Mood-Based Song Recommendations with Facial Recognition

Music plays a vital role in our everyday life. Life without music cannot be imagined. Music changes our mood; Whatever our mood might be, the only thing we do in all of our moods is to listen to music. We also listen to music when working, driving, travelling and even when reading a comic or a story. Music can induce a clear emotional response in its listeners. The pitch and rhythm of the music are managed in the areas of the brain that deal with emotions and mood. Thus, music plays an important role in enhancing our mood. As elders have said “Face is the Index of the Mind”, the mood of a person can be known by looking at the face of the person. The abstract of this system/ project is to build an automated system that builds playlists and plays the songs according to the mood of the user by directly discerning the facial emotions of the user. This model requires a camera to capture the face of the user and then the mood of the user is recognized by CNNs. Then the playlist is recommended to the user based on the discerned “Mood” of the user. This disposes of the tedious and monotonous task of physically gathering tunes into various records and helps in creating a suitable playlist dependent on a person's passionate highlights. Hence, the proposed system can be used to build a music recommendation system based on the facial emotion gestures of the user.

INTRODUCTION

Music prompts a reasonable passionate reaction in its audience. Melodic inclinations have been exhibited to be exceptionally associated with character qualities and mind-sets. Facial emotions are the most common and natural methods of passing on feelings, temperaments and sentiments. Convolutional Neural network, as a Deep Learning Neural Network, assumes a critical part in face image recognition. Cognition technology of CNN and Music Recommendation System based on Facial Emotion Gestures is created to distinguish a model that perceives facial articulations and prescribes music as indicated by comparing mind-set of the user or client. Human beings have the innate capacity to see somebody's face and conjecture their mind-set. This capacity if learnt by an electronic gadget - computer, humanoid robot or a mobile gadget - can have important applications in reality. Music, an instrument for stirring emotions and feelings, is undeniably more remarkable than language. Music is

something which takes advantage of our emotional centre as human beings [1]. Accordingly, paying attention to good music can assist us with lifting our mind-set from a negative sense to a positive sense. For example, focusing on lively tunes when the individual is feeling grim can assist him with arising his difficulty and start feeling better. This framework proposes one such application, emotion-based music recommendation. Emotion of the client can be effortlessly speculated by taking a gander at his/her face. For this reason, face detection and emotion recognition, examining the fiducial highlights from his/her face is essential.

The issues related with face detection incorporate foundation components, lighting conditions, posture and facial demeanor. This space of face detection and emotion detection is as of now a functioning space of examination because of advancement of Virtual Reality and Augmented Reality. Constant face detection and recognition frameworks have restricted usefulness because of the fluctuating nature of pictures as a result of
the issues related like foundation, enlightenment, and so on Thus, innovative work for arrangements identified with these issues is a continuous work. Using regular music players, a client expected to actually mastermind his playlist and select tunes that would diminish his/her attitude and energetic experience. This task was work genuine and an individual every now and again went up against the trouble of showing up at an appropriate once-over of songs. Different frameworks which recognize the disposition of the client by utilizing facial appearance have their time and memory intricacy generally high and subsequently flop in accomplishing an ongoing presentation. Regardless of whether they perceive the temperament of the client then their choice of melodies for making a playlist is with the end goal that it will simply pick tunes mirroring the current mind-set of the client and won't attempt to improve his mind-set in any capacity. In this way, if the client is dismal, In the current frameworks, client is furnished with a rundown of melodies with pitiful emotion which can corrupt his/her mind-set further and can prompt misery. Along these lines, the framework proposed will distinguish the emotion of the client from his facial articulations. It will then, at that point furnish the client with a playlist of melodies, paying attention to which the client will feel good.

Convolutional Neural Network

A CNN gets a picture as a contribution to the type of a 3D Matrix. The underlying two measurements contrast with the width and height of the image in pixels while the third one identifies with the RGB potential gains of each pixel. CNNs comprises of the accompanying successive modules (every one may contain more than one layer)

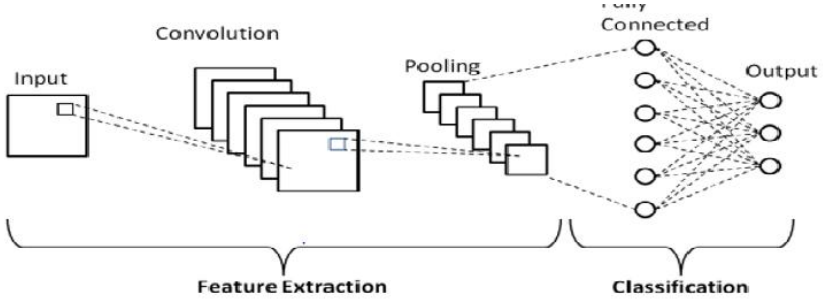
Convolution

ReLu activation function

Pooling

Fully connected layers

Output layer



Convolution Layer:

The part connected with doing the convolution movement in the underlying portion of a Convolutional Layer is known as the Kernel/Channel. Convolution activity is a component savvy network increase activity. Convolutional layers take the three-dimensional information framework and they pass a channel (otherwise called convolutional channel) over the image, applying that to a little window of pixels at the same time (for instance, 3x3 pixels) and this window, being moved until the entire picture has been separated. The convolutional action registers the dab consequence of the pixel regards in the current channel window close by the heaps described in the channel. The yield of this movement is the last tangled picture. The focal point of picture request CNN's is that as the model trains what it really does is that it learns the characteristics for the channel matrices that enable it to remove huge features (shapes, surfaces, concealed districts, etc) in the image. Each convolutional layer applies one

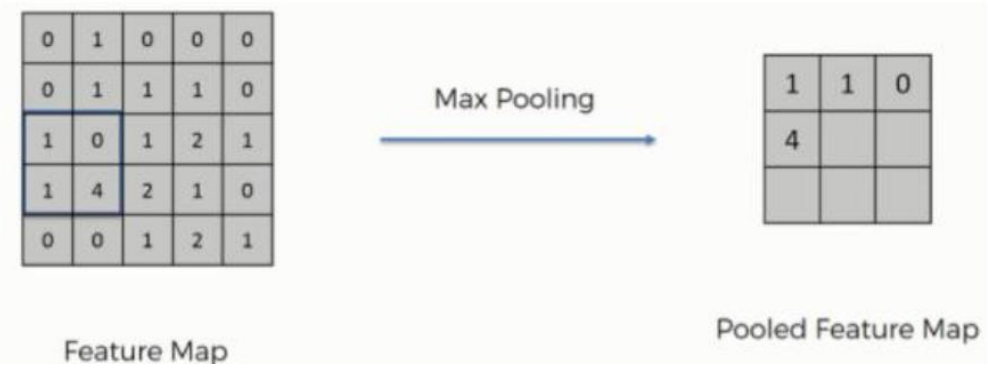
new channel to the tangled image of the past layer that can eliminate one more part. Accordingly, as more channels are stacked, the more features the CNN can remove from an image.

The three components that go into the convolution activity are:

Input image

Feature detector

Feature map



METHODOLOGY

The methodology for the **Mood-Based Song Recommendations with Facial Recognition** project involves integrating facial emotion recognition with a song recommendation system. The process begins with the collection and preparation of data. A pre-trained emotion detection model, such as FER (Facial Emotion Recognition), is utilized to identify emotions like Happy, Sad, Angry, and Neutral from facial expressions. Alternatively, a custom model can be trained using datasets like FER2013, which consists of labeled facial images. The images are preprocessed by resizing them to 48x48 grayscale format and normalizing pixel values to ensure compatibility with the model.

For real-time emotion detection, the Haar Cascade classifier is employed to detect faces in the video stream. The detected face is then cropped, converted to grayscale, resized, and fed into the emotion detection model, which predicts the user's mood based on facial expressions. Simultaneously, a song recommendation dataset is created, mapping each mood to a curated list of songs. This dataset serves as the basis for selecting songs relevant to the detected emotion.

The system operates in real-time, capturing live video feed through a webcam. Each frame is processed to detect faces and analyze emotions. Once an emotion is detected, it is mapped to the corresponding mood in the song dataset, and a list of recommended songs is displayed. The interface shows the detected emotion on the video frame, while song recommendations are printed in the console or displayed in the UI. This ensures an interactive user experience, where the recommendations update dynamically based on the user's mood.

Finally, the project can be enhanced by incorporating additional features like integrating streaming services such as Spotify to play songs, expanding the dataset for

more diverse moods, or training a more sophisticated deep learning model for improved accuracy. This methodology ensures a practical and engaging implementation of real-time emotion-based song recommendations.

Code Implementation

Step - 1 : Install Necessary Libraries
Step - 2 : Load Pre-trained Emotion Detection Model
Use a pre-trained model such as FER (Facial Emotion Recognition).

```
import cv2
import numpy as np
from keras.models import load_model
import pandas as pd

# Load pre-trained emotion detection model
model = load_model('emotion_detection_model.h5')  # Replace with your model file
emotion_labels = ['Angry', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', 'Neutral']
```

```
# Load Haar Cascade for face detection
face_classifier = cv2.CascadeClassifier(cv2.data.haarcascades + 'haarcascade_frontalface_default.xml')
```

Step - 3 : Create a Song Recommendation Dataset

Prepare a dataset (CSV file) with songs for each mood.

```
# Example song dataset
data = {
    'Mood': ['Happy', 'Sad', 'Angry', 'Neutral'],
    'Songs': [
        ['Happy Song 1', 'Happy Song 2', 'Happy Song 3'],
        ['Sad Song 1', 'Sad Song 2', 'Sad Song 3'],
        ['Angry Song 1', 'Angry Song 2', 'Angry Song 3'],
        ['Neutral Song 1', 'Neutral Song 2', 'Neutral Song 3']
    ]
}
song_df = pd.DataFrame(data)
```

Step - 4 : Real-Time Facial Recognition and Song Recommendation

```
def recommend_songs(emotion):
    """Recommend songs based on emotion."""
    for mood in song_df['Mood']:
        if mood == emotion:
            return song_df[song_df['Mood'] == mood]['Songs'].values[0]

# Open webcam
cap = cv2.VideoCapture(0)
```

```
while True:
    ret, frame = cap.read()
    if not ret:
        break

    gray_frame = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
    faces = face_classifier.detectMultiScale(gray_frame, scaleFactor=1.3, minNeighbors=5)
```

```
    for (x, y, w, h) in faces:
        cv2.rectangle(frame, (x, y), (x+w, y+h), (255, 0, 0), 2)
        roi_gray = gray_frame[y:y+h, x:x+w]
        roi_gray = cv2.resize(roi_gray, (48, 48), interpolation=cv2.INTER_AREA)

        if np.sum([roi_gray]) != 0:
            roi = roi_gray.astype('float') / 255.0
            roi = np.expand_dims(np.expand_dims(roi, axis=-1), axis=0)

            predictions = model.predict(roi)
            emotion_index = np.argmax(predictions)
            emotion = emotion_labels[emotion_index]

            # Display emotion
            cv2.putText(frame, emotion, (x, y-10), cv2.FONT_HERSHEY_SIMPLEX, 1, (0, 255, 0), 2)

            # Recommend songs
            songs = recommend_songs(emotion)
            print(f"Detected Mood: {emotion} | Recommended Songs: {songs}")
        else:
            cv2.putText(frame, 'No Face Detected', (20, 60), cv2.FONT_HERSHEY_SIMPLEX, 1, (255, 0, 0), 2)

    cv2.imshow('Mood-Based Song Recommender', frame)

    # Break loop on 'q' key press
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break
```

```
cap.release()
cv2.destroyAllWindows()
```

Conlusion :

The Mood-Based Song Recommendation System with Facial Recognition is an innovative and interactive application that combines facial emotion recognition with personalized song suggestions. By utilizing a pre-trained emotion detection model and a curated song dataset, the system effectively identifies the user's mood in real-time and recommends songs that resonate with their emotions. This project demonstrates the potential of combining computer vision and machine learning to enhance user experiences in entertainment and personalization.

The system's real-time functionality, affordability, and scalability make it a versatile solution for various applications, including music streaming platforms, mental health tools, and personalized entertainment systems. Additionally, the project highlights the importance of integrating user-centric technologies like facial recognition to create more engaging and intuitive interfaces.

This implementation serves as an excellent foundation for further enhancements, such as integrating live music playback, adding more moods and song categories, or refining emotion detection using advanced deep learning models. Ultimately, this project bridges technology and creativity, showcasing how artificial intelligence can enrich everyday experiences.