Real-Time Emotion Detection using CNN Integrated with Music Recommendation and Selection

[1]M.Swathi Reddy
Assistant Professor,Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women,Hyd
Email: swathi.madireddy@gmail.com

[2]Chinthala Pavani
UG Student,Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women,Hyd
Email: pavanichinthala20@gmail.com

[3]Kodithala Shirisha
UG Student,Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women,Hyd
Email: shirishak236@gmail.com

[4]Nenavath Pravalika
UG Student,Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women,Hyd
Email: npravalika222@gmail.com

***Abstract*-In the modern digital era, personalization has become a prerequisite to enhancing how we relate to technology, particularly in the entertainment sector. This project proposes a smart music recommendation and playback system based on facial emotion recognition to provide music that resonates with the emotional state of an individual. The concept is straightforward but potent: by scanning facial expressions in real time, the system selects music corresponding to the user's mood, providing a more personalized and interactive listening experience. The system operates by detecting the user's facial expressions using a webcam or camera. It employs pre-trained Convolutional Neural Network (CNN) models, including CNNModel.h5, to identify emotions such as happiness, sadness, anger, surprise, or neutrality. These models have been trained on emotion recognition datasets to deliver accurate and timely results. When the emotion is detected, it's matched against a particular set of songs designed to suit that mood. The app is developed using the Django framework to enable everything from linking the machine learning model to managing the music recommendation process. Users may listen to the recommended tracks or opt to replace them with their own choice. This project integrates artificial intelligence with web development in a novel manner, providing an enjoyable and interactive product.that makes technology come across as more emotionally intelligent and responsive in nature.**
KEYWORDS-Recommendation System, Emotion Recognition, CNN Models, Face Recognition, Webcam.

I.INTRODUCTION

Music is a part of our daily lives. It helps us express emotions, uplift mood, and bring comfort during tough times. Music is a universal language, which connects people across cultures and regions.Music has a deep emotional impact.A single song can make us smile,bring back memories or even move us to tears without saying a single word.

People from different places and cultures connect through music, making it a powerful way to share feelings. Over the years, music has changed a lot, and today, we can listen to any song we like with just a few clicks. In today's digital era, music streaming platforms offer users access to an enormous collection of songs. However, this abundance often makes it difficult for users to choose music that suits their current mood or activity. Typically, music recommendations are based on past user interactions. Recommendation systems aim to address this challenge by offering personalized music suggestions based on user preferences, behaviors, and contextual data. Among these, emotion-based music recommendation systems provide a more human-centered approach by aligning music suggestions with the user's emotional state.Music is a powerful medium that influences emotions, productivity, and mental health. Emotionally relevant music can promote relaxation, relieve stress, enhance focus, and even aid in therapy. Despite advances in recommendation technologies, systems that detect real-time emotions and recommend music accordingly are still evolving.This paper presents a Mood-Based Music Recommendation System Using Facial Emotion Recognition, implemented using the Django web framework. The system uses OpenCV to capture real-time facial images from a webcam and perform preprocessing. Emotion detection is handled by a pre-trained Convolutional Neural Network (CNN), whose architecture is stored in cnnmodel.json and weights in cnnmodel.h5. By using a pre-trained model, the system avoids resource-intensive training, thereby enhancing both speed and accuracy. Once the user's emotion,such as happiness, sadness, anger, surprise, or fear is detected, the system recommends songs that match the identified emotional state, delivering a seamless and personalized music experience.

II. LITERATURE REVIEW

Music has a profound impact on human emotions and brain activity. Studies have shown that individuals often use music to regulate arousal levels and mood, which significantly affects psychological states. Two primary functions of music include helping people achieve a positive emotional state and increasing self-awareness. Furthermore, musical preferences are closely associated with personality traits and emotional conditions.The brain processes musical attributes such as meter, timbre, rhythm, and pitch in regions responsible for emotion and mood regulation. This strong link between music and emotion presents opportunities to use music as a tool for emotional intervention and mental well-being support. Since music conveys rich emotional information, understanding a person's emotional state enables the customization of musical experiences to enhance user satisfaction and psychological comfort.Traditionally, music recommendation systems rely on manual inputs, wearable devices, or audio feature classification. However, these methods often suffer from high computational costs and lack real-time adaptability. To overcome these drawbacks, various studies have explored facial expression recognition as a promising alternative.

Our approach adopts automatic facial emotion recognition using a Convolutional Neural Network (CNN) trained on the FER2013 dataset, enabling the real-time classification of facial expressions into basic emotions such as *happy, angry, sad, surprise,* and *neutral*. The user's image is captured through a built-in webcam, and based on the detected emotion, an appropriate music playlist is automatically generated. Pygame and Tkinter are used for interactive music playback. This method demonstrates better performance in terms of computational efficiency and accuracy when compared to traditional approaches [1][2].

In [3], Yusuf Yaslan et al. proposed an emotion-based music recommendation system that uses physiological signals captured via wearable devices integrating Galvanic Skin Response (GSR) and Photoplethysmography (PPG) sensors. While effective, such hardware-dependent systems may not be accessible or affordable for all users.

In [4], Ayush Guidel et al**.** demonstrated that facial expressions are a reliable indicator of human emotional states. Their system categorized emotions into *happy, sad, anger, excitement, surprise, disgust, fear,* and *neutral*. Face detection was achieved using a CNN, highlighting the effectiveness of deep learning in facial emotion recognition. They emphasized that music, often described as the "language of emotions," can be tailored based on facial emotion analysis.

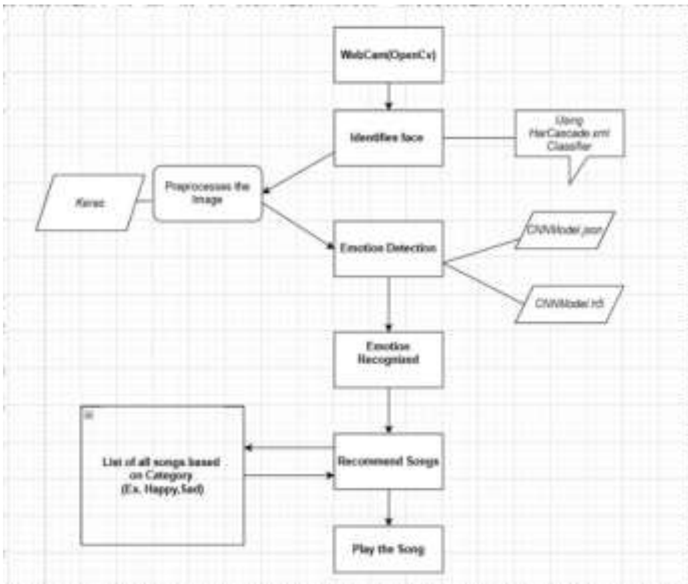
In [5], the authors explored the use of multimodal data facial expressions, voice features, and textual inputs for more accurate emotion detection. The integration of image, audio, and text data improves classification accuracy by capturing various emotional cues. Emotions such as *happy, sad, disgust, anger, fear,* and *surprise* were categorized in this study, which reflects the complexity and depth of emotional understanding in humans.

Furthermore, Ramya Ramanathan et al. [6] proposed an intelligent music player that uses emotion recognition to group a user's local music library based on emotion-related metadata, particularly lyrics. Their study offered insights into different methodologies for detecting human emotions and implementing emotion-based music players. It also provided a brief overview of system functionalities, including emotion classification and playlist generation. It generates playlist based on emotion recognized from the local playlist what you are using. Your local playlist helps in identifying your intrest.

III. METHODOLOGY

This paper focuses on one of the most interesting applications of machine learning — detecting a person's emotion by simply looking at their face. The core idea is to understand the emotion a person is showing and recommend songs that match their mood. To achieve this, the system uses a pretrained Convolutional Neural Network (CNN) model that was trained earlier using thousands of facial images showing different emotions such as happy, sad, angry, and more. The training part is very important because, just like a human need to see examples to learn emotions, the model also needs to be trained with examples of each type of facial expression to understand and recognize them later. Now, when a new image is captured through the webcam, it might not be in the same format or quality as the images the model was originally trained on. So before giving it to the model, we follow a few steps this is called image pre-processing. We crop the face area from the image, resize it to the right shape (32x32 pixels), normalize the values (so all images have similar brightness/contrast levels), and reshape it to match the models expected format. This step is important because it helps the model recognize emotions more accurately.

System Architecture :



The system architecture is designed to offer a smart and seamless music experience based on the user's facial emotions, all built using the Django framework. The process starts when the webcam accessed through OpenCV captures the user's facial image. Haarcascade is then used to detect facial features and landmarks. This image is passed to a Convolutional Neural Network (CNN) model that has been pre-trained to recognize emotions such as happy, sad, angry, and more.Once the emotion is identified, Django serves as the backbone for integrating the front-end interface with the back-end logic. The recognized emotion is passed to a music recommendation module, which selects songs that match the user's current mood. Django efficiently handles user interaction, data flow, and real-time communication between the camera input, emotion detection module, and music player—delivering a responsive and personalized music playback experience.

A. FaceDetection

Using a webcam to take a picture of the user's face is the first and most fundamental step in the suggested system. The application turns on the webcam and snaps a picture of the user's face after obtaining consent for camera access. The raw image acquired at this point, however, usually includes more than just the face; it might also include elements of the background, additional objects in the scene, or even more than one face if there are multiple people present. These extra components have the potential to add noise and lower the precision and accuracy of emotion recognition.To ensure that only relevant facial data is used, the system employs a face detection mechanism using the Haar Cascade Classifier provided by the OpenCV library. This classifier is a pre-trained model specifically designed to recognize the structural patterns of the human face by identifying key features such as the eyes, nose, and mouth. It operates by scanning the input image and matching visual patterns to known face-like structures through a cascade of increasingly complex stages, thus reliably pinpointing the region where the face is located.Once a face is detected, the system isolates it by cropping the image to include only the region containing the face, effectively removing all extraneous visual information. This step not only enhances the clarity of the input data but also significantly improves the performance and speed of the subsequent emotion recognition model. The resulting cropped face image serves as the primary input for the emotion classification module, ensuring that the system processes only the most relevant visual data.Once the facial region has been accurately isolated, the system proceeds with preprocessing the cropped face image to prepare it for emotion recognition.

B. Image Preprocessing

The image goes through a critical preprocessing step to get it ready for the Convolutional Neural Network (CNN) to classify emotions after the face has been precisely identified and cropped. Because the CNN model was trained on inputs in a particular structure and format, this step is crucial. This stage's first task is to resize the cropped face image to 32 by 32 pixels. The processing pipeline is made simpler and faster by this resizing, which guarantees consistency with the model's training data and permits uniform input size. In machine learning tasks, standardizing image dimensions is especially crucial because it helps avoid discrepancies that could otherwise compromise prediction accuracy.

After being resized, the picture is converted into a numerical array that shows each pixel's intensity. Images are recorded and saved as numerical data since they are made up of red, green, and blue (RGB) values. Normalization is the next preprocessing step. To scale each pixel value, which normally falls between 0 and 255, into a range between 0 and 1, it is divided by 255. Because it speeds up convergence and lowers the possibility of exploding or vanishing gradients during prediction, this normalization procedure is a widely used best practice in neural networks.

After normalization, the image is reshaped into a four-dimensional array with the format 1x32x32x3, where '1' denotes the batch size (meaning that only one image is processed at a time), '32x32' denotes the image's dimensions, and '3' denotes the RGB color channels. This modified format guarantees that the input closely complies with the CNN model's structural specifications. The image is now in the best possible condition to be fed into the CNN in order to identify emotions. By carefully executing this preprocessing pipeline resizing, normalizing, and reshaping—the system ensures maximum compatibility and performance efficiency, forming a robust bridge between raw facial data and high-accuracy emotion classification.

C. Emotion Detection

Following thorough preprocessing and format conversion, the facial image is fed into a Convolutional Neural Network (CNN) that has already been trained to identify a range of human emotions. Large-scale datasets like FER2013, which contain thousands of labeled facial images displaying a variety of emotional expressions like happiness, sadness, anger, surprise, and neutrality, were used to train this CNN model.The CNN has acquired the ability to recognize subtle patterns and visual cues in facial features that are connected to various emotions as a result of this intensive training. The CNN goes through several layers as it processes the new image, extracting increasingly complex and abstract features with each step. For example, deeper layers identify particular facial features like the curvature of a smile, raised eyebrows, or widened eyes, while lower layers may detect edges and textures.The model then gives each potential emotion a probability score based on this hierarchical understanding. The likelihood that the provided image reflects each emotion category is indicated by these probabilities. The final prediction is the emotion with the highest probability. For instance, the model might identify the emotion as "happy" if it notices an upward curvature of the lips and bright, widened eyes.

This identified emotion serves as a crucial input for the system's music recommendation module, which comes next. The system can improve the user experience by dynamically creating a playlist of songs that complement and possibly elevate the user's current mood by precisely determining their emotional state.

D. Music Recommendation

Once the facial image is preprocessed and converted into the proper format, it is sent to a Convolutional Neural Network (CNN) model that has already been trained to recognize human emotions. This model was trained on thousands of facial images labeled with different emotional expressions such as *happy*, *sad*, *angry*, *surprised*, and others. Since it already "knows" what these emotions look like based on its training, it compares the features in the new input image to the patterns it has learned. The CNN goes layer by layer, identifying facial features like the curve of a smile, the position of the eyebrows, or the openness of the eyes. Based on these, it assigns a probability to each emotion category. The emotion with the highest probability is selected as the predicted emotion of the user. For example, if the model sees a wide smile and bright eyes, it may detect the emotion as *happy*. This detected emotion is then used in the next step to recommend music that matches the user's mood.

E. Music Playback and Termination

Following the emotion recognition and recommendation phase, the user is presented with a dynamically generated list of songs associated with the detected emotional state. Upon selection of a desired track, the system initiates audio playback using the playsound() function from Python's standard libraries. This function is executed within a separate multiprocessing.Process thread to ensure that the main application thread (Django web server) remains non-blocking and responsive to further user interactions.The asynchronous nature of the playback allows users to engage with other functionalities of the application while the song plays in the background. To enhance user control, the interface also includes a "Stop" link. When this is triggered, the corresponding audio process is programmatically terminated using p.terminate().This module effectively closes the loop of the system’s operational pipeline — starting from webcam-based image capture, progressing through image preprocessing, facial emotion detection via a pretrained CNN model, music recommendation based on emotion, and finally real-time audio playback with manual stop functionality.

IV. RESULTS AND ANALYSIS.

Fig 4.1: Interface

This is the initial user interface of the system which consists of home and Emotion Detection Webcam. We can access webcam by clicking on it, next it opens the webcam.Where we can give our image.

Fig 4.2: Permission for Webcam

When we click on “Emotion Detection Webcam” on the user interface,the webcam prompts us for permission. We can click on “Allow this time” to grant access.once permission is granted, the webcam opens and get ready to capture our face.After positioning our face infront of the webcam, we need to click the “Take Snapshot” to record the image.When we click on “Emotion Detection Webcam,” the backend activates the webcam using OpenCV's cv2.VideoCapture() function. After the user positions their face, clicking “Take Snapshot” captures a frame with cv2.read().This image is thenpreprocessed and fed into a trained CNN model to detect the user's emotion. Based on the detected emotion, the system plays a suitable music track to match the mood.

Fig 4.3: Capturing Image

Now , the image gets captured, and given to further processing.

Fig 4.4: Emotion Detected

It recognizes the emotion and mentions which emotional state you are in.

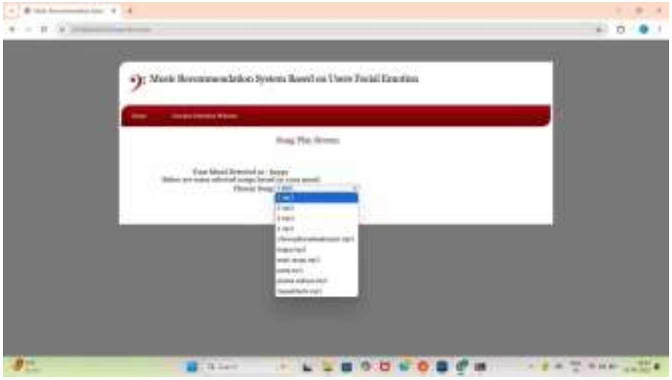


Fig 4.5: Recommended songs

Based on the recognized emotion it recommends songs. We can play the song of our interest.

V. CONCLUSION

In this paper, we successfully created a music recommendation system that understands human emotions just by looking at their facial expressions. It starts by capturing the user's photo using the webcam, then detects the face, processes the image, and identifies the user's emotion using a trained machine learning model. Based on this emotion, it suggests a list of songs that match the mood. The user can then select and play a song and even stop it whenever they want. This system works smoothly and quickly, making it feel natural, like talking to a friend who knows your mood and plays music accordingly. Overall, this project combines the power of deep learning with web development to offer a personalized and emotion-aware experience. It shows how machines can understand human feelings and respond in meaningful ways.

VI. FUTURE SCOPE

It can be used to enhance the systems to detect emotions continuously from a live webcam stream instead of capturing a single image.Connect with platforms like Spotify or YouTube via APIs to recommend and stream a broader and more dynamic range of songs.Train the emotion detection model on more comprehensive datasets to improve recognition across different facial types, lighting conditions, and demographics.

VII.REFERENCES

[1] https://www.researchgate.net/publication/354855186_Music_Recommendation_Based_on_Face_Emotion_Recognition.

[2] Amya Ramanathan, Radha Kumaran, Ram Rohan R, Rajat Gupta, and Vishalakshi Prabhu, *An intelligent music player based on emotion recognition*, 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions, 2017.

[3] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, Emotion-based music recommendation system using wearable physiological M. Athavle et al. ISSN (Online) : 2582-7006 International Conference on Artificial Intelligence (ICAI-2021) 11 Journal of Informatics Electrical and Electronics Engineering (JIEEE) A2Z Journals sensors, IEEE transactions on consumer electronics,vol.14,no.8,May 2018.https://doi.org/10.1109/TCE.2018.2844736

[4] AYUSH Guidel, Birat Sapkota, Krishna Sapkota, Music recommendation by facial analysis, February 17, 2020..

[5] Gupte A, Naganarayanan A and Krishnan M Emotion Based Music Player-XBeats International Journal of Advanced Engineering Research and Science 3 23685

[6] Ramya Ramanathan, Radha Kumaran, Ram Rohan R, Rajat Gupta, and Vishalakshi Prabhu, an intelligent music player based on emotion recognition, 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions2017. https://doi.org/10.1109/CSITSS.2017.8447743

[5] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 196–203, May 2018. https://doi.org/10.1109/TCE.2018.2844736

[6] S. J. Shabu, C. Janaardhan, K. Bhaskar, A. V. A. Mary, J. Refonaa, and S. Dhamodaran, "Music Recommendation System based on Facial Expression," *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2023, pp. 908–912. https://doi.org/10.1109/ICESC56218.2023.10198745

[7]] A. Alrihaili, A. Alsaedi, K. Albalawi, and L. Syed, "Music recommender system for users based on emotion detection through facial features," *2019 12th International Conference on Developments in eSystems Engineering (DeSE)*, Kazan, Russia, 2019, pp. 1014–1019. https://doi.org/10.1109/DeSE.2019.00188