Live Emotion Based Music Recommender System

Pooja M
Dept. of AI & ML
RNSIT
Bengaluru, Karnataka, India
pooja.m@rnsit.ac.in

Kavyashree H L
Dept. of AI & ML
RNSIT
Bengaluru, Karnataka, India
kavyashree.hl@rnsit.ac.in

Pratham M
Dept. of Computer Science
Christ University
Bengaluru, Karnataka, India
pratham.m@mca.christuniversity.in

Meghana G
Dept. of AI & ML
RNSIT
Bengaluru, Karnataka, India
1rn21ai072.meghanag@rnsit.ac.in

***Abstract*—The Live Emotion Based Music Recommendation improves the user experience by evaluating facial expressions and recommending songs that correspond with the identified mood. This study use deep learning approaches to identify emotions in real time and incorporates a convolutional neural network (CNN) for face emotion classification. Video frames are captured, face characteristics are processed, and identified emotions are mapped to carefully selected Spotify playlists by the algorithm. A web interface built using Flask shows real-time emotion recognition and dynamically modifies song suggestions.**

***Index Terms*—Emotion Recognition, Music Recommendation, Convolutional Neural Network (CNN), Deep Learning, Real-Time Processing**

I. Introduction

Music is a great instrument for improving mood and psychological well-being since it has a significant impact on human emotions. Because they rely on user preferences, history, and collaborative filtering processes, traditional music recommendation systems usually fail to capture a listener's emotional state in real time. Emotion-based music recommendation systems have surfaced to overcome this constraint, utilizing developments in computer vision, deep learning, and artificial intelligence. By combining facial expression detection with an automated song selection process, this study offers a fresh method to real-time emotion-based music recommendation.

The suggested method employs a convolutional neural network (CNN) model trained on the FER-13 dataset to reliably detect emotions from facial expressions. The device uses a camera to capture live video feeds and analyzes frames to detect emotions such as happiness, sadness, rage, and surprise. The system utilizes the Spotify API to obtain a personalized playlist based on the identified emotion in order to ensure that song selections align with the user's mood. The program is designed as a web-based platform that combines Flask for backend operations and OpenCV for image processing in order to offer a seamless and interesting user experience.

Issues involving dataset imbalances, computing overhead, and real-time processing limitations on edge devices continue to exist despite their effectiveness. This study aims to optimize the system's performance by utilizing efficient deep learning models and real-time feedback systems. By bridging the gap between emotion recognition and music selection, this work enhances human-computer interaction and produces a recommendation system that is more intuitive and emotionally flexible.

In addition, by incorporating deep learning for facial expression detection, the system is better able to identify minute changes in emotions, which improves the accuracy and personalization of the recommendation process. By using the FER-13 dataset, which offers a wide range of facial expressions, the model's ability to make predictions between users is enhanced. The approach also guarantees a dynamic and constantly updated music collection by utilizing the Spotify API, in contrast to conventional static databases. Numerous real-world applications, such as stress management, mental health assistance, and personalized user experiences in entertainment platforms, can benefit from the suggested methodology. In order to create a more robust and adaptable music recommendation system, future advancements will focus on expanding the dataset's emotional diversity, reducing computational costs, and enhancing model efficiency.

II. Literature Survey

Pushpavalli Mohan and Hsien-Tsung Chang's research, ”Deep Learning Based on Face Emotion Recognition using an Artificial Neural Network,” explores the use of deep learning for facial emotion recognition. It draws attention to earlier techniques that relied on static images, which reduced accuracy. By incorporating dynamic datasets like CK+ and FER, the study enhances real-time detection. Models such as MobileNet and ANN are evaluated for recognizing seven emotions, emphasizing deep learning's role in improving accuracy for applications in security and healthcare.

The paper “Music Recommendation System Using Facial Detection-Based Emotion Analysis” by Manvitha Sri Guthula

979-83315-7746-9/26$31.002026 © IEEE

and Dr. Monali Bordoloi explores facial emotion recognition for music recommendation. It examines earlier studies on music emotion recognition (MER) and facial emotion recognition (FER), emphasizing the accuracy of deep learning techniques like CNNs. Studies on Viola-Jones for face detection and ANNs for emotion classification are discussed, along with research on music emotion recognition techniques. These works provide the foundation for developing emotion-aware music recommendation systems.

The paper "Emotion-Based Music Recommendation System" by Vijay Prakash Sharma and others explores CNN-based facial emotion recognition for music recommendations. It reviews existing techniques like ANNs, Markov Decision Processes, and Reinforcement Learning. Studies on audio-visual emotion recognition and wearable sensors highlight alternative approaches. The research addresses gaps in real-time facial expression-based music recommendation, proposing a deep learning-driven system for automated mood-based playlist generation.

The paper "A Brand-New CNN-Based Emotion-Based Music Suggestion System" by Tejaswini Priyanka V & Others presents a novel way to improve user experiences. It uses a Convolutional Neural Network (CNN) to recognize faces and identify user emotions in real time from camera broadcasts. By using OpenCV to extract facial features, emotional states such as surprise, anger, sadness, and happiness can be accurately classified. Customized playlists based on the user's mood are connected to the identified emotions.

III. Research Methodology

The systematic method of conducting research, which includes the concepts, plans, and methods for collecting, analyzing, and understanding data in order to meet the goals of the study, is known as research methodology. The methods are as follows.

A. System Overview

The proposed Emotion-Based Music Recommendation System seeks to improve user experience by using facial expressions to identify emotions and suggesting appropriate music in response. The system uses a structured pipeline that combines deep learning and computer vision to provide tailored song recommendations based on the real-time detection of facial emotions.

The system divides emotions into seven groups using a deep learning model that was built using the FER-2013 dataset. It incorporates the Spotipy API to retrieve Spotify song suggestions, OpenCV for real-time video capture, and Flask for the web-based interface. By using multithreading, the technology guarantees an uninterrupted experience by enabling lag-free simultaneous video processing and music retrieval.

A webcam is used to record live video, Haarcascade is used to identify faces, and the extracted facial area is then fed into a CNN-based model. After the model predicts the emotion, the Spotipy wrapper is used to translate it to a playlist genre or

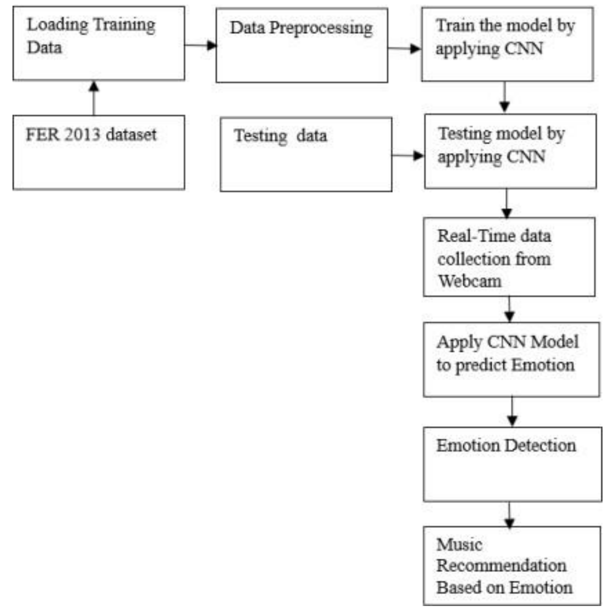

Fig. 1. Flow diagram

a collection of suggested songs that are dynamically retrieved from the Spotify API.

The model architecture follows a sequential CNN structure, incorporating Conv2D layers with varying filter sizes, MaxPooling layers for feature extraction, and Dropout layers to reduce overfitting. The final dense layer employs a softmax activation function to classify emotions. The training pipeline includes image normalization, grayscale conversion, and batch processing using ImageDataGenerator.

Multithreading is used to improve efficiency and provide real-time emotion detection. By providing distinct threads for frame processing, music searching, and video streaming, the system significantly improves responsiveness.

The web interface, built using HTML and CSS offers a visually appealing and intuitive user experience. Users can see their detected emotion in real time and receive dynamically updated song recommendations. The system runs efficiently on local devices but can be optimized further for cloud deployment and cross-device accessibility.

IV. Implementation

A. Tools and Languages

We have used Python, HTML, CSS programming languages. The frameworks and libraries include TensorFlow, Keras, OpenCV, Flask, and Spotipy. For model training and experimentation, use Google Colab, Jupyter Notebook, or Visual Studio code.

B. Collection of Datasets

The Data plays a crucial role in building an effective emotion-based music recommendation system. The FER 2013 dataset was selected as the main source for training. The dataset, containing 35,887 grayscale images categorized into seven emotions, obtained from Kaggle repository and other online sources. A deep learning model for emotion recognition was trained using this dataset and subsequently included into the music recommendation system.

Fig. 2. FER 2013 Dataset Example

C. Data Pre-processing

Data preprocessing plays a important role in developing an emotion-based music recommendation system by ensuring that the input data is clean, structured, and optimized for model training. This process involves handling both facial expression images from the FER 2013 dataset and song metadata retrieved from the Spotify API.

For facial expression analysis, the images were first converted to grayscale and resized to 48x48 pixels to maintain uniformity and reduce computational complexity. To enhance model performance, pixel values were normalized to fall within a range of 0 to 1. Data augmentation techniques such as rotation, flipping, and zooming were applied to artificially expand the dataset and improve generalization. OpenCV's Haar cascade classifier was used for real-time face detection, ensuring that only relevant facial features were processed for emotion classification.

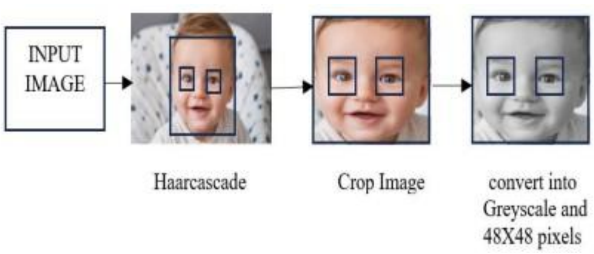

Fig. 3. Image Conversion

D. Training the Model using CNN

A Convolutional Neural Network (CNN) model is trained using the processed dataset with the goal of classifying facial expressions into seven emotion groups. The model adopts a sequential architecture, starting with a Conv2D layer consisting of 32 filters with a 3×3 kernel, using the ReLU activation function. Additional convolutional layers with 64 and 128 filters further refine spatial feature extraction. Max-pooling layers with a 2×2 pool size are included to reduce dimensionality while maintaining key image features. Dropout layers are incorporated to prevent overfitting, improving the model's generalization. Before classification, the network is flattened and passed through a dense layer of 1024 neurons, followed by another dropout layer. The final softmax activation function categorizes images into seven distinct emotions. The model is compiled using categorical cross-entropy as the loss function and Adam optimizer. Multiple training epochs improve model performance, and validation data is used to monitor accuracy.

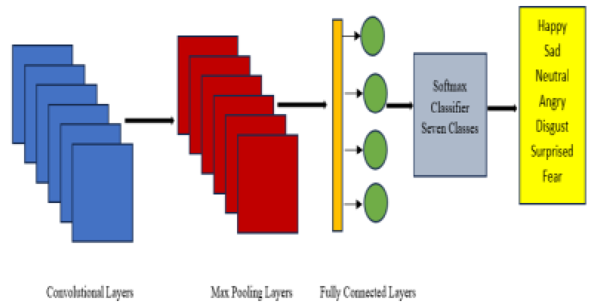

Fig. 4. Convolutional Neural Network

Formulas:

1. Softmax Activation Function:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{1}$$

where:

- z_i is the input for the i-th class.
- n is the number of classes.
- e^{z_i} ensures all values are positive and sum to 1.

2. ReLU (Rectified Linear Unit) Activation Function:

$$f(x) = \max(0, x) \tag{2}$$

- If $x > 0$, the output is x.
- If $x \leq 0$, the output is 0.

E. Testing the Model

The pre-trained model is loaded, and test images are processed in the same way as the training data to maintain uniformity in dimensions and normalization. These test images are then fed into the model, which predicts emotion labels. Metrics like precision, recall, F1-score, and a confusion matrix are used to assess performance and measure accuracy by comparing the expected outputs with the actual labels. This stage guards against overfitting and guarantees the model's effective generalization. where:

- TP – True Positive
- FP – False Positive
- TN – True Negative
- FN – False Negative

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

F. Webcam-Based Real-Time Data Collection

For real-time data collection, live video frames are captured by a webcam, and the frames are processed using a convolutional neural network (CNN) to identify emotions. To enhance feature extraction, each frame of the camera's continuously streaming video is converted to grayscale. A Haar Cascade classifier recognizes faces in the image. The extracted face is resized to 48x48 pixels in order to be compatible with the trained emotion categorization model.

The CNN model with pre-trained weights predicts the user's sentiment by classifying the input image into one of seven predefined emotion categories, such as angry, disgusted, afraid, happy, neutral, sad, and surprised.

To enhance efficiency, multithreading is implemented, ensuring smooth video streaming and real-time processing without lag. The FPS (Frames Per Second) module is used to optimize performance, allowing seamless facial recognition and emotion classification. The system captures approximately 15-20 FPS, ensuring real-time detection of facial expressions and accurate emotion classification.

G. Music Recommendation Based on Emotion

Once the user's emotion is detected using the webcam-based facial expression recognition system, the next step is to recommend music tailored to their current mood. The system integrates with Spotify's API to fetch curated playlists corresponding to different emotional states. A predefined music dictionary maps each detected emotion to a specific playlist ID on Spotify.

To authenticate and obtain songs from various playlists, the application makes use of the Spotipy library. While the getTrackFeatures function retrieves important information like the track name, album, and artist, the getTrackIDs function collects track IDs from a specified playlist. Pandas DataFrames is then used to store the recovered data in an organized fashion and CSV files are generated for every playlist.

When an emotion is identified, the relevant playlist is requested, and the top 15 suggested songs are shown for real-time recommendations. A customized and engaging experience is produced by this dynamic system, which makes sure that consumers are given music recommendations that correspond with their present emotional state. By effectively managing video streaming and API queries, multithreading enhances performance and enables lag-free, fluid interaction.

H. User Interface and Front-End Integration

The front-end of the emotion-based music recommendation system is designed using HTML and CSS creating an intuitive and responsive user interface. The page is titled "Live Emotion Music Based Recommender System" and follows a structured layout that divides the screen into two primary sections: 'The Emotion Detector' and Song Recommendations. The Emotion Detector section displays a real-time webcam feed that captures and processes facial expressions using OpenCV, streaming the video dynamically from the Flask server.

On the right, the Song Recommendations section dynamically updates based on the detected emotion. A function fetches real-time JSON data from the server every 100 milliseconds, extracting and displaying relevant song details such as track name, album, and artist in a structured table. Styling is achieved using Bootstrap 5 and custom CSS, ensuring a visually appealing and responsive interface. With everything considered, this interface effectively connects the Spotify API-based recommendation engine with the emotion detection pipeline, guaranteeing an intuitive interface for mood-based music selection in real time.

V. Working of CNN

The given Fig. 5 depicts the working of Convolutional Neural Network (CNN).

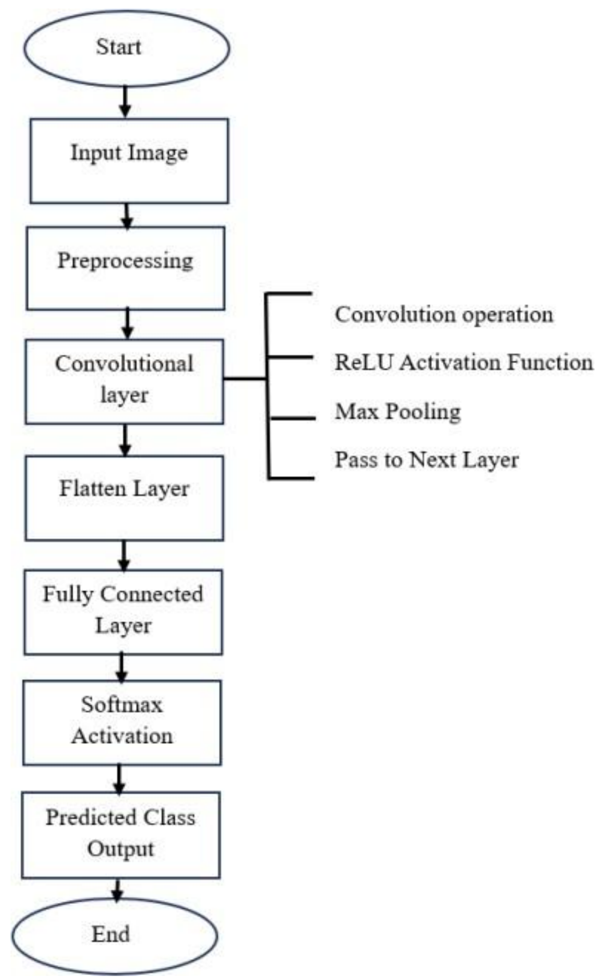

Fig. 5. Convolutional Neural Network

VI. Discussion of Results

The Emotion-Based Music Recommendation System successfully integrates facial emotion recognition with personalized music recommendations using deep learning and Spotify's API. The system captures real-time frames at an average rate of 15-20 frames per second and processes the detected emotions to curate music playlists corresponding to the user's emotional state.

Facial expressions may be accurately classified into seven categories by the CNN (Convolutional Neural Networks)-based deep learning model: angry, disgusted, afraid, happy, neutral, sad, and surprised. Lighting, face visibility, and user interaction all affect how accurately emotions can be detected. It achieves an average accuracy of 86.53% in optimal conditions.

VII. Conclusion and Future Work

The Live Emotion Based Music Suggestion System effectively combines an intelligent music recommendation algorithm with facial expression identification. It determines the user's emotional state through real-time image evaluation and recommends music that fits their mood. By providing a seamless and entertaining way of enjoying music, the combination of deep learning for emotion recognition and an automatic song selection system improves the user experience. This research demonstrates how AI-driven solutions may be used to understand human emotions and modify things accordingly, making them a helpful tool for mental health and happiness.

Future developments can concentrate on enhancing the emotion detection model's efficiency and accuracy by integrating cutting-edge deep learning architectures, including transformer-based models for improved face feature extraction.

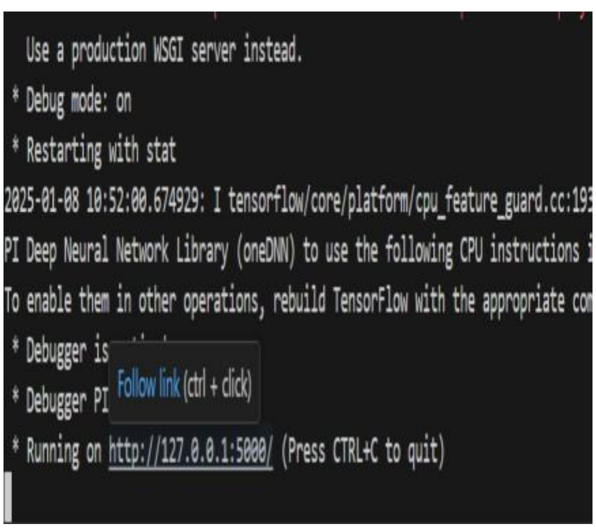

Fig. 6. Output

The Fig. 6 shows the terminal output of running a Flask-based web application for the Emotion-Based Music Recommendation System. Because the application runs in debug mode, code changes can be made in real time without the server having to be restarted. "Running on http://127.0.0.1:5000/" is a message that shows that the Flask server is successfully hosted on the local computer and can be accessed using the supplied URL. The debugger, which provides detailed logs and traceback information is enabled to assist in error detection. Users can interact with the running application by clicking on http://127.0.0.1:5000/ in their web browser. Pressing CTRL + C at the terminal will end the process if needed. Once the web page is opened, users can access the emotion detection module, which captures facial expressions using a webcam.

Fig. 7. Output

The Fig. 7 represents the output of the Live Emotion-Based Music Recommender System. The interface consists of two main sections: the Emotion Detector on the left and Song Recommendations on the right. The system has successfully identified the user's emotion as "Happy," as indicated by the label above the detected face within the green rectangular box. Based on this detected emotion, the system generates a curated list of songs associated with happiness, displaying relevant details such as Name, Album, and Artist. User-friendly navigation is ensured by the neat and organized user interface, which makes interacting with the recommendation system easy.

Furthermore, the system's capacity to identify emotions can be enhanced by including a multi-modal approach, such as voice and physiological data analysis. Accessibility and scalability can also be improved by implementing a mobile-friendly version and cloud-based deployment, which will make the system more useful for real-world applications.

VIII. Acknowledgment

We are grateful for entire Artificial Intelligence and Machine Learning department for their help and guidance.

References

[1] Pushpavalli Mohan and Hsien-Tsung Chang (2023). Deep Learning Based on Face Emotion Recognition using an Artificial Neural Network. 2023 International Conference on Artificial Intelligence and Power Engineering (AIPE).

[2] Manvitha Sri Guthula and Dr. Monali Bordoloi (2024). Music Recommendation System using Facial Detection based Emotion Analysis. 2024 International Conference on Emerging Techniques in Computational Intelligence (ICETCI).

[3] Vijay Prakash Sharma, Gaded, A. S., Chaudhary, D., Kumar, S., & Sharma, S. (2021). Emotion-Based Music Recommendation System. 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), Amity University, Noida, India, Sep 3-4, 2021.

[4] Tejaswini Priyanka V, Ramesh, G., Reshma Reddy, Y., & Gomathy, S. (2023). A Novel Emotion based Music Recommendation System using CNN. In Proceedings of the 7th International Conference on Intelligent Computing and Control Systems (ICICCS-2023).

[5] Srivastava, D., Bhardwaj, A., Puri, S., Bhagat, S. D., & Verma, M. (2024). Emotify: An AI-Powered Emotion-Based Music Recommendation System. 2024 4th International Conference on Intelligent Technologies (CONIT), 979-8-3503-4990-0/24/$31.00.

[6] Chin, Y.-H., Lin, S.-H., Lin, C.-H., Siahaan, E., Frisky, A., & Wang, J.-C. (2014). Emotion Profile-Based Music Recommendation. In 7th International Conference on Ubi-Media Computing and Workshops, 1–5.

[7] Nagaraj, P., Yeswanth, G. S., Sudar, K. M., Charan, G., Krishna, G. V., & Charan, N. S. (2024). Relevant Musical Schema-Based Human Emotion Controller Using Deep Learning Techniques. 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS).

[8] Subbulakshmi, N., Chandru, R., Ariffa Begum, S., & Lakshmireddy, B. (2024). AI based Emotion Aware Music Composition using Spotify App. 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL), 979-8-3503-9615-7/24/$31.00.

[9] Tejashwini, S. G., and D. Aradhana. “Multimodal deep learning approach for real-time sentiment analysis in video ” International Journal of Advanced Computer Science and Applications 14, no. 8 (2023).