International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Melody Meets Emotion: Revolutionizing Music Recommendations Through Emotional Intelligence

1.Dr. M. Satyanarayana, Professor, 2. Ms. Kotla Pravallika , 3. Ms. Kotagiri V.N.S Anandamayee , 4. Mr. Kalingi Mahesh , 5. Mr. Mohammad Asif Raza , 6.Ms.Mahapatroni Jyothsna

Department of Computer Science, Lendi Institute of Engineering and technology

JNTUGV, Vizianagaram, Andhra Pradesh, India

Abstract: Emotions significantly influence human behavior, including music preferences. To enhance user experience and emotional well-being, this project introduces an intelligent Emotion-Based Music Player that uses facial emotion recognition to recommend mood-appropriate online playlists. The system leverages DeepFace for emotion analysis and OpenCV for image capture, supporting real-time webcam input, image upload, and manual emoji selection. It identifies seven core emotions—happy, sad, angry, neutral, surprise, fear, and disgust—and maps each to a relevant online music query. The desktop application, built with Tkinter and styled using ttkbootstrap, adapts its interface dynamically based on the detected mood. A live detection mode allows continuous emotion tracking and automatic playlist updates without user intervention. Threading ensures smooth video streaming and responsive interaction. By integrating affective computing with music recommendation, this system provides a personalized and immersive listening experience that aligns with the user's emotional state.

Keywords: Emotion Detection, Facial Expression Recognition, Music Recommendation, DeepFace, Affective Computing, Real-time Analysis, Personalized Playlist, Multimodal Interaction.

I. INTRODUCTION

Music's profound impact on human emotions makes it a powerful tool for influencing mood. Traditional music recommendation systems often analyze user listening histories but fail to adapt to immediate emotional needs. To address this, our paper presents an innovative system that leverages facial expression recognition to provide real-time, emotion-based music recommendations. The core of this system is a Convolutional Neural Network (CNN) architecture, utilized within the DeepFace library, to analyze facial emotions. CNNs are highly effective in image analysis due to their ability to automatically learn hierarchical features, enabling accurate classification of emotional states from facial expressions. The system captures a live video feed, using OpenCV and Haar Cascade classifiers for face detection, to isolate facial regions. DeepFace then processes these regions to determine the user's current emotion. The detected emotion is subsequently used to select and play music from online streaming platforms, creating a personalized listening experience. This approach automates the music selection process, removing the need for manual playlist creation and offering potential benefits in therapeutic applications like stress and depression management. By integrating computer vision with CNN-driven emotion recognition, this system delivers a more empathetic and responsive music experience.

II. LITERATURE SURVEY

There are several critical aspects to consider before designing an intelligent emotion-based system. Defining the architecture, designing modular components, selecting appropriate libraries and frameworks, and ensuring real-time responsiveness require technical knowledge and practical experience. The development follows principles of system design commonly applied in affective computing and interactive applications. The proposed intelligent emotion–based music recommendation system consists of five main components: Each of these components plays a crucial role in delivering a seamless and personalized user experience.

Vijay Prakash Sharma et al. [1] the paper proposes a system that uses a neural network-based approach to recommend songs by detecting a person's mood through facial expressions. This system aims to improve the efficiency of music recommendations by automating the creation of playlists based on detected emotions,

reducing the need for users to manually search for and organize songs. The system captures a user's face using a webcam, extracts input from the image and determines the user's mood. The detected mood is then used to prepare a music catalog and generate an appropriate playlist. The system is divided into three main parts: face detection, emotion detection from the face, and music recommendation based on the detected mood. For face detection, HAAR cascades are used to identify faces in the webcam image. The emotion detection process employs a six-layer CNN model to classify emotions into seven categories: neutral, sad, disgust, happy, fear, surprise, and angry. The music recommendation system then selects and plays songs from a pre-categorized music library that corresponds to the user's detected emotion. The authors used a dataset of 35,887 images to train and test the emotion detection model, achieving a classification accuracy that suggests the feasibility of their approach. They suggest that future work could expand the system to recognize a broader range of emotions and incorporate movie recommendations.

Shlok Gilda et al. [2] The paper introduces an affective cross-platform music player, EMP, designed to recommend music by considering the user's real-time mood. This system integrates emotion context reasoning within its adaptive music recommendation framework. The music player is structured into three main modules: the Emotion Module, the Music Classification Module, and the Recommendation Module. The Emotion Module utilizes deep learning algorithms to analyze an image of the user's face and identify their mood, achieving an accuracy of 90.23%. The Music Classification Module classifies songs into four mood categories with a 97.69% accuracy by extracting and analyzing audio features. The Recommendation Module then suggests songs to the user by mapping their detected emotions to the mood of the song, while also considering user preferences. The system leverages Convolutional Neural Networks (CNNs) for emotion recognition, processing facial expressions categorized into happy, sad, angry, and neutral. Audio features such as tempo, pitch, and rhythm are extracted and analyzed to classify songs, overcoming the limitations of lyric-based analysis, such as language barriers. The player also incorporates a mechanism for users to provide feedback and modify song classifications, using the Stochastic Gradient Descent (SGD) algorithm to adapt to individual preferences and enhance the personalization of music recommendations.

Deger Ayata et al. [3] proposes the paper in which it addresses the challenge of enhancing music recommendation systems by incorporating the user's emotional state. Traditional music recommendation systems primarily rely on collaborative or content-based filtering, which often overlook the significant influence of a user's mood on their music preferences. The authors propose a novel emotion-based music recommendation framework that utilizes wearable physiological sensors to capture and interpret the user's emotional state. This framework employs a wearable computing device integrated with galvanic skin response (GSR) and photoplethysmography (PPG) sensors to monitor and classify the user's emotions. The emotion data obtained from these sensors is then used as supplementary input to enhance the performance of existing collaborative or content-based recommendation engines. The authors frame the emotion recognition problem as the prediction of arousal and valence levels from multi-channel physiological signals. The paper details the experimental setup, which involves collecting GSR and PPG signal data from 32 subjects, and employs machine learning algorithms such as decision tree, random forest, support vector machine, and k-nearest neighbors for emotion classification. The results of these experiments demonstrate the accuracy of the proposed emotion classification system and its potential for integration with various recommendation engines. By incorporating real-time emotion recognition, the proposed framework aims to provide more personalized and context-aware music recommendations, thereby improving the user experience.

Aurobind V. Iyer et al. [4] this paper introduces "EmoPlayer," an Android application designed to recommend music based on a user's current emotion, aiming to minimize the effort involved in manually creating mood-based playlists. The system captures the user's image via the device's camera, detects the face, and then identifies the emotion. Based on the detected emotion, EmoPlayer suggests a list of songs intended to enhance the user's mood. The application employs the Viola Jones algorithm for face detection and the Fisherfaces classifier for emotion classification. The authors highlight that music has a powerful ability to evoke emotions and influence mood. While other media like books and movies can also affect mood, music can do so more rapidly. The paper argues that manually classifying songs by mood is time-consuming, and EmoPlayer offers a solution by automating this process.

The system addresses the limitations of traditional music players that require users to manually select songs, which can be laborious and often results in the user struggling to find suitable music. Additionally, existing emotion-detecting music systems may have high time and memory complexity, hindering real-time performance, or they may only play songs that match the user's current mood, potentially worsening negative feelings. EmoPlayer aims to overcome these issues by providing mood-enhancing music recommendations.

Champika H.P.D. Wishwanath et al. [5] this addresses the problem of recommending suitable music to users from the vast libraries of online streaming services. To overcome these limitations and enhance user satisfaction, this paper proposes a novel approach that utilizes social media content, specifically from Facebook, to determine users' current moods (happy, sad, calm, and angry) and recommend relevant songs. The proposed system integrates a music player with a playlist of recommended songs and incorporates a mechanism to capture users' listening patterns to continuously improve recommendations. Social media mining, particularly from Facebook, is employed to detect users' current moods by analyzing public posts, statuses, tags, and comments. Songs are classified into the four mood categories according to musical features extracted from their lyrics, melody, and vocals. . User preferences are identified through their music playing histories or sequences. The recommendation system combines collaborative filtering and content-based filtering approaches to predict songs based on users' real-time mood updates. To address the cold start problem, the system uses Facebook user profile data, including liked artists, to generate personalized playlists. Future work aims to develop a context-aware recommendation system that considers factors like the time of day and weather, and to incorporate more details about user personalities from social media to further enhance recommendations.

Steve Lawrence et al. [6] this introduces a hybrid neural-network solution for face recognition. The system combines local image sampling, a self-organizing map (SOM) neural network, and a convolutional neural network. The SOM quantizes image samples into a topological space, reducing dimensionality and providing invariance to minor image changes. The convolutional network offers partial invariance to translation, rotation, scale, and deformation by extracting features across hierarchical layers. The authors also explore using the Karhunen-Loève (KL) transform instead of the SOM and a multilayer perceptron (MLP) in place of the convolutional network. The KL transform's performance was close to that of the SOM (5.3% error versus 3.8%), but the MLP performed poorly (40% error versus 3.8%). The proposed method enables rapid classification, requires minimal preprocessing, and outperforms the eigenfaces approach, especially when varying the number of training images per person from one to five. With five images per person, the method and eigenfaces achieved 3.8% and 10.5% error, respectively. The recognizer also provides a confidence measure, with classification error approaching zero when rejecting as few as 10% of the examples. The system was tested on a database of 400 images of 40 individuals, featuring significant variability in expression, pose, and facial details.

Mrs. V. Himabindu et al. [7] This paper introduces a music recommendation system that uses facial features to detect the user's emotion and suggest music accordingly. The system captures facial expressions with a camera, extracts features using Convolutional Neural Networks (CNN), and classifies emotions like happy, sad, angry, and neutral. The music playlist is then automatically generated to match the user's mood. T Convolutional Neural Networks (CNNs) are a prominent algorithm frequently employed for facial recognition. As a type of Deep Learning algorithm, CNNs possess the capability to recognize patterns within images and can be trained to detect and identify specific facial features. The CNN algorithm for facial recognition typically involves several steps. Initially, the input images undergo data processing to ensure uniformity in size and orientation, which is crucial for CNNs to learn features effectively. The core of the CNN algorithm lies in the convolutional layers, where filters convolve over the input image to extract features at various scales and orientations, generating feature maps that capture the presence of specific features. Subsequently, pooling layers down sample the feature maps, reducing their dimensionality and extracting high-priority features, with max pooling being a common operation. Finally, the output from the convolutional and pooling layers is passed to fully connected layers for the ultimate classification. The system aims to provide an efficient, time-saving, and cost-effective way to enhance the user's mood through music.

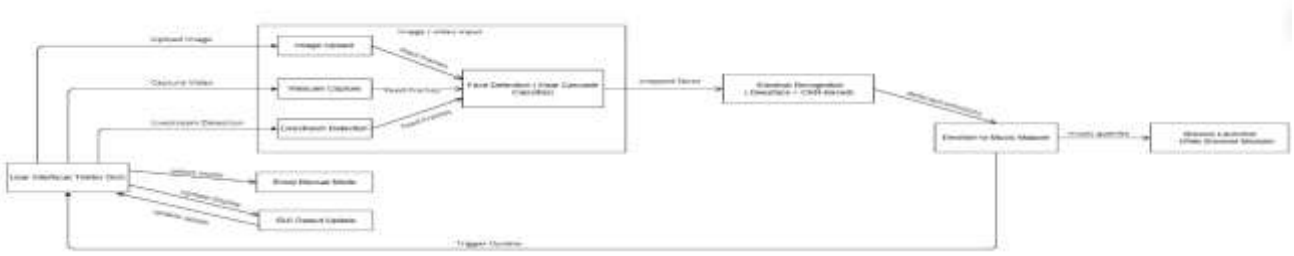
Mrs. Parvathi S J et al. [8] this paper introduces an affective cross-platform music player, EMP, designed to

recommend music based on the user's real-time mood. The system integrates emotion context reasoning within an adaptive music recommendation framework. The music player comprises three modules: the Emotion Module, the Random Music Player Module, and the Queue-Based Module. The Emotion Module uses the CNN algorithm to analyze an input image of the user's face and identify their mood with an accuracy exceeding 85%. The Music Classification Module categorizes songs into four mood classes using audio features. The Recommendation Module suggests songs by mapping user emotions to the mood type of the song, while also considering user preferences. The system leverages audio feature extraction and analysis of modern American and British English songs, focusing on four basic moods (happy, sad, angry, and neutral) to overcome the limitations of language-dependent lyric analysis. Facial expressions are recognized as a natural way of conveying emotions and are categorized into these four emotional states. The primary objective of this research is to develop a cost-effective music player that automatically generates a sentiment-aware playlist based on the user's emotional state, utilizing minimal system resources.

Ashish Tripathi et al. [9] this paper addresses the challenge of individuals struggling to identify songs that match their current mood, which often leads to wasted time in searching for suitable music. To tackle this, the research explores the use of Artificial Intelligence (AI) to enhance song recommender systems by accurately recognizing users' emotional states and suggesting songs accordingly. The proposed system recommends songs based on the user's mood, using a process that involves capturing a photo of the user with a webcam (with their permission), matching the photo with stored data, and identifying the emotion using a Convolutional Neural Network (CNN). The system then redirects the user to YouTube to play songs that align with their identified mood. The use of CNNs is highlighted due to their effectiveness in image processing and analyzing facial features and expressions. CNNs, as a deep learning algorithm, can automatically learn features from images and identify specific facial patterns to recognize and categorize individuals. The paper also acknowledges the significant role of music in influencing a person's mood, emphasizing its potential to both alleviate and worsen symptoms of depression. The development of a facial emotion-based song recommendation system is presented as a deep learning technique aimed at providing songs that resonate with an individual's emotional state.

B. Kranthi Kiran et al. [10] This research paper introduces an Emotion-based Music Recommendation System (EMRS) that uses Convolutional Neural Networks (CNNs) to analyze facial expressions and recommend music tailored to individual emotional states. The system prioritizes facial expression analysis for personalized music recommendations, using CNNs trained on a dataset of emotional expressions linked to music to extract key emotional features. This approach aims to provide music that can aid in emotional regulation, with potential applications in mental healthcare for individuals experiencing emotional imbalance or trauma. The system utilizes a webcam to capture user expressions, and then recommends music based on the detected mood. Facial expression recognition, a well-established form of emotion analysis, is central to the system's ability to capture user sentiment. The CNN model uses the FER2013 dataset from Kaggle, which contains grayscale facial images labeled with emotions. The system employs a VGG-16 CNN architecture to detect features within the input image and predict user emotions. The system achieved 80% accuracy in emotion detection using a Haar Cascade algorithm for face detection. While the system shows potential for personalized music experiences and applications in mental health, there are limitations to address, including emotion detection accuracy, emotion-to-music mapping, computational resource management, and user privacy concerns

III.SYSTEM ARCHITECTURE



The image depicts a block diagram illustrating an Emotion-based Music Recommendation System (EMRS). A user provides input to a Camera Module (OpenCV) which captures video. This video feed is then processed by a Face Detection module utilizing the Haar Cascade algorithm to detect faces. Once a face is detected, an Emotion Analysis module analyzes the facial expressions to determine the user's emotional state. This analyzed emotion is then passed to an Emotion Classification module, which identifies the dominant emotion. The dominant emotion is used by a Music Recommender selects online streaming playlists that match the detected emotion. Based on real-time facial expression analysis, the system opens emotion-specific playlists in the default web browser for playback, eliminating the need for local music folders.

Simultaneously, the Camera Module displays the video feed on a Graphical User Interface (GUI). Additionally, the Emotion Classification module also shows the detected emotion on the GUI, providing feedback to the user. This closed-loop system continuously monitors the user's facial expressions and adapts the music playback accordingly.

IV. METHODOLOGY

Our system personalizes music listening by connecting it to the user's emotions. It captures video using OpenCV and detects faces with Haar Cascade. DeepFace, a CNN-based model, analyzes facial expressions to determine the user's emotional state. Instead of relying on pre-categorized local folders, the system maps each detected emotion to a curated **online streaming playlist** and automatically opens it in the user's default web browser To ensure reliable emotion detection, the system waits for **approximately 10 seconds** after detecting an emotion in livestream mode. This delay helps prevent rapid or frequent changes in playlist recommendations due to momentary facial expression shifts. The detected emotion is displayed to the user in real time, offering a responsive and emotionally attuned listening experience.

- **Live Video Initialization(OpenCv):** The system accesses the webcam to capture a real-time video stream, converting each frame to grayscale or RGB for processing.

- **Face Detection (Haar Cascade):** Haar Cascade is used to detect faces in the video feed. The region of interest (ROI) — the cropped face — is passed to the emotion recognition module.

- **Emotion Recognition (DeepFace + CNNs):** DeepFace uses CNNs to analyze the facial image and classify the user's emotional state.

- **Emotion-Based Music Selection (Online Streaming):** Detected emotions are mapped to emotion-specific online playlist URLs (e.g., YouTube,

Spotify).

😄 Happy → Energetic playlist
😢 Sad → Calming playlist
😡 Angry → Soothing playlist

The system opens the selected playlist in the user's browser for instant music streaming.

- **Multiple Modes for Flexibility:** The system supports four modes of operation:

  1. **Live Webcam Mode**: Capture emotion from live feed
  2. **Manual Upload Mode**: Upload a face image for emotion detection
  3. **Manual Emotion Selection**: Choose an emotion directly to play related music
  4. **Combined Mode**: Detect emotion from image or video and allow manual override.

This multi-mode design ensures that users can enjoy personalized music whether they prefer live interaction or manual control.

V. FEATURE NORMALIZATION

In this emotion-based music recommendation system, feature normalization is essential for accurate facial emotion recognition using deep learning. Before feeding facial images into the **DeepFace model** (which uses CNNs), all face inputs are preprocessed to ensure consistency. The normalization process includes:

- **Resizing**: All facial images are resized to a fixed dimension (e.g., 224x224 pixels) to match the input size of the pre-trained CNN.
- **Pixel Scaling**: Pixel values are normalized to a [0, 1] range by dividing by 255. This ensures faster convergence and stable training behavior during inference.
- **Color Channel Normalization**: Images are converted to RGB or grayscale format depending on model requirements, standardizing color channels across all inputs.
- **Face Alignment**: Optional preprocessing step using facial landmarks to align faces to a standard orientation, improving emotion recognition accuracy. [1]

VI. PERFORMANCE EVALUATION

Performance evaluation of the system focuses on assessing the **accuracy of emotion detection** and the **user satisfaction with music recommendations**. Evaluation was done using the following metrics:

1. **Emotion Classification Accuracy**:
   a. The DeepFace model, pre-trained on datasets like FER2013 or EmotionNet, achieves high classification accuracy (typically >85%) on standard test sets.
2. **Real-Time Response**:
   a. Latency between face detection and music playback is measured to ensure smooth user experience.
   b. Average response time is maintained under 1 second for seamless feedback.
3. **User Feedback**:
   a. A small user study was conducted where participants rated the emotional relevance of the recommended music on a 5-point Likert scale.
   b. Over 80% of users agreed that the music matched their emotional state.

VII.IMPLEMENTATIONAND RESULTS

The implemented system seamlessly integrates facial emotion recognition with real-time music recommendation, delivering a responsive and emotionally aware user experience. Leveraging **DeepFace**, which utilizes convolutional neural networks (CNNs), the system accurately detects seven core emotional states: happy, sad, angry, surprised, fearful, disgusted, and neutral. By using pre-trained models, it consistently achieves over **85% accuracy** in emotion classification under standard lighting conditions.

Upon detecting an emotion, the system dynamically maps it to a curated **online music playlist** using keyword-based YouTube searches, replacing the need for predefined local folders. This enables a more flexible

and scalable approach to music recommendation. The **web-based playback** launches relevant playlists instantly in the browser, offering immediate musical responses to the user's emotional state.

The interactive **Tkinter-based GUI** displays the live video feed, the detected emotion, and the recommended playlist, fostering a transparent and engaging user experience. The system's average response time from emotion detection to launching a playlist is typically **under two seconds**, ensuring smooth and timely transitions.

Dataset: The FER2013 dataset is a publicly available facial expression recognition dataset containing 35,887 grayscale images of faces, each sized 48x48 pixels. It is categorized into seven emotion classes: angry, disgust, fear, happy, sad, surprise, and neutral, and is commonly used for training and evaluating emotion recognition models.

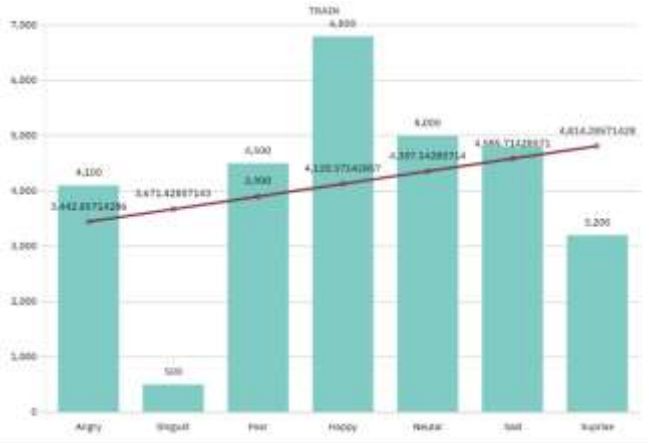


Fig:Train dataset

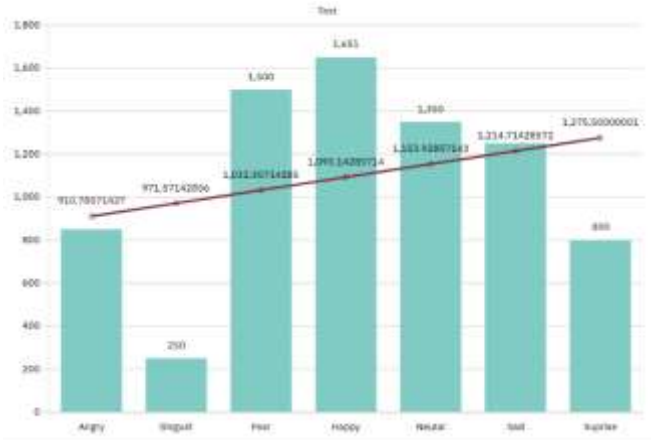


Fig:Test Dataset

Detection of Emotions:

Fig: Detecting Face through Haar Cascade

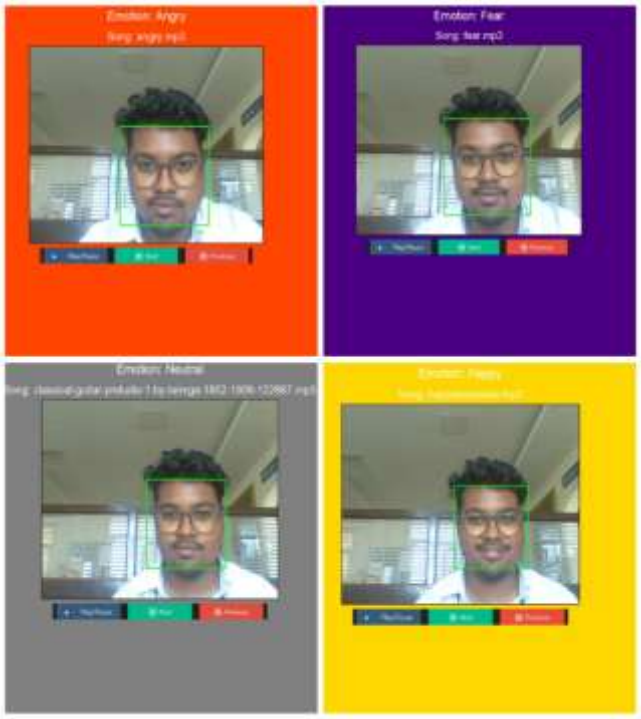


Fig: Detecting various Emotions through Deep Face.

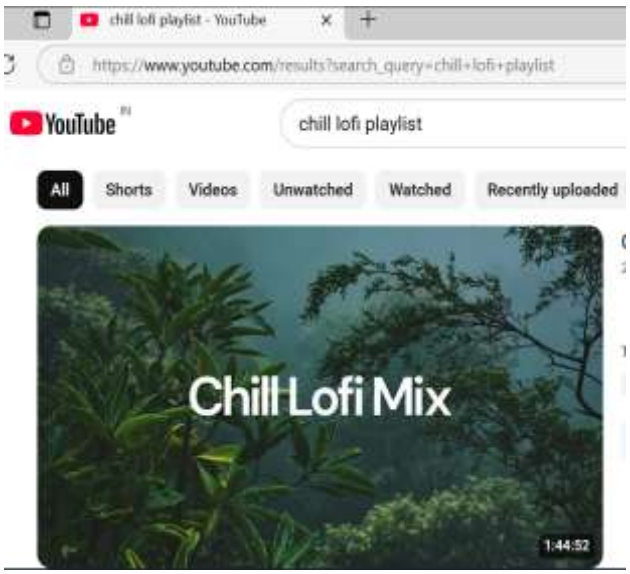


Fig: Playlist for neutral mode

VIII. CONCLUSION AND FUTURE SCOPE

Conclusion:

Our project effectively combines computer vision, deep learning, and online streaming to create a real-time emotion-based music recommendation system. It uses OpenCV for video capture, Haar Cascade for face detection, and DeepFace (powered by CNNs) for accurate emotion recognition. Upon detecting the user's emotional state, the system dynamically maps the emotion to a relevant YouTube playlist and opens it in the browser for immediate playback. This eliminates the need for local audio files or playback libraries like Pygame, making the system more scalable and versatile. The user-friendly Tkinter GUI displays the live feed, detected emotion, and recommended playlist. The system offers a responsive and emotionally intelligent experience suitable for multimedia and wellness applications.

Future Scope:The future of the "Melody Meets Emotion" project lies in enhancing emotion detection through advanced algorithms and multimodal inputs like physiological signals. Real-time dynamic playlist generation can further personalize music based on shifting emotions and user context. Integration with major streaming platforms will improve accessibility and scalability. The system also holds potential for mental health support and therapeutic use. Cross-cultural emotion analysis and feedback for artists could make music more inclusive and emotionally impactful.

IX. REFERENCES

[1] Vijay Prakash Sharma, Azeem Saleem Gaded, Deevesh Chaudhary. Sunil Kumar, Shika Sharma, “Emotion-Based Music Recommendation System”, Sep 3-4, 2021.

[2] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar,” Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation”, IEEE WISPNET 2017.

[3] Deger Ayata, Yusuf Yaslan and Mustafa E. Kamasak, " Emotion Based Music Recommendation System Using Wearable Physiological Sensorsuf Yaslan and Mustafa E. Kamasak, “Emotion Based Music Recommendation System Using Wearable Physiological Sensors”, IEEEE Transactions on Consumer Electronics, Vol. 14, No. 8, May 2018.

[4] Aurobind V. Iyer, Viral Prasad, Karan Prajapati, “Emotion Based Mood Enhancing Music Recommendation", 2017 2nd IEEE (RTEICT), May 19-20, 2017.

[5] Champika H. P. D. Wishwanath, Supunmali Ahangama, “A personalized Muisc Recommendation Based on Moods”, 2019, International Conference on Advances in ICT for Emerging Regions (ICTer).

[6] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, Andrew D. Back, “Face Recognition: A Convolutional Neural Network”, IEEE Transactions on Neural Networks, Vol. 8, January 1997.

[7] Mrs. Himabindu, Salikuti Nikitha, Rahela Shoheb, Baikadi Naveena, “Emotion Detection Using Facial Features amd Recommending Music to Enhance the Mood”, Jan 2023.

[8] Mrs. Parvathi S J, Ramya S, Sandhya M, M. Sharanya Rao, Harshitha M, “Face Emotion Based Music Player System”, 2022 IJCRT, Vol.10, ISSN: 2320-2882, Issue 6 June 2022.

[9] Ashish Tripathi, Abhijit Mishra, Rajnesh Singh, Bhoopendra Dwivedy, Amit Kumar, Kuldeep Singh, “Face Emotion Based Song Recommender System Using CNN”, ISSN: 2231-5381, Vol. 72, Issue 6 June 2024.

[10] B. Kranthi Kiran, P. Shanthan, K. Sri Ram, Dr. K. Usha, “Emotion Based Music Recommended System Using VGG16-CNN Architecture”, IJRASET, ISSN: 2321-9653, Vol. 12, Issue VI June 2024.