2025 2nd International Conference on Integration of Computational Intelligent System (ICICIS)
Pune, India. Sep 09-12, 2025

Facial Expression-Driven Music Recommendation Using Convolutional Neural Networks and Gradio Interface Integration

Sai Srikar Devasani
Department of CSE(AI&ML)
Vardhaman College of Engineering
Hyderabad,Telangana,India
saisrikardevasani@gmail.com

Dr D Rajeshwari
Department of CSE(DS)
Sri Indu Institute of Engineering and Technology
Hyderabad,Telangana,India
rajeshwaricse546@gmail.com

Manzoor Mohammad
Department of CSE(AI&ML)
Vardhaman College of Engineering
Hyderabad,Telangana,India
manzoor@vardhaman.org

M.A.Jabbar
Department of CSE(AI&ML)
Vardhaman College of Engineering
Hyderabad,Telangana,India
Jabbar.meerja@gmail.com

Burra. Vijaya Babu
Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India
vijay_gemini@kluniversity.in

MD Reshma
Department of ECE
Sri Indu Institute of Engineering and Technology
Hyderabad,Telangana,India
reshumd2017@gmail.com

2025 2nd International Conference on Integration of Computational Intelligent System (ICICIS) | 979-8-3315-3736-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICICIS65613.2025.11371105

***Abstract*—Music recommendation systems have become an important part of improving user experience in today's world of personalized technology. This paper introduces a unique way of recommending music based on facial expressions using deep learning techniques. By leveraging computer vision and convolutional neural networks (CNNs), the system analyzes real-time facial expressions captured via a webcam to infer emotional states. These emotional cues are then mapped to corresponding music genres or tracks that align with the user's current mood. The system is designed to offer dynamic and personalized music recommendations, enhancing user engagement and satisfaction. The project combines advances in facial expression recognition with music recommendation algorithms, creating a seamless integration of emotion-aware music suggestions. This approach not only improves the relevance of recommendations but also adapts to the emotional context of the user, offering a more intuitive and responsive user experience.**

***Index Terms*—Facial Expression Recognition, Music Recommendation, Deep Learning, Computer Vision, Convolutional Neural Networks (CNNs), Emotion Analysis and Personalized Music Suggestions.**

I. Introduction

Music is an important part of life, often reflecting emotions and shaping experiences [1]. Traditional music recommendation systems are mainly based on user preferences and past listening history, which may not always match the current mood of users [2]. With advances in deep learning and computer vision, facial expression recognition has become a reliable way to understand emotions. This project introduces a facial expression based music recommendation system that uses facial expressions to detect the user's current emotion and suggest suitable music tracks that fits their mood [4]. Convolutional Neural Networks (CNNs) are trained for facial expression recognition, enabling the system to accurately classify emotions such as Happy, Sad, Angry, Disguist, Fear, Surprise and Neutral. Various technologies are integrated to ensure accurate emotion detection and personalized music recommendations, enhancing the overall user experience by aligning musical suggestions with the user's current emotional state [6].

A. Background and Motivation

Music plays an important role in shaping emotions, often serving as a source of comfort, relaxation or motivation. Traditional music recommendation systems primarily focus on analyzing past user interactions, listening history and predefined preferences to suggest songs [5]. Although this approach can be effective in identifying general musical tastes, it falls short when it comes to adapting to the user's real-time emotional state. This limitation can lead to recommendations that, although relevant in a broader sense, may not reflect what the user truly needs or wants to hear at a specific moment. With advances in deep learning, particularly in the field of computer vision, emotion recognition through facial expressions has become increasingly accurate and reliable. Convolutional Neural Networks (CNNs) are very effective in recognizing facial expressions with great accuracy. Using this technology, this system aims to create a music recommendation system that suggests songs based on the emotions of the user, making the listening experience more personalized and engaging [13].

B. Problem Statement

Building a facial expression based music recommendation system using deep learning comes with several challenges,

mainly due to the complexity of recognizing facial expressions and matching them to emotions in real time [10].

Key challenges include:

- Fast Real Time Processing: The system needs to quickly recognize facial expressions and recommend music without delays, requiring an optimized model and fast processing [14].
- Matching Emotions to Music: Choosing the right songs for detected emotions requires a well defined system that accurately links moods with music genres [16].
- Lighting and Environment Issues: Changes in lighting, camera angles and background distractions can reduce the accuracy of facial expression detection [17].

To overcome these challenges, a Convolutional Neural Network(CNN) is used to recognize different facial expressions. CNNs are very good at understanding images and can learn to spot small differences in facial features [3]. This helps the system accurately detect emotions like Happy, Sad, Angry, Disguist, Fear, Surprise and Neutral. Real-time detection is optimized using OpenCV and Gradio, ensuring efficient processing. The system is integrated with the Spotify API to provide seamless music recommendations based on detected emotions.

C. Objectives

The main goal is to build a system that recommends music based on real-time facial emotion recognition. The following objectives outline the key steps taken to achieve this:

- To train a CNN model to accurately classify facial expressions into different emotional categories such as Happy, Sad, Angry, Disguist, Fear, Surprise and Neutral [19].
- Use OpenCV to capture and process facial expressions in real time [22].
- Implement Gradio to provide a simple and interactive interface for smooth user interaction [23].
- To map the emotions detected to suitable music genres or tracks using the Spotify API [24].
- To improve recognition accuracy through data augmentation, fine tuning and preprocessing technique [25].
- To develop a responsive and user-friendly system that delivers music suggestions based on real-time emotional input [26].

II. LITERATURE REVIEW

Londhe et al. (2020) proposed a music recommendation system that detects emotions through facial expressions using Convolutional Neural Networks (CNN), aiding in emotion-based music selection [1].

Jaiswal et al. (2020) implemented a deep learning approach to detect human emotions from facial expressions, utilizing a CNN model that provided effective classification of basic emotions [2].

Krupa et al. (2020) developed a facial emotion recognition model using CNN, where images were preprocessed and classified to identify emotional states [3].

Florence and Uma (2020) introduced a real-time facial expression-based music recommendation system using CNN architecture that linked emotions to music tracks [4].

Gilda et al. (2021) presented a smart music player that recommends songs based on detected facial emotions using deep learning, aiming to enhance user experience [5].

Kevin and Rajeev (2021) proposed a mood-based music player using facial expression recognition, incorporating image processing and CNN for real-time emotion detection [6].

Zhalehpour and Amini (2021) developed an emotion-based music recommendation system through facial expression recognition, integrating it with playlist generation [7].

Amogh et al. (2021) conducted a survey on various deep learning approaches for emotion detection and highlighted CNN as the most effective method [8].

Fessahaye et al. (2020) introduced T-RECSYS, a context-aware music recommender system that personalizes suggestions based on user context and preferences [9].

Shaees et al. (2020) employed transfer learning for emotion detection, leveraging pre-trained models for improved accuracy in facial expression classification [10].

Lathiya et al. (2021) provided a review of music recommendation systems using machine learning techniques, focusing on content and collaborative filtering [11].

Dinesh and Ramesh (2021) used a hybrid filtering approach to develop a music recommendation system that considers both user preferences and item similarity [12].

Singh and Dubey (2021) designed an emotion recognition system using CNN and LSTM for facial expressions, achieving high accuracy in sequential data modeling [13].

Moswedi and Ajoodha (2022) classified music mood using audio features and machine learning techniques to match songs with emotional states [14].

Prasad et al. (2022) implemented a real-time facial emotion recognition system using CNN, achieving high accuracy under well-lit conditions [15].

Verma et al. (2021) developed a real-time deep learning-based facial emotion recognition system, integrating image acquisition and expression classification [16].

Rumiantcev and Khriyenko (2020) introduced an emotion-aware music recommender system that adapts suggestions based on real-time emotion analysis [17].

Roy and Majumder (2021) presented a CNN-based facial expression recognition system trained on static images for emotion classification [18].

Sharma and Kaur (2020) proposed a deep learning-based music recommendation system that selects tracks according to facially detected emotional states [19].

Harshitha et al. (2021) implemented a real-time facial emotion recognition model using deep CNN for applications such as music personalization and surveillance [20].

Li et al. (2020) introduced an attention-based CNN model for facial emotion recognition in videos, providing better focus on expressive regions of the face [21].

Banu et al. (2021) designed a music recommendation system based on user emotion detection using facial features and mapped moods to genre types [22].

Anand and Shukla (2020) implemented a real-time facial expression recognition model using deep learning techniques, enabling responsive music control [23].

Patel et al. (2021) developed an adaptive music recommendation system that identifies user emotion and dynamically adjusts music suggestions accordingly [24].

Dixit and Shukla (2021) employed CNN for facial emotion recognition and evaluated model performance on datasets representing various emotions [25].

Tripathi and Saxena (2021) proposed an emotion-based music recommendation system using machine learning techniques like decision trees and k-nearest neighbors [26].

Prathibha and Mounika (2021) created a music recommendation system using facial emotion recognition and linked emotional states to curated playlists [27].

Shinde and Patil (2020) developed a facial expression-based smart music system that analyzed user expressions to automate playlist management [28].

Sharma and Arora (2021) introduced a real-time deep CNN-based facial emotion recognition system for intelligent multimedia interaction [29].

Yadav and Vishwakarma (2020) reviewed various techniques used for emotion recognition through facial expressions, emphasizing deep learning frameworks [30].

Febriana et al. (2021) proposed a facial expression recognition model using a Bidirectional LSTM-CNN architecture, improving temporal emotion analysis [31].

III. MATERIALS AND METHODS

A Facial expressions based system was developed to recommend personalized music playlists by analyzing user emotions using deep learning techniques. This implementation involved various tools and methods that ensured high accuracy in emotion detection and smooth integration with music recommendation APIs [9].

A. Features and Capabilities of the CNN Model

To detect emotions from facial expressions, we trained a Convolutional Neural Network (CNN). CNNs are powerful models used in image-related tasks because they can automatically detect and learn patterns from input images [22]. They are especially useful in identifying subtle changes in facial features that represent different emotional states. The main features of CNN model include:

- Automatic Feature Extraction: CNN automatically focus on important facial parts such as the eyes, eyebrows and mouth. These areas are crucial in determining expressions and emotions [23].
- Emotion Classification: Once trained, the model is capable of accurately classifying an input image into one of several emotional categories, including Happy, Sad, Angry, Disgust, Surprise, Fearful and Neutral [24].
- Custom Training: The CNN model was trained on a well-known facial emotion dataset, making it reliable for emotion detection in real-time conditions [25].
- Lightweight and Fast: CNN was optimized to work quickly, making it suitable for real-time use in a web application, giving fast results to users [30].

B. Dataset used and its Features

The FER-2013 (Facial Expression Recognition 2013) dataset is widely used for training deep learning models to recognize facial emotions. It was originally introduced in a Kaggle competition and contains around 35,887 grayscale images, each with a resolution of 48x48 pixels [26]. These images capture a diverse range of human expressions, making the dataset suitable for building a robust emotion recognition model. The dataset includes seven emotion categories: Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral. Each image is labeled with one of these emotions, helping the model to learn patterns associated with different facial expressions. Since the images contain faces from various individuals, the dataset ensures diversity which improves the model's ability to generalize across different users. To effectively train and evaluate the deep learning model, the dataset is split into training and test sets [27]. The training set allows the model to learn from labeled examples, while the test set is used to check how well the model recognizes emotions on new and unseen images. This structured approach ensures that the system performs reliably when deployed for real-world facial expression based music recommendations.

C. Hugging Face Hub for Saving and Hosting the Model

The Hugging Face Hub was used to store, manage and share the trained Convolutional Neural Network (CNN) model, providing a convenient and efficient platform for managing machine learning models and datasets [11]. By hosting the model in Hugging Face, it can be easily accessed and deployed without requiring local storage or manual setup. One of the key benefits of using Hugging Face is its built-in version control, which allows us to save and update new versions while keeping older ones for comparison [15]. This is especially helpful during fine-tuning or testing new data augmentation techniques. Hugging Face works well with TensorFlow and PyTorch, making it simple to upload, download and fine tune models. It also supports collaboration which allows team members to access and test the model from different devices. Additionally, the platform ensures security and reliability, safely storing the model and making it easy to use in the Gradio interface for deployment and interaction [18].

D. Interface Design for Music Recommendation by Facial Expression Detection

Gradio was used to design a simple, user-friendly and interactive web-based interface that connects the trained CNN model and the end user [20]. The interface allows users to upload an an existing image or capture a live photo directly using their device's webcam. This ensures that users can

interact with the system in a way that is most convenient to them. Once the image is submitted, the model processes it and displays the predicted emotion on the screen. Based on the detected emotion, the system recommends a suitable Spotify playlist instead of a single song [21]. For example, a "Happy" expression triggers an energetic playlist, "Sad" suggests a soothing or calm playlist and "Angry" links to relaxing music. The entire application runs directly in a web browser and doesnot require installation. This makes it highly accessible and easy to use across different devices and platforms [25].

Using only a facial expression, users can receive music recommendations that are more relevant and emotionally connected, creating a more personalized and enjoyable listening experience.

IV. PROPOSED SYSTEM

The proposed system is designed to recommend a suitable music playlist based on a user's facial expression using deep learning techniques. The system works by following a step-by-step process that includes capturing an image, recognizing the emotion from the face and recommending music based on the detected emotion [27].

The system starts with the user providing an input image through the Gradio interface either by uploading a photo or using a webcam. This image is preprocessed (resized and normalized) and passed to a trained CNN model. This CNN model was trained on the FER-2013 dataset which contains thousands of labeled images of facial expressions [29]. The model can identify one of the seven basic emotions: Angry, Disgust, Fear, Happy, Sad, Surprise and Neutral and outputs the most probable emotion label based on the facial features detected in the image [30]. It analyzes facial features like the eyes, eyebrows and mouth to determine which emotion best matches the users expression. Once the emotion is detected, the system maps that emotion to a corresponding Spotify playlist [31]. This mapping is predefined with each emotion linked to a curated playlist that fits the mood. For example, a cheerful playlist for "Happy" or a calming one for "Sad" [7]. The Spotify playlist is then displayed as a clickable link for the user to listen to directly. By combining facial expression recognition with emotion based music recommendation, the system offers a personalized and engaging user experience. Users don't need to search for music manually, instead they simply show how they feel and the system does the rest. Technologies like Gradio help in creating an easy-to-use web interface, while Hugging Face is used to store and access the trained model safely and efficiently [8].

The following steps summarize the system workflow as depicted in Fig.1:

- User Input: The system allows the user to upload an image, paste it or capture it directly using a webcam. This makes the system flexible and easy for anyone to use.
- Image Preprocessing: The input image undergoes a few adjustments to help the model understand it better. These include converting the image into grayscale (removing colors to focus on expressions), resizing it to a standard size and normalizing the values.
- Display Fun Fact: To make the experience more fun and engaging, a random fun fact is shown to the user while the system processes the image. This helps keep the user entertained and reduces the waiting time.
- Emotion Detection: After preprocessing, the cleaned image is passed to a trained CNN (Convolutional Neural Network) model. This model looks at facial features and identifies the emotion shown in the image.
- Playlist Recommendation:The system uses the detected emotion to pick a matching spotify playlist. Each emotion is mapped to a corresponding Spotify playlist that aligns with the user's emotional state.
- Output Display: In the end, the system displays the fun fact, predicted emotion and the recommended Spotify playlist in a clean and interactive interface.

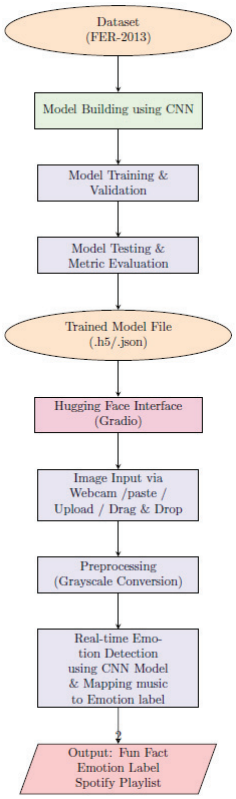

Fig. 1. Workflow of the Proposed Facial Expression-Driven Music Recommendation Using Convolutional Neural Networks and Gradio Interface Integration

V. EXPERIMENTAL RESULTS AND DISCUSSION

The implemented system was evaluated using the FER-2013 dataset, achieving an accuracy of 71.13% in emotion classification using the CNN model, which demonstrated a reliable performance for real-time emotion recognition from facial expressions [12]. It allows users to upload an image or

use a webcam for input. It has the ability to provide real-time predictions through a web-based interface built using Gradio. Once the user uploads or captures an image using their webcam, the system performs image preprocessing, detects the dominant facial emotion using the trained CNN model and provides a Spotify playlist based on that emotion [28]. This integration ensures that the user receives a personalized music recommendation that matches their mood.

The user experience is further enhanced by including a fun fact, which appears immediately after image submission, keeping the user engaged while the system processes the image. The final output displays the fun fact, the predicted emotion and a Spotify playlist link.

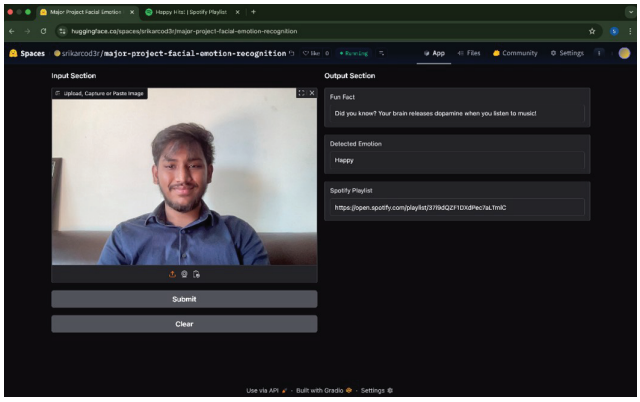

Fig. 2. Emotion Based Music Recommendation Interface

Fig.2 shows the final version of the user interface where the user has submitted an image through the webcam. The system detected the emotion as “Happy”, displayed a relevant fun fact and recommended a Spotify playlist link with energetic and uplifting songs.

VI. CONCLUSION

This paper discussed an optimal approach to effectively demonstrate how facial expressions can be used to recommend music related to the user’s emotional state. By training a Convolutional Neural Network (CNN) on the FER-2013 dataset, this system is able to accurately identify seven basic emotions. The model is deployed using Hugging Face Hub, while the user interface is built with Gradio allowing for simple and interactive real-time usage. The system not only provides emotion detection, but also an engaging user experience by instantly presenting a matching Spotify playlist and a fun fact. Achieving an accuracy of 71.13%, the model shows promising performance for practical applications.

Further enhancements will focus on fine-tuning the model with more diverse and balanced datasets, increasing the accuracy of emotion classification and improving robustness. Additionally, incorporating a wider range of emotions, refining playlist mapping and enabling user feedback mechanisms will further enrich the overall user experience.

REFERENCES

[1] Londhe, A., Gupta, S. and Bhutada, P., 2020. Music recommendation system based on emotion recognition from facial expression using CNN. International Research Journal of Engineering and Technology (IRJET),7(6), pp.4280-4283.

[2] Jaiswal, P., Bhardwaj, A. and Sharma, A., 2020. Emotion detection from facial expressions using deep learning. International Journal of Scientific & Technology Research, 9(4), pp.1428-1432.

[3] Krupa, D.P., Nikam, S. and Bhosale, S., 2020. Facial emotion recognition using CNN. International Journal of Engineering Research & Technology (IJERT), 9(07), pp.203-206.

[4] Florence, A. and Uma, G.V., 2020. Real-time facial expression based musicnrecommendation system using CNN. International Journal of Recent Technology and Engineering (IJRTE), 9(4), pp.382-387.

[5] Gilda, A., Khandge, A. and Shinde, A., 2021. Smart music player based on facial emotion recognition. International Journal of Creative Research Thoughts (IJCRT), 9(5), pp.472-476.

[6] Kevin, J. and Rajeev, R., 2021. Mood based music player using facial expression recognition. International Journal of Computer Applications,183(1), pp.33-36.

[7] Zhalehpour, S., & Amini, M. (2021). Emotion-Based Music Recommendation System Using Facial Expression Recognition. Journal of AI and Data Mining, 9(2), 215–222.

[8] Amogh, A., Sultana, S. and Sharma, D., 2021. Emotion detection using deep learning approaches: A survey. Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS),pp.1622-1626.

[9] Fessahaye, M., Musaye, A., Kalinga, A. and Pairet, E., 2020. T-RECSYS:A context-aware music recommender system. Proceedings of the International Conference on Intelligent Human Systems Integration (IHSI),pp.83-88.

[10] Shaees, M., Shafi, I. and Raza, M., 2020. Emotion detection using transfer learning. International Journal of Advanced Computer Science and Applications (IJACSA), 11(5), pp.327-332.

[11] Lathiya, R., Dave, M., Patel, R. and Parmar, B., 2021. A Review on Music Recommendation System using Machine Learning. International Research Journal of Engineering and Technology (IRJET), 8(5), pp.3366-3369.

[12] Dinesh, D., & Ramesh, M. (2021). Music recommendation using hybrid filtering. Journal of Applied Science and Computations, 8(1), pp.76-81.

[13] Singh, V. and Dubey, A.K., 2021. Emotion recognition system based on facial expressions using CNN and LSTM. Materials Today: Proceedings, 47, pp.4753-4756.

[14] Moswedi, R.M. and Ajoodha, R., 2022. Mood classification of music using audio features. Proceedings of the SAICSIT Conference, pp.62-69.

[15] Prasad, R., Pradhan, A. and Sinha, R., 2022. CNN based approach for real time facial emotion recognition. International Journal of Innovative Research in Computer and Communication Engineering, 10(1), pp.22-28.

[16] Verma, S., Gupta, A. and Singh, D., 2021. Deep learning based real time facial emotion recognition system. Journal of Information and Computational Science, 11(8), pp.487-492.

[17] Rumiantcev, S. and Khriyenko, O., 2020. Emotion-aware music recommender system. International Conference on Computer Science and Artificial Intelligence, pp.162-169.

[18] Roy, A. and Majumder, D., 2021. Emotion recognition from facial expressions using CNN. Materials Today: Proceedings, 47, pp.1981-1986.

[19] Sharma, R. and Kaur, H., 2020. Music recommendation system based on facial emotion recognition using deep learning. International Journal of Computer Sciences and Engineering, 8(12), pp.18-24.

[20] Harshitha, R., Divya, R., & Lakshmi, S. (2021). Facial expression recognition using deep CNN for real-time applications. International Journal of Advanced Trends in Computer Science and Engineering, 10(3),pp.1772–1777.

[21] Li, Y., Zeng, J., & Wang, Z. (2020). Attention-based CNN for facial emotion recognition in videos. IEEE Access, 8, pp.68091-68100.

[22] Banu, M., Raju, A., & Prasad, K. (2021). Music Recommendation System based on User Emotion Detection. Journal of Physics: Conference Series,1716, 012016.

[23] Anand, A. and Shukla, R., 2020. Real-time facial expression recognition using deep learning. International Journal of Computer Applications, 177(26), pp.21-25.

[24] Patel, N., Sharma, A., & Patil, A. (2021). Adaptive Music Recommendation System Using Emotion Recognition. Journal of Emerging Technologies and Innovative Research (JETIR), 8(6), pp.504-510.
[25] Dixit, A., & Shukla, S. (2021). Facial Emotion Recognition Using Convolutional Neural Network. Materials Today: Proceedings, 46, pp.6302–6307.
[26] Tripathi, P. and Saxena, R., 2021. Emotion based music recommendation system using machine learning techniques. International Journal of Innovative Research in Technology, 8(2), pp.1-7.
[27] Prathibha, C., & Mounika, B. (2021). Music recommendation using facial emotion recognition. International Journal of Research in Engineering, Science and Management, 4(6), pp.164–167.
[28] Shinde, A., and Patil, A. (2020). Facial Expression Based Smart Music System. International Journal of Engineering Research & Technology (IJERT), 9(12), pp.1-5.
[29] Sharma, P., and Arora, A. (2021). Real-time facial emotion recognition using deep convolutional networks. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(1), pp.100-104.
[30] Yadav, S., & Vishwakarma, D. (2020). A review of emotion recognition using facial expressions. International Journal of Information Management Data Insights, 1(1), 100002.
[31] Febriana, R., Halima, B.M., Christinaa, M., Ramdhana, D. and Andry, 2021. Facial expression recognition using bidirectional LSTM- CNN. Journal of Physics: Conference Series, 1869(1), p.012084.