Xpressify: Facial Emotion Recognition and Music Recommendation System

Devi R.[1], Balasakthi[2], Harish[3], Maithireyan[4], Vendhan[5]

Department of Information Technology, Government College of Technology, Coimbatore, India.

E-mail: [1]r.devi@gct.ac.in, [2]bala.71772118105@gct.ac.in, [3]hari.71772118113@gct.ac.in, [4]mait.71772118126@gct.ac.in, [5]vendhanvelusamy@gmail.com

Abstract

Facial Emotion Detection and Music Recommendation System is a real-time facial emotion detection and personalized music recommendation system. It employs a trained Convolutional Neural Network (CNN) trained on the FER2013 database for seven emotion recognitions, i.e., anger, disgust, fear, happiness, sorrow, surprise, and neutrality. Everything is accomplished through Artificial Intelligence (AI) and Computer Vision technology. React and Flask are used to create the frontend and backend, while OpenCV is used to detect the facial emotion of the user through the webcam. The system does not pre-downloading playlists but dynamically uses the YouTube Data Application Programming Interface (API) to download appropriate music tracks depending on the emotional state of the user. Measures of precision, recall, accuracy (79.07%), and loss values are used to deploy the proposed system. The values indicate the viability of the application of Xpressify for use in real-time affect classification and customized music. With dynamic, interactive, and emotive-response music recommendations, the integration provides a more realistic experience for the user and is applicable in entertainment, mental health, and human-computer interaction.

Keywords: Facial Emotion Recognition, Music Recommendation System, Convolutional Neural Network (CNN), YouTube API Integration, Real-Time Emotion Detection.

1. Introduction

Designing inherently compelling digital experiences demands attention to human emotion and the performance of understanding. Emotion-sensitive systems operating in real time have become deeply integrated into well-being, intelligent environments, and individual entertainment during the past few decades. In these systems, emotion-adaptive music

249

recommender systems have the potential to have profoundly positive effects on emotional well-being and individual happiness.

One effort to bridge the gap left behind by conventional music recommendation systems those that functioned on the basis of static behavior data under normal circumstances but lacked the ability to explore a user's dynamic emotional status is Xpressify: Facial Emotion Recognition and Music Recommendation System. Traditional systems tend to make blanket or mismatched suggestions because they fail to react to dynamic user states or become overly sensitive to real-time emotional undertones.

This project proposes a combined platform of real-time AI music recommendation and live facial emotion recognition as the solution to these issues. It is built with an efficient, lightweight Convolutional Neural Network (CNN) trained on the FER2013 dataset, fine-tuned for strong emotion recognition from live webcam video. The CNN detects distinctive facial features corresponding to seven universal emotions: happiness, sadness, anger, surprise, fear, disgust, and neutrality. Model building is performed in a manner that allows for rapid inference to facilitate real-time revision of the recommendations based on the user's expression. The algorithm determines related songs by invoking the YouTube Data API when detecting an emotion. With a React-based interface and ReactPlayer-based playback within its toolkit, the integration removes previous playlists and widens the set of music available to more accurately reflect the user's current mood.

The backend is developed using Python Flask, which is data transfer efficient and cross-platform scalable. Data augmentation, resizing, normalizing, and grayscale conversion are among the several automated steps of data preprocessing done to make the CNN uniform under different lighting conditions and appearances of the user. By capturing the mood detected and offering real-time interaction with recommended music, the popular streaming-based user interface provides real-time feedback to the user.

Emotion recognition module and song recommendation module evaluation are the targets of evaluation for this project. Precision, recall, F1-measure, accuracy, and loss are some common parameters to evaluate system performance scientifically. Real-world environment robustness is guaranteed by other strict tests under varied illumination conditions, camera orientations, and face orientations.

In effect, Xpressify constructs an affective smart music recommender system by combining computer vision, deep learning, and large-scale web technology in a unique way. Since it is based on end-to-end architecture design, real-time computation, adaptive personalization, and large-scale experimentation, the system has no rival for real-world usage in encouraging human-computer interaction, psychological health, and enjoyment.

2.   Related Work

We can translate human feelings onto individual music experiences using the most recent developments in facial emotion detection systems and music suggestion systems. Previous studies grouped emotions like anger, fear, happiness, sadness, and surprise using Darwin's theory of universal facial expressions [1][2][3]. Building on that, current systems utilize deep learning algorithms to improve recognition accuracy as well as automatic music suggestion.

2.1 Facial Emotion Recognition

As deep learning-powered CNNs enable real-time emotion classification, they substitute traditional handcrafted feature extraction techniques [4][7]. FER2013-type training data on CNNs enable them to recognize emotions with or without occlusion and varying lighting conditions through eye movement and facial feature extraction, including mouth shape [6]. In order to enable emotion-adaptive recommendations, some models such as VGG16-based CNNs employ Spotify APIs for dynamically creating playlists [6]. Other methodologies employ MobileNet, which performs very little computation on facial emotion detection using webcams [9].

2.2 Music Recommendation Systems

Legacy recommenders are not effective in real-time and incorporate content-based methods and collaborative filtering [3][5]. Affective computing is integrated into newer technology so that the user's mood is also dynamically factored into music recommendations [5]. Personalization has been maximized by leveraging listening history-based, tempo-based, and energy-based playlist generators [4]. Hybrid approaches such as Emo Melodies generate emotion-sensing playlists based on a blend of collaborative filtering and content-based filtering [5].

2.3 Integrated Emotion-Based Music Recommendation

Automatic mood-based music recommendation systems have been established in some research by merging facial emotion recognition with music recommendation. Bakariya et al. [1] introduced CNN-based six emotion detection using a blend of Flask and OpenCV for real-time recommendation and classification over FER2013. Some other works involve real-time detection via webcam, in which real-time playlist creation is triggered as a result of emotion classification through an external API [2].

To facilitate feature extraction in dynamic environments, systems combine CNN architectures with conventional image processing [3]. For reasons of maximum computational efficiency, others use Tkinter and Pygame for local music playback through GUI [8]. According to consideration of musical feature and facial expression analysis based on rhythm, melody, and user feedback, individualization is facilitated by a two-track model of analysis [10].

2.4 Challenges and Research Gaps

Though notable advancements have been achieved, real-time emotion-aware music recommendation is a challenging task. Occlusions, dynamic facial expressions, and lighting create detection challenges [7]. Emotional mapping to music is also difficult due to the subjective nature of music preference [9]. Hybrid deep learning and API fusion is also utilized by researchers to address privacy issues related to facial data collection and latency for real-time processing [6].

3.   System Design and Implementation

Figure 1 shows the system architecture of the proposed system. This model is an end-to-end real-time face emotion detection and dynamic music recommendation system in a combined manner. The user interface, validation, emotion-music mapping, CNN-based emotion detection, music recommendation using the YouTube API, data pre-processing, and system integration are some of its dependent modules to be executed in a pipeline format.

3.1 Data Preprocessing

Data preparation is the very first practical step of the raw face image pre-processing approach for emotion recognition strength. FER2013 images are resized to 48×48 pixels via resizing and grayscale conversion for uniformity and light computational overhead. Pixel values are normalized to the [0,1] range via MinMax scaling to accommodate accurate model training. Data augmentation methods like rotation, zoom, brightness adjustment, and flipping around the x-axis are used to help the model generalize rather than overfit. Data is also divided into test, validation, and training subset dataset in an 80-10-10 ratio for easy testing of the model. At last, preprocessed images are reshaped into the format (samples, 48, 48, 1).
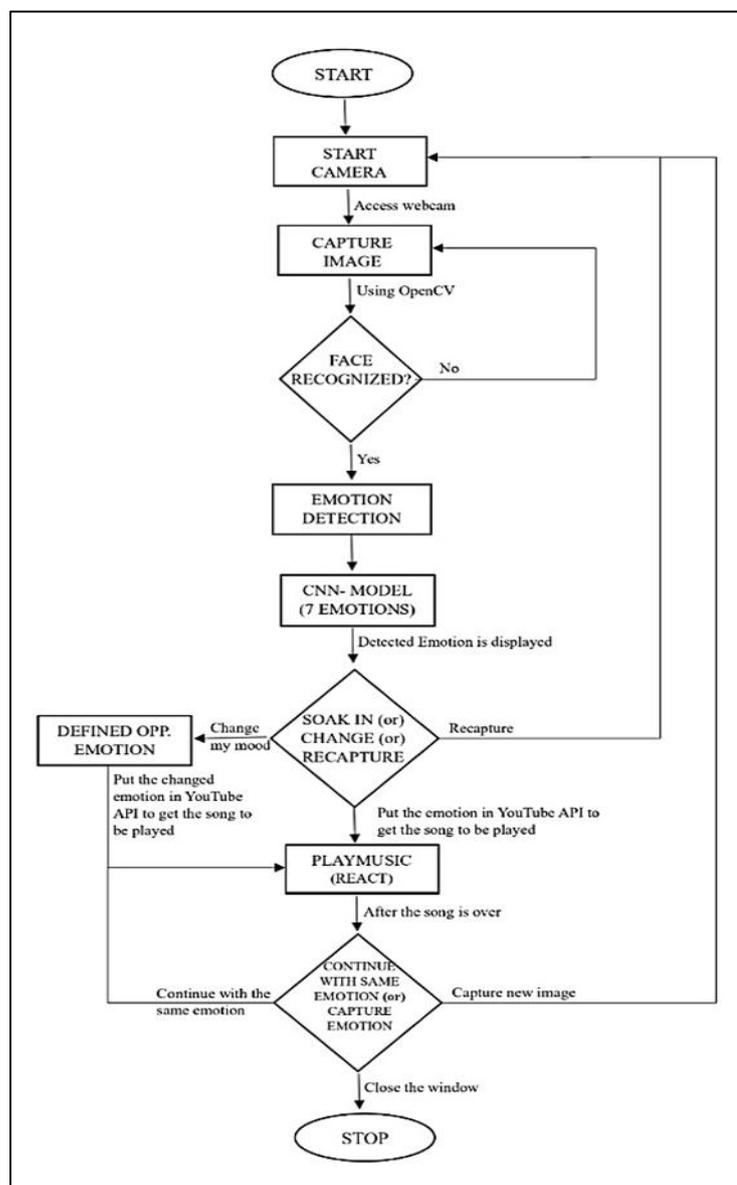

Figure 1. System Architecture

3.2 Convolutional Neural Network (CNN) for Emotion Detection

Our model is composed of two different CNN architectures: Model-A has fifteen layers, and Model-B has twelve layers. Model-A uses convolutional layers with 3×3 filters, ReLU activation, and 2×2 max pooling. These layers have 64, 128, 256, and 512 filters in each respective layer. Dropout (rate 0.3–0.5) and batch normalization are used to stabilize training and avoid overfitting. Model-B is the thinner model that incorporates fewer filters (32, 64, 128) and tries to propagate information at a higher speed but at the cost of reasonable levels of accuracy. The OAHEGA and FER2013 datasets were utilized to train the models in a manner that compared the performance between a deeper and a faster model and to identify which one is better in terms of performance for real-time emotion detection.
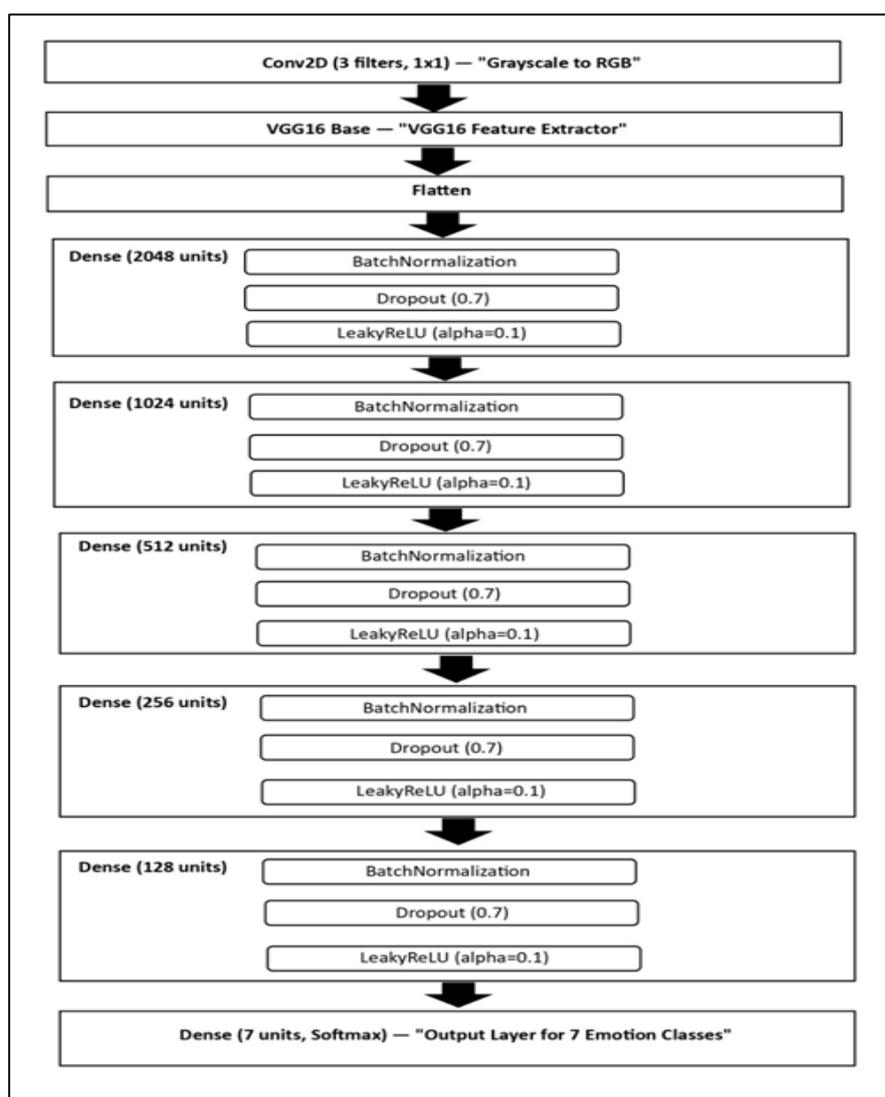

Figure 2. CNN Model

3.3 Emotion – Music Mapping

It is from this module that the emotions detected are transformed into suggested music. An emotion-to-music mapping governs the emotions of joy, sadness, and anger to their respective musical counterparts of energetic, calm, or strong songs, which is system-helpful. Mappings may be given by the user so that further personalization of language or artist is enabled, making the experience dynamic and personalized. The equivalent music genre is retrieved from the mapping and passed to the recommendation module for processing on an image mapped to an emotion category by CNN. The system is integrated with the YouTube Data API to promote dynamic and diverse music selection.

3.4 Music Recommendation using YouTube API

A dynamic search query, i.e., "happy upbeat songs," is constructed from the output emotion and the equivalent music genre related to it. This data is used to make a query to the YouTube API, which responds with matching music videos in real-time. Fallback suggestions are used in case of failure or no match. ReactPlayer is used to inline downloaded music videos within the app for easy playback of media inside the UI. Real-time and smooth user interaction is the responsibility of the UI design.

3.5 User Interface

The video from the live webcam feed is captured with the assistance of OpenCV, such as real-time face detection and utilization of face regions for emotion detection. The detected emotion is also mapped on the screen using the skip, pause, or play control available in the YouTube API playback. The previously detected emotions and their corresponding music suggestions are also updated. Feedback is given so that users can analyze how accurate the music suggestions were and how well the emotion recognition performed, allowing for enhancements in future predictions and suggestions from the system.

3.6 Validation Module

To ensure the system is reliable and accurate, a robust validation system is employed. Metrics like precision, accuracy, recall, F1-score, and confusion matrices are utilized to validate the CNN model and assess how well the model generalizes when handling unseen, new data. The emotion module and music suggestion module are continuously fine-tuned with

user feedback, and any misclassified examples or poor music suggestions are monitored for closer detailed analysis.

Besides, there are also robustness tests used in bug discovery and system performance optimization under various conditions.

It is the last step for the project to integrate all the modules into a working system.

3.6.1   Evaluation Metrics

To assess the performance of the CNN-based emotion recognition system, standard classification metrics were used, defined as follows:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{1}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{2}$$

$$Recall \quad = \frac{TP}{(TP+FN)} \tag{3}$$

$$F1\ Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

Where:

TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

3.7 System Integration

OpenCV recognizes faces in real time, and it supplies CNN with emotion classification data. After identifying the emotion, the system fetches the corresponding musical genre and issues a respective music video call to the YouTube API. The recommended songs are rendered using ReactPlayer, with interactivity presented by the user interface through the provision of playback controls, feedback, and detected emotions. Axios and CORS are used to enable communication between the back end and the front end in a manner that allows data to be shared between the React-based front end and the Flask-based back end without any problems.

4. Results

4.1 Data Preprocessing Output

The 35,887 records dataset FER2013 in three columns, underwent a critical preprocessing stage for deployment with the CNN model.

Pixel values were converted to number arrays, values were scaled, and the pictures were rescaled in a way that allowed for quick training of the model After pre-processing, the dataset was converted in the following manner:

- **Emotion Encoding:** To facilitate model interpretation, labels were converted to integer values (for example, 0 for anger, 1 for disgust, etc.).
- **Pixel Transformation:** In order to normalize the intensity values of pixels within the [0,1] range, they were removed, rearranged into 48×48 matrices, and then divided by 255.
- **Data Splitting:** To ensure a fair assessment of the model, the dataset was split into 80% training, 10% validation, and 10% testing.
- **CNN Reshaping:** Images were rearranged into a 4D array with dimensions (samples, width, height, and channels); for grayscale images, the channel dimension was set to 1.

4.2 CNN Model Output

The CNN model trained for 50 epochs, showing steady improvement:

Table 1. CNN Model Accuracy

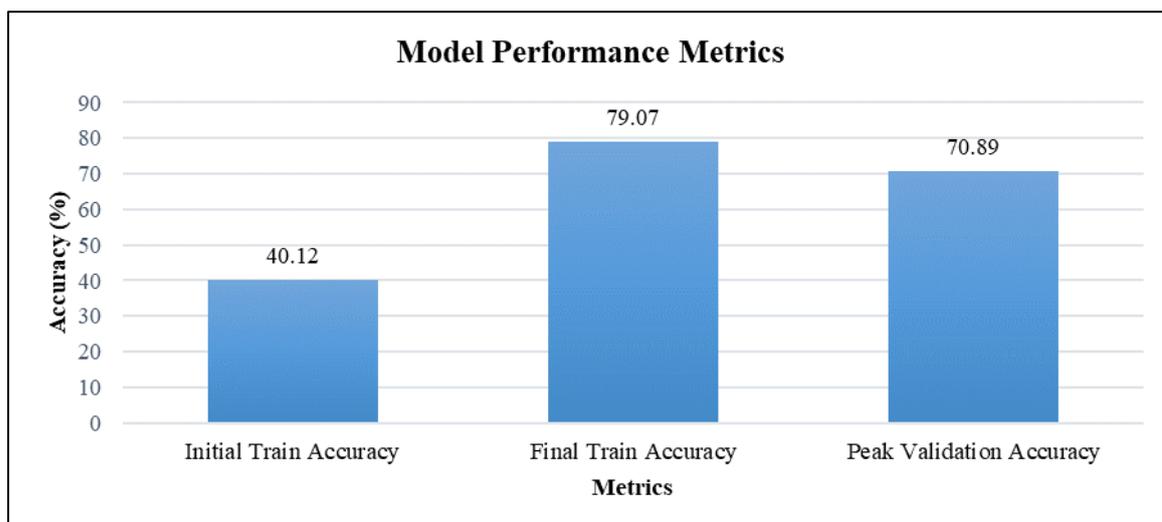| Accuracy	Percentage
Initial Training Accuracy	40.12% (Epoch 1)
Final Training Accuracy	79.07% (Epoch 50)
Peak Validation Accuracy	70.89%
Minimum Validation Loss	1.05

Figure 3. CNN Model Accuracy

Figure 3 shows an accuracy of 79.07% on the new test, with the model correctly classifying facial expressions into seven emotion categories (Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral). Although predictive of areas for improvement such as hyperparameter tuning and sophisticated data augmentation techniques, the results validate the potential of CNN in complicated emotion identification. Figure 4 illustrates the output of the CNN model.
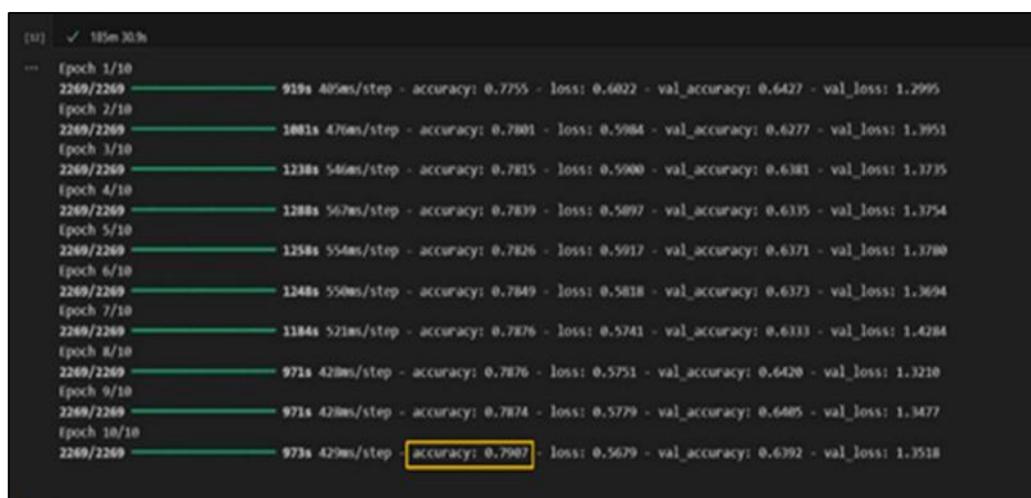

Figure 4. CNN Model Output

4.3 Final Output

Real-time personalized song recommendations are given by the module for music recommendation based on facial emotion recognition with the YouTube API. It identifies the

user's mood by implementing the CNN model and employs it to dynamically create search queries with mood keywords (e.g., "happy songs," "sad music").

The following steps make up the recommendation process:

- **Emotion Recognition:** The CNN determines the user's present emotional condition.
- **Query Generation:** The system uses the identified emotion to create a search query.
- **Request for YouTube API:** The query retrieves a list of pertinent videos.
- **Song Selection & Playback:** The most relevant video is embedded and played using ReactPlayer for song selection and playback.
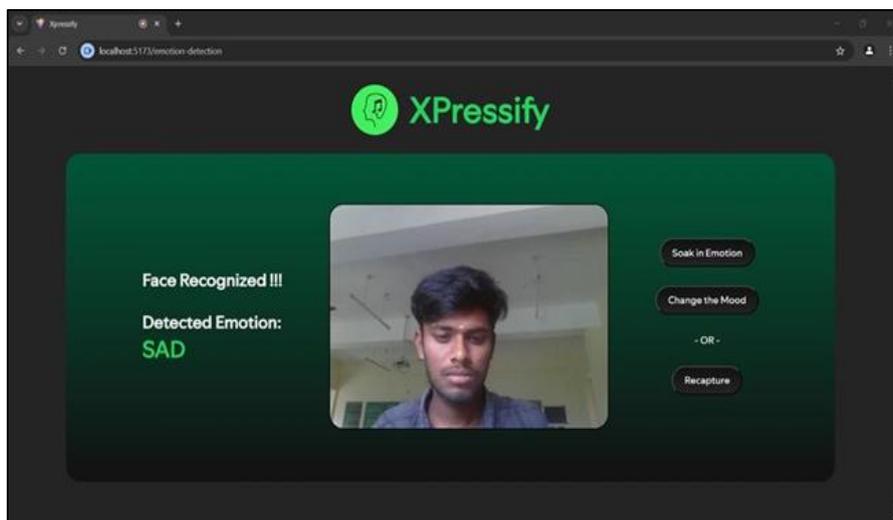

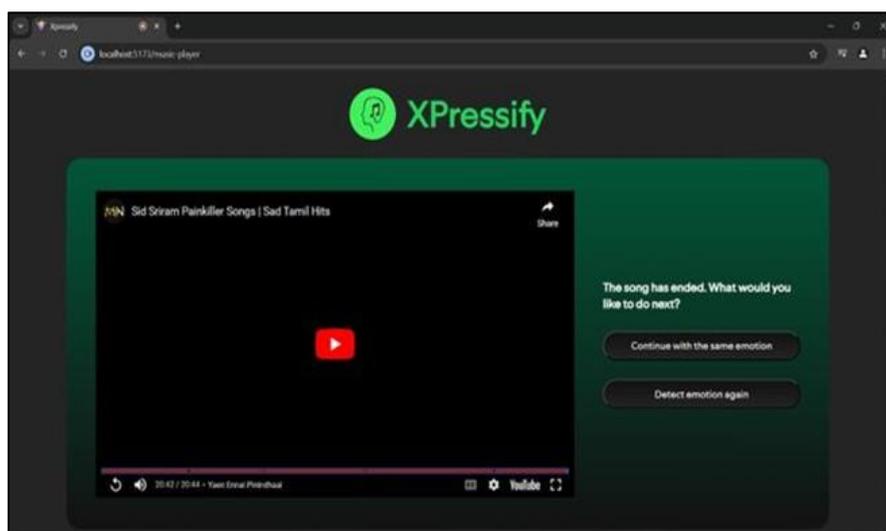**Figure 5.** Final Output- Detected Emotion: Sad

Figure 6. Final Output- Corresponding Music Recommendation: Sad Tamil Hits

The system successfully recognizes the user's emotion as sadness, as seen in Figure 5. Based on the identified emotion, as shown in Figure 6, the system suggests a matching music playlist (Sad Tamil Hits).

5. Comparative Analysis

Table 2. Comparative Table

Feature	Xpressify	S. Shalini et al. (2021)	M. Athavle et al. (2021)	S. Bobade et al. (2023)
Dataset	FER2013 + OAHEGA	FER2013	FER2013	Not specified
Emotions Detected	7 (Anger, Disgust, Fear, Happy, Sad, Surprise, Neutral)	6 (Not clearly listed)	6 (Likely from FER)	5 (Limited)
Accuracy	79.07%	Not reported	Not reported	~71%
Music Recommendation	YouTube API (Dynamic & Real-time)	Static playlist	Local playback	Firebase playlist
Real-Time Capability	Yes	No	Partial	Yes
UI Framework	React (Web-based)	Not specified	Tkinter	Basic UI

6. Conclusion

This project employs state-of-the-art deep learning methods, i.e., Convolutional Neural Networks (CNNs), to recognize facial emotions and recommend music accordingly. An effective preprocessing pipeline was also used to train a model on the FER2013 dataset. The model recognized happiness, sorrow, anger, surprise, fear, disgust, and neutrality with an accuracy of 79.07%. Personalized music recommendations based on emotions can be accessed by integrating a recommendation engine with over a thousand songs/videos via the YouTube API. The development and test data were small. Face variation, face occlusion, and illumination are factors that can independently have a real-world effect on emotion recognition accuracy. Because the system is trained on the FER2013 dataset, there may be some quirks of emotion or culture that were not accounted for. Additionally, the mapping of emotion to music genres is based on past correlations that are not directly transferable to the same level of emotional maturity or personal taste in any given user. Within these constraints, the project integrates emotion detection and music recommendation with personalization, creating an

interactive and novel application that enhances user engagement and emotional health. For long-term music taste adaptation and personalization, future research can explore reinforcement learning applications and user ratings. Consequently, the Facial Emotion Recognition and Music Recommendation System assumes the existence of computer vision and recommendation systems intersecting to develop a smart program. It is a demonstration of artificial intelligence's usage in improving human-computer interaction and provides an interactive emotion-based system with scope for improvement.

References

[1] Bakariya, Brijesh, Arshdeep Singh, Harmanpreet Singh, Pankaj Raju, Rohit Rajpoot, and Krishna Kumar Mohbey. "Facial emotion recognition and music recommendation system using CNN-based deep learning techniques." Evolving Systems 15, no. 2 (2024): 641-658.

[2] Florence, S. Metilda, and Mohan Uma. "Emotional detection and music recommendation system based on user facial expression." In IOP conference series: Materials science and engineering, vol. 912, no. 6, p. 062007. IOP Publishing, 2020.

[3] Shalini, Shantha K., R. Jaichandran, S. Leelavathy, R. Raviraghul, J. Ranjitha, and N. Saravanakumar. "Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques." Turkish journal of computer and mathematics education 12, no. 2 (2021): 912-917.

[4] Patel, Kevin, and Rajeev Kumar Gupta. "Song playlist generator system based on facial expression and song mood." In 2021 international conference on artificial intelligence and machine vision (AIMV), IEEE, (2021): 1-6.

[5] Pareeka, Shreela, Ritika Rai, Rishika Tyagi, Priyanshu Singh, and Vishakha Chauhan. "Emo Melodies-Music Aligned with Facial Emotions." SSRN Electronic Journal (2024).

[6] Chidambaram, G., A. Dhanush Ram, G. Kiran, P. Shivesh Karthic, and Abdul Kaiyum. "Music recommendation system using emotion recognition." International Research Journal of Engineering and Technology (IRJET) 8, no. 07 (2021): 2395-0056.

[7] Bagwan, Raheena, Sakshi Chintawar, Komal Dhapudkar, Alisha Balamwar, and Sandeep Gore. "Facial Emotion Recognition using Convolution Neural Network." International Journal of Trend in Scientific Research and Development (ijtsrd) 5, no. 3 (2021): 800-801.

[8] Athavle, Madhuri, Deepali Mudale, Upasana Shrivastav, and Megha Gupta. "Music recommendation based on face emotion recognition." Journal of Informatics Electrical and Electronics Engineering (JIEEE) 2, no. 2 (2021): 1-11.

[9] S. Bobade et al., "Music recommendation system using facial expression," in Proceedings of the International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, Nov. (2023): 1–5.

[10] S. Kulkarni et al., "Facial emotion based music recommendation system," in Proceedings of the International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, May (2023): 1–6.

[11] Athavle, Madhuri. (2021). Music Recommendation Based on Face Emotion Recognition. Journal of Informatics Electrical and Electronics Engineering (JIEEE). 2. 1-11.

[12] S. K. Sana et al., "Facial emotion recognition based music system using convolutional neural networks," Mater. Today: Proc., vol. 56, 2022, Art. no. 03.131.

[13] N. P. Mankale, "Music Recommendation System Based on Facial Emotions," International Journal of Research Publication and Reviews, vol. 4, no. 10, Oct. (2023): 457–462.

[14] D. Tiwari, H. Kumar, and S. Garg, "Music recommendation system using facial expression," International Research Journal of Engineering and Technology (IRJET), vol. 11, no. 05, May (2024): 2145–2149.

[15] K. Lekha Sree and P. Praveen Kumar, "Music Recommend System Using Facial Emotion Recognition," International Journal of Creative Research Thoughts (IJCRT), vol. 11, no. 10, Oct. (2023): 688–693.

[16] A. Pradhan et al., "Facial emotion based music recommendation system," International Journal of Engineering Applied Sciences and Technology, vol. 7, no. 12, Apr. (2023): 68–70.

[17] M. K, H. R. Jain, A. R. Chandra, and J. Bhansali, "Music Recommendation Based on Facial Expression," International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), vol. IX, no. XI, p. 18, Nov. 2020.

[18] More et al., "Melomaniac: Emotion based music recommendation system," International Journal of Advance Research and Innovative Ideas in Education (IJARIIE), vol. 7, no. 3, (2021): 1323–1329.

[19] Kabani, Hafeez, Sharik Khan, Omar Khan, and Shabana Tadvi. "Emotion based music player." International journal of engineering research and general science 3, no. 1 (2015): 2091-2730.

[20] Bansal, Dhruvisha, Pinkal Bhatt, Megha Dusane, Avneet Saluja, and Kushal Patel. "Emotion based music playing device." International research journal of engineering and technology (IRJET) 7, no. 06 (2020).

[21] Kabani, Hafeez, Sharik Khan, Omar Khan, and Shabana Tadvi. "Emotion based music player." International journal of engineering research and general science 3, no. 1 (2015): 2091-2730.

[22] Sawant, Srushti, Shraddha Patil, and Shubhangi Biradar. "Emotion based music system." International Journal of Innovations & Advancement in Computer Science, IJIACS ISSN (2018): 2347-8616.

[23] Agrawal, Charu, Meghna Varma, Anish Varshaney, Khushboo Singh, Chirag Advani, and Diwakar Yagyasen. "Emotions Based Music Player." International Research Journal of Engineering and Technology (IRJET) 8, no. 7 (2021): 2584-2587.

[24] Singh, Sushant Kumar. "Emotion-Aware Music Recommendation System using Facial Expression Analysis." International Journal of Performability Engineering 21, no. 6 (2025).

[25] Gilda, Shlok, Husain Zafar, Chintan Soni, and Kshitija Waghurdekar. "Smart music player integrating facial emotion recognition and music mood recommendation." In 2017 international conference on wireless communications, signal processing and networking (wispnet), IEEE, (2017): 154-158