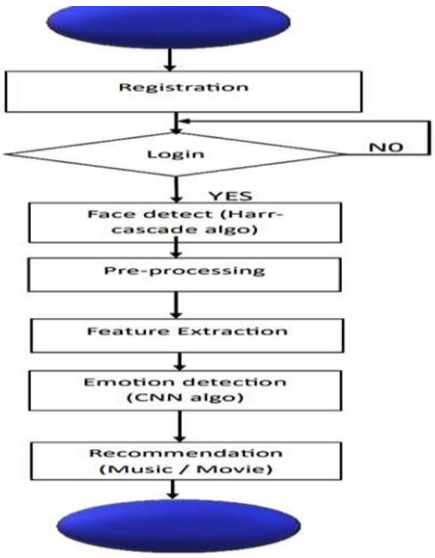

Figure 2 Flow Chart

Algorithms

Haar Cascade Algorithm

Harr-Cascade algorithm is use for real time face detection and Features extraction. It works by looking for patterns that are common in faces, like the way the area around the eyes or mouth shape. This is useful for emotion detection.

Convolutional Neural Network (CNN) Algorithm: A Convolutional Neural Network

(CNN)is a type of network used in deep learning, mainly for recognize faces and detect emotions by analyzing facial features. CNN is used for the training the dataset. CNNs are used for image classification and recognition because of its high accuracy.

The architecture of a CNN involves multiple layers

- Sequential Layer: A Sequential layer is a way to build a model step by step, one layer after another. It runs layers in the order you add them.
- Convolutional Layer: which are particularly effective for processing data with a grid-like structure, such as images.
- Max Pooling Layer: A max pooling layer is a fundamental building block in Convolutional Neural Networks (CNNs). Its primary function is to down sample the spatial dimensions (width and height) of the input feature maps while retaining the most important information.
- Dropout Layer: Which is used to convert images of 2D pixel into 1D vector.
- Flatten Layer: The primary role of a flatten layer is to take a multi-dimensional input tensor and reshape it into a one-dimensional tensor. This is done by taking all the elements of the input tensor and arranging them into a single vector.
- Dense Layer: A dense layer takes all the input and connects it to all the outputs. It helps the computer learn by adjusting the strength of these connections.

Advantages

- *Real-time Emotion Detection*: The system can recognize human emotions instantly through facial expressions using a camera, enabling real-time interaction.
- *Personalized Recommendations*: Based on the user's detected emotion, the system suggests appropriate music and movies, enhancing user satisfaction and emotional engagement.
- *Non-Invasive Interaction*: Users do not need to manually input their mood or preferences the system automatically identifies emotions, making it user-friendly.
- *Wide Application Scope*: This system can be applied in entertainment, therapy, education, and smart devices to improve human-computer interaction.
- *Emotion Aware Technology*: Integrating emotion detection with recommendation systems makes the platform intelligent, adaptive, and user-centric.

4. CONCLUSION

This system uses facial expressions to choose and play music that matches your mood. It does this by analyzing your face with a special network (CNN), identifying your emotion, and then selecting the right type of music. The system captures images from your device's camera to recognize your expressions, making it a personalized experience for each user.

REFERENCES

[1] Li, Huihui. *Research on facial expression recognition based on cognitive machine learning* [D]. Guangzhou: South China University of Technology, 2019.

[2] Zou, Jianchen, and Deng, Hao. *An automatic facial expression recognition method based on convolutional neural network* [J]. Journal of North China University of Technology, 2019, 31(5): 51-56.

[3] Yao, L. S., Xu, G. M., and Zhap, F. *Facial Expression Recognition Based on CNN Local Feature Fusion* [J]. Laser and Optoelectronics Progress, 2020, 57(03): 032501.

[4] Chauhan, Shavak, Mangrola, Rajdeep, and Viji, D. *Analysis of Intelligent Movie Recommender System from Facial Expression*. In: 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021.