iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59432

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Music Recommendation System based on Facial Expressions by using CNN

Jeevan Babu Maddala[1], S. Pavithra[2], T. Vijay Kumar Reddy[3], V. Mounika[4], T. Maidhili[5]
Department of Computer Science & Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh

Abstract: ***In today's world, the integration of emotion detection and personalized music recommendation has become increasingly prevalent across diverse fields. This project endeavors to develop a cutting-edge music recommendation system that tailors playlists to users' emotional states, as discerned from their facial expressions in real-time. Leveraging the Facial Expression Recognition 2013 (FER 2013) dataset, which encompasses a wide array of emotional expressions including happiness, sadness, anger, fear, disgust, neutrality, and surprise, serves as the foundation for training our model. The project employs the Mediapipe algorithm for feature extraction from facial expressions and utilizes a Convolutional Neural Network (CNN) to classify these emotions accurately. For seamless integration and user interaction, streamlit web browser frameworks are harnessed to facilitate music recommendations within an intuitive interface. Furthermore, the proposed system capitalizes on real-time face detection through camera input using OpenCV, ensuring timely and responsive interaction. By automatically generating music playlists based on the user's current emotional state, our system aims to significantly reduce computational time and overall costs compared to existing literature. In essence, this project presents a comprehensive solution that combines real-time facial expression analysis with personalized music recommendation, offering enhanced efficiency and accuracy while minimizing computational overhead.***

Keywords: Facial Expression Recognition (FER 2013), Convolutional Neural Networks (CNN), Opencv, Camera, Mediapipe, Feature Extraction, Emotion Detection, Streamlit

I. INTRODUCTION

In today's digital realm, the fusion of technology and user-centricity has paved the way for personalized experiences that cater to individual preferences and needs. One such realm where personalization holds significant importance is in the domain of music streaming applications. We all appreciate the convenience of recommendations tailored to our tastes, simplifying the process of discovering new music while enjoying our favorite tunes. However, what if these recommendations could transcend mere preferences and adapt dynamically to our current mood? Understanding the significance of mood in shaping our music preferences, this project endeavors to develop a Facial Expression-Based Dynamic Music Recommendation System. By harnessing the power of facial expression recognition, our system aims to decipher the user's emotional state in real-time and recommend songs that align with their mood. The importance of this project lies in its ability to revolutionize the way we interact with music streaming applications. Rather than relying solely on explicit user preferences or historical data, our system taps into the subtle cues conveyed through facial expressions to provide personalized recommendations. This not only enhances user engagement but also fosters a deeper emotional connection with the music being listened to.

In daily life, our Facial Expression-Based Dynamic Music Recommendation System offers immense utility and convenience. Imagine starting your day with an uplifting dynamic playlist tailored to your morning mood, transitioning seamlessly to calming tunes during work breaks, and winding down with soothing melodies in the evening. Whether you're feeling joyful, stressed, or contemplative, our system ensures that your music selection complements your emotional state, enhancing your overall well-being and productivity throughout the day. In this article, we embark on a journey to explore the development and implementation of our Facial Expression-Based Dynamic Music Recommendation System. We delve into the methodology behind facial expression recognition, the utilization of the Facial Expression Recognition 2013 (FER 2013) dataset, and the integration of this technology into music streaming applications. Furthermore, we examine the potential implications and benefits of this innovative approach in enhancing user experience and satisfaction. Through this project, we aim to showcase the transformative potential of leveraging emotional intelligence in music recommendation systems, ushering in a new era of personalized and emotionally resonant music experiences for users worldwide.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

All Recent studies have delved into the integration of facial expression recognition technology within music recommendation systems and musical therapy interventions, showcasing promising advancements in many domains.

1) Gokul Krishnan K, Parthasarathy M, Sasidhar D, and Venitha E presented a project titled "Emotion Detection and Music Recommendation System Using Machine Learning," which introduces an automated music playlist feature that considers the user's emotion as a crucial factor. This project is designed as an Android application, leveraging the front camera of the user's mobile device to detect their emotion and incorporate it into music requests made to popular streaming services. Additionally, the application tracks the songs played to enhance future recommendations through learning from user preferences. The innovation lies in the complete automation of the process, allowing users to simply enjoy the music without any manual intervention. Notably, the detector ceases power usage once the user begins enjoying the music, optimizing energy consumption.

2) Arto Lehtiniemi and Jukka Holm introduced a concept titled "Animated Mood Picture in Music Recommendation," where users engage with a series of images to obtain music recommendations based on the genre associated with the pictures. Developed by Nokia Research Center, this innovative music system utilizes textual meta labels to identify genres and analyze audio content for processing.

3) Ramya Ramanathan et al's paper introduces an intelligent music player utilizing emotion recognition, emphasizing the significance of emotions in human interactions. Initially, local music selections are categorized based on the conveyed emotion, primarily considering song lyrics. The paper highlights the efficacy of emotion detection methodologies for developing emotion-based music players, proposing an algorithm for automatic playlist generation based on facial expressions to mitigate manual effort and time consumption. This algorithm aims to enhance system efficiency and accuracy while reducing computational costs, validated against user-dependent and user-independent datasets.

4) The effectiveness of emotion classification with SVM is evaluated using confusion matrices from Kohn-Kanade and FEEDTUM databases. Additionally, Kyogu Lee and Minsu Cho proposed a Mood Classification system from Musical Audio, employing user-dependent models and selecting adjectives related to music moods from USPOP and Last.fm data, which were clustered into three categories using PCA.

5) The Emotional module comprises Dataset Description and Model Description sections, with a multi-layered CNN utilized for image feature evaluation. This includes layers such as Input, Convolutional, Dense, and Output layers. The music classification module involves Preprocessing and feature Description of music dataset, while the recommendation module generates playlists using mapping techniques. Renuka S. Deshmukh et al. proposed facial emotion detection using machine learning algorithms for emotion classification, with plans to further develop an emotional expression API to output classified emotions.

6) Shlok Gilda and colleagues investigate the application of convolutional neural networks (CNNs) for emotion recognition, leveraging CNNs' ability to mimic human brain function in visual analysis. However, due to the computational demands and intricacy of CNNs, optimizing network efficiency is crucial.

III.PROPOSED METHODOLOGY

A. Data Collection

The Facial Expression Recognition 2013 (FER 2013) dataset stands as a pivotal resource within the realm of facial expression analysis and emotion recognition.

It provides researchers with a comprehensive collection of grayscale facial images, each measuring 48x48 pixels, depicting individuals expressing various emotional states. With approximately 35,887 images categorized into seven distinct emotion classes including anger, disgust, fear, happiness, sadness, surprise, and neutrality, this dataset serves as a foundational benchmark for training and evaluating algorithms in automatic facial expression recognition tasks.

The compilation of the FER 2013 dataset was spearheaded by the Facial Expression Recognition and Analysis (FERA) lab at the University of California, San Diego.

The FER2013 dataset is divided into training and testing datasets comprising 24,176 and 6,043 grayscale images, respectively. Each image is labeled with one of five emotions: happiness, sadness, anger, surprise, or neutrality. The dataset features both posed and unposed headshots for improved efficiency.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

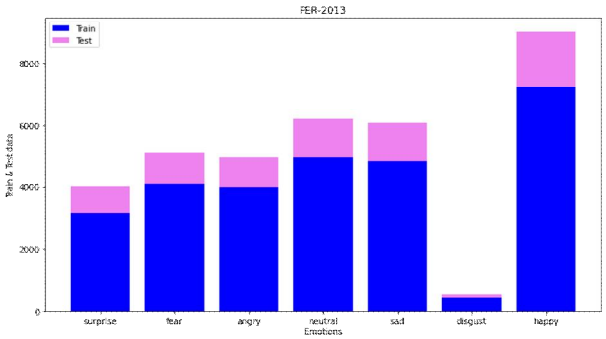


Fig. 1 Distribution of train/test dataset per class



Fig. 2 FER2013 Dataset Sample Images

B. Capturing Images

As the initial step in emotion recognition, capturing images using OpenCV entails initializing the webcam, continuously retrieving frames, and performing subsequent image processing and analysis. This foundational process serves as the basis for detecting facial expressions, a crucial component in determining the emotional state of an individual. Through leveraging OpenCV's capabilities, this step lays the groundwork for subsequent stages in the emotion recognition pipeline, facilitating accurate and real-time assessment of emotions from visual cues.

C. Image PreProcesssing

Pre-processing the captured images in emotion recognition involves converting them to grayscale to simplify analysis, reducing noise and isolating facial regions through face detection algorithms, facilitating accurate emotion assessment.

D. Feature Extraction

MediaPipe acts as a fundamental component for extracting and analyzing facial features. Through its versatile framework, it offers a robust set of tools and pre-trained models tailored for facial landmark detection and tracking. These models are adept at precisely identifying key facial landmarks such as the eyes, nose, and mouth, which serve as vital indicators of emotional expression. Utilizing MediaPipe's capabilities, our system seamlessly tracks the movement and changes in these facial landmarks in real-time video streams. It’s versatile set of tools and modules can be combined with heuristic methods to enhance the performance and functionality.

This dynamic tracking enables the continuous analysis of users' facial expressions, capturing subtle nuances and variations indicative of different emotional states. By leveraging MediaPipe's functionality, our project empowers the system to accurately interpret users' emotional cues from their facial expressions, thereby enabling personalized music recommendations aligned with their mood and emotional context.

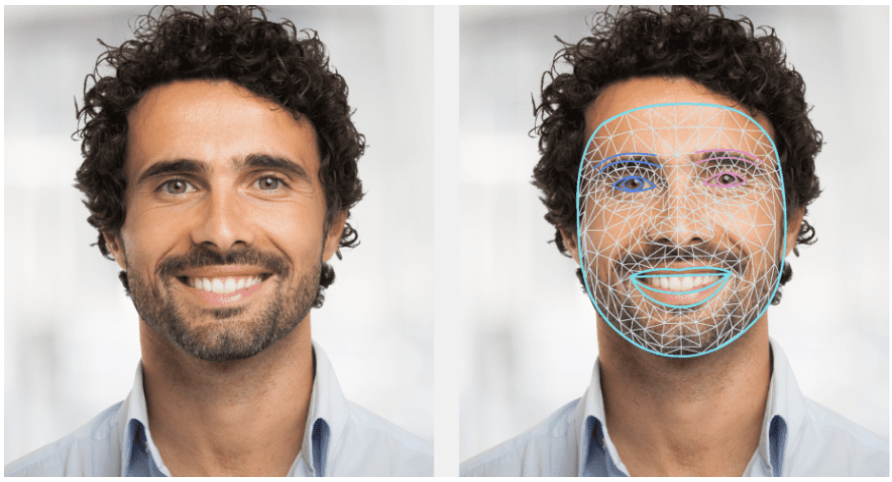


Fig. 3 Feature Extract ion Using Mediapipe

E. Emotion Detection Using CNN

Emotion detection using Convolutional Neural Networks (CNNs) is a process that begins with the preparation of the dataset. In this case, the FER2013 dataset is employed, which comprises grayscale images depicting facial expressions labeled with corresponding emotions. To ensure effective training, the dataset undergoes preprocessing steps aimed at standardizing the image size, normalizing pixel values, and introducing augmentation techniques to diversify the dataset. Following data preparation, a CNN architecture is designed and implemented. Typically, CNN architectures consist of multiple layers, including convolutional layers responsible for feature extraction and fully connected layers for classification.

A Convolutional Neural Network (CNN) architecture includes layers for input, convolution, activation, pooling, and fully connected layers, culminating in an output layer. The input layer receives raw data, followed by convolutional layers that extract features. Activation functions introduce non-linearity, while pooling layers down sample feature maps. Fully connected layers further extract high-level features, and the output layer produces the final predictions. CNNs excel at automatically learning and extracting complex patterns from data, making them well-suited for tasks such as image classification and emotion recognition.

During training, the model learns to discern intricate patterns within facial expressions associated with different emotions. This learning process is facilitated through optimization algorithms like stochastic gradient descent, which iteratively adjusts the model's parameters to minimize the difference between predicted and actual emotions. To optimize the model's performance, hyperparameters such as learning rate, batch size, and network architecture are fine-tuned. Once training is complete, the model's accuracy is evaluated on a separate test set to assess its ability to generalize to unseen data. Finally, with a trained model in hand, real-world applications can leverage it to automatically analyze facial expressions and infer underlying emotions, enabling advancements in fields such as affective computing and human-computer interaction.

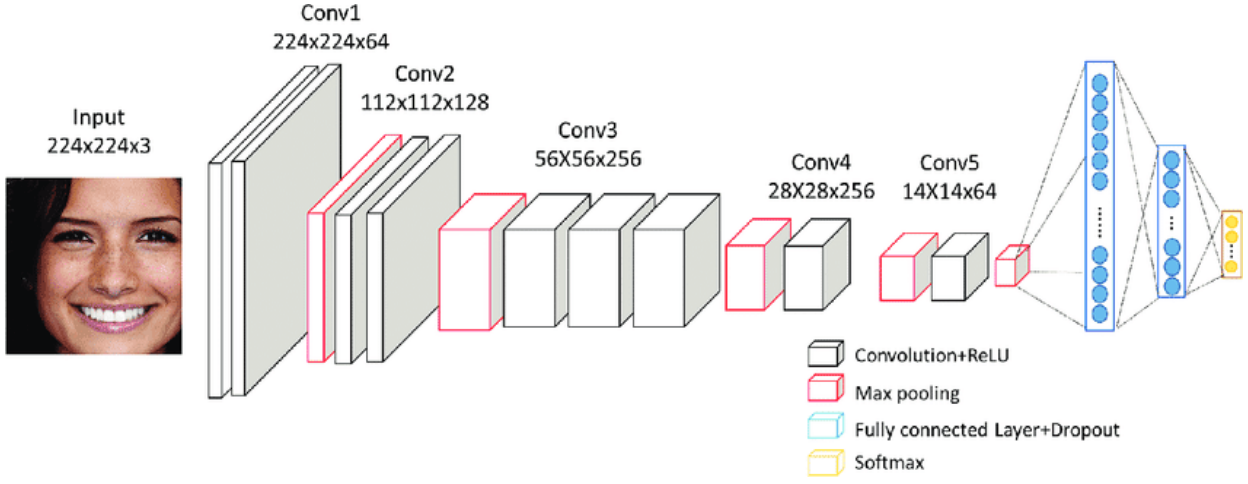


Fig. 4 CNN Architecture

F. Music Recommendation

Upon determining the user's emotional state, they are empowered to personalize their experience by selecting their preferred language from a list featuring Telugu, Hindi, English and music streaming platform from a curated list featuring Spotify, YouTube, and SoundCloud. Once their choices are made, the interface intuitively recommends songs aimed at enhancing the user's mood, creating a tailored listening experience. In accomplishing this task, we harness the capabilities of the Streamlit web framework, an open-source Python library that enables rapid development of interactive web applications for data science and machine learning projects. By integrating Streamlit into our project, we empower users to effortlessly navigate and personalize their experience. This allows for dynamic selection of language and music streaming services, enhancing user engagement and satisfaction.

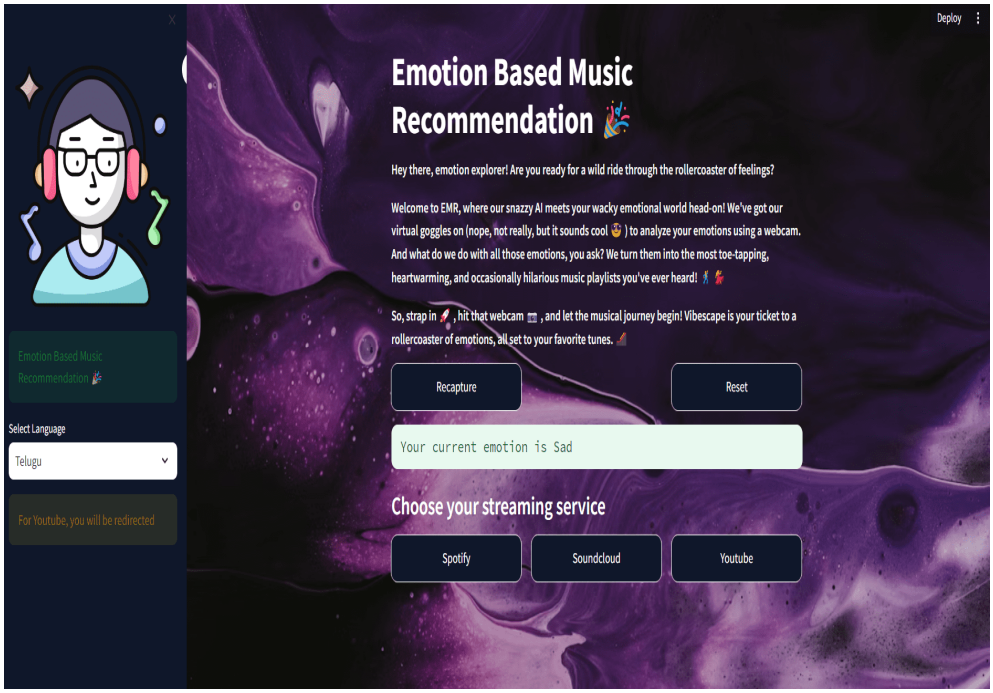


Fig. 5 Interface After Capturing User Emotion

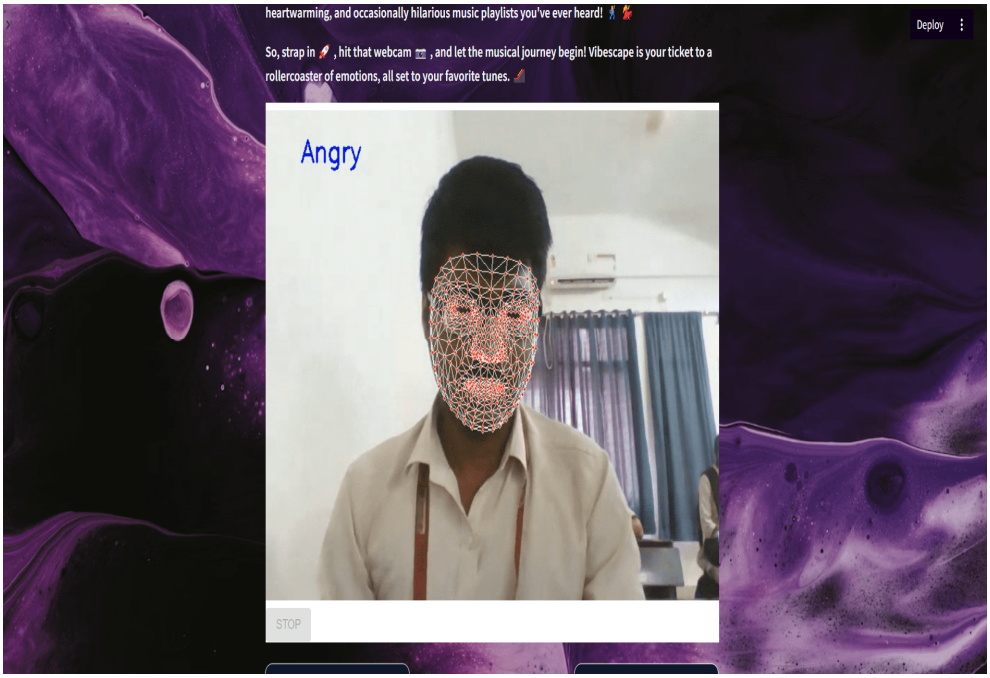


Fig.7 Detection of the "Angry" Emotion

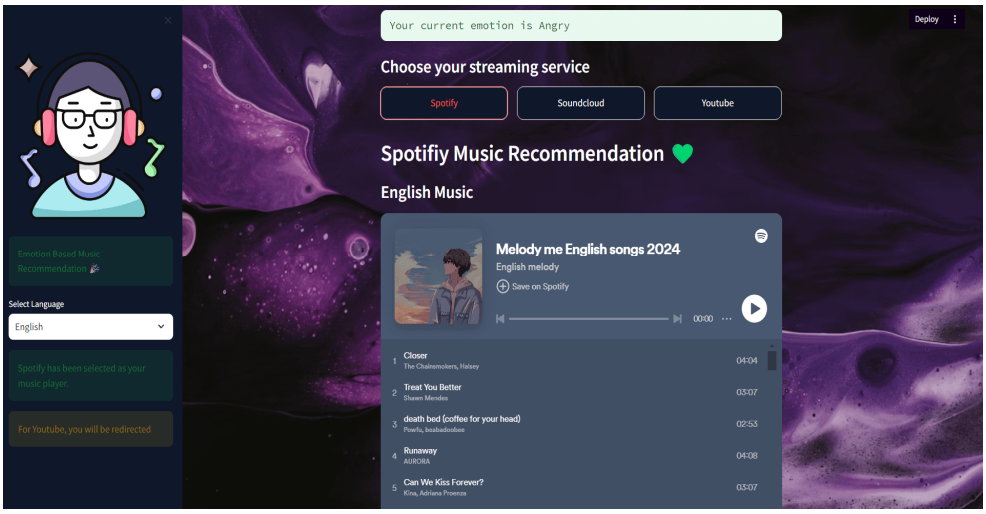


Fig. 8 Spotify Playlist Recommendation for Anger

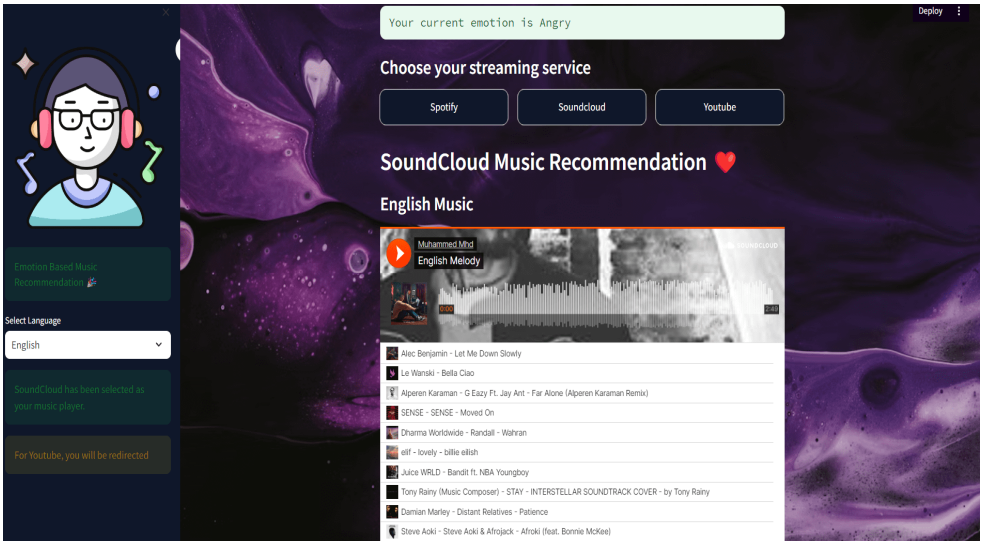


Fig. 9 SoundCloud Playlist Recommendation for Anger

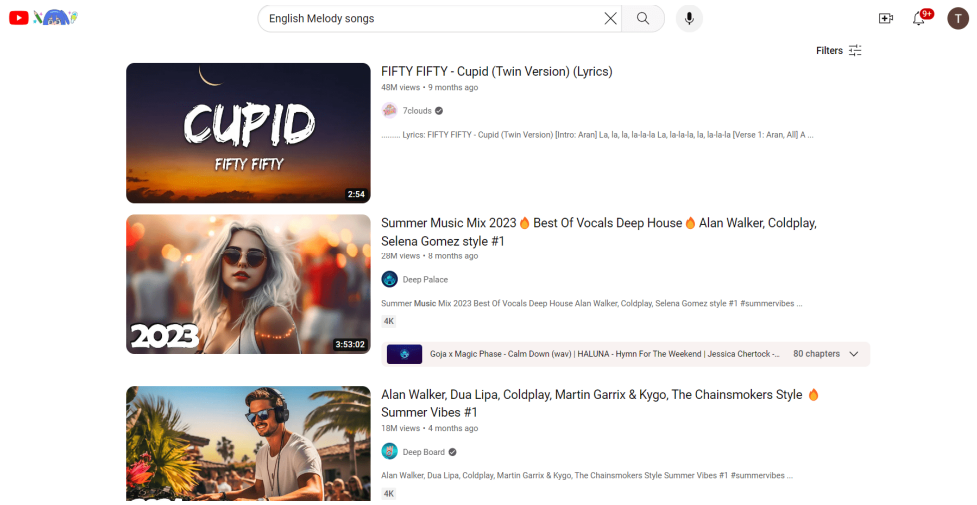


Fig. 10 Youtube Playlist Recommendation for Anger

IV. RESULTS

In essence, this project revolves around detecting the user's emotions by analyzing facial expressions captured in real-time through a camera. Leveraging advanced technology, the system interprets the user's mood based on their facial cues. Once the emotion is identified, the system generates customized music playlists tailored to match the user's emotional state. These playlists span across a variety of music streaming platforms, offering users a diverse selection of songs that resonate with their feelings. This seamless integration of emotion detection and music recommendation enhances the overall user experience, providing a personalized and immersive journey through music.

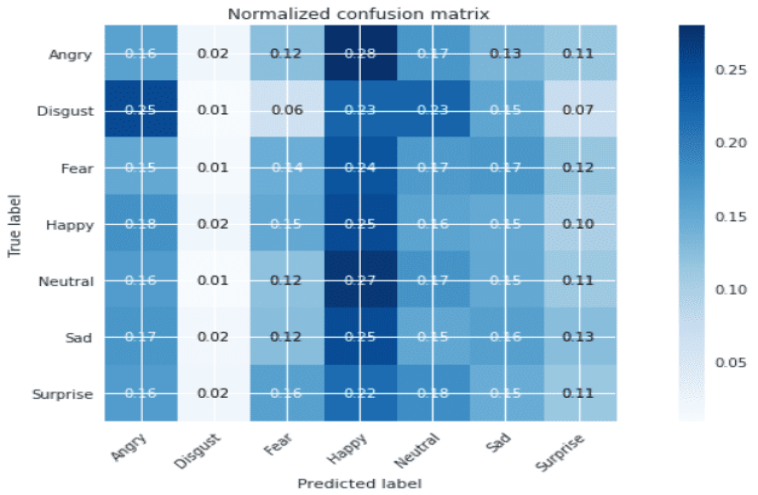


Fig. 15 Normalized Confusion Matrix

	Metric Name	Train	Validate
0	Loss	0.336	1.448
1	Accuracy	0.975	0.900
2	Precision	0.928	0.661
3	Recall	0.898	0.622
4	Auc	0.994	0.898
5	F1 Score	0.913	0.640

Fig. 16 Model Evaluation Scores

V. CONCLUSIONS

In conclusion, our project underscores the effectiveness of combining Mediapipe and Convolutional Neural Network (CNN) techniques for music recommendation based on facial expressions. While Mediapipe offers simplicity and efficiency in facial feature detection, the CNN model provides advanced emotion recognition capabilities. Through this amalgamation, we have developed a robust music recommendation system adept at accurately discerning users' emotions from facial expressions and tailoring personalized playlists accordingly. In comparison to existing systems, our project introduces notable improvements. One such enhancement is the dynamic adaptation of music playlists within streaming applications, ensuring users are continually exposed to a diverse array of tracks. Additionally, our user interface prioritizes accessibility, allowing users to select from multiple languages (including Telugu, Hindi, and English) and choose their preferred music recommendation platform (SoundCloud, Spotify, or YouTube). Noteworthy is our model's impressive accuracy rate, reaching up to 92%, affirming its efficacy in accurately identifying user emotions. Our study emphasizes the significance of employing diverse methodologies to augment the accuracy and efficacy of emotion-based music recommendation systems. Moving forward, further research avenues could focus on refining and optimizing these algorithms, as well as exploring additional modalities like voice tone and physiological signals to enhance emotion detection and music recommendation precision. Ultimately, our project highlights the transformative potential of facial expression analysis in reshaping music interactions, delivering customized listening experiences aligned with users' emotional states.

VI. LIMITATIONS AND FUTURE SCOPE

A. *Limitations*

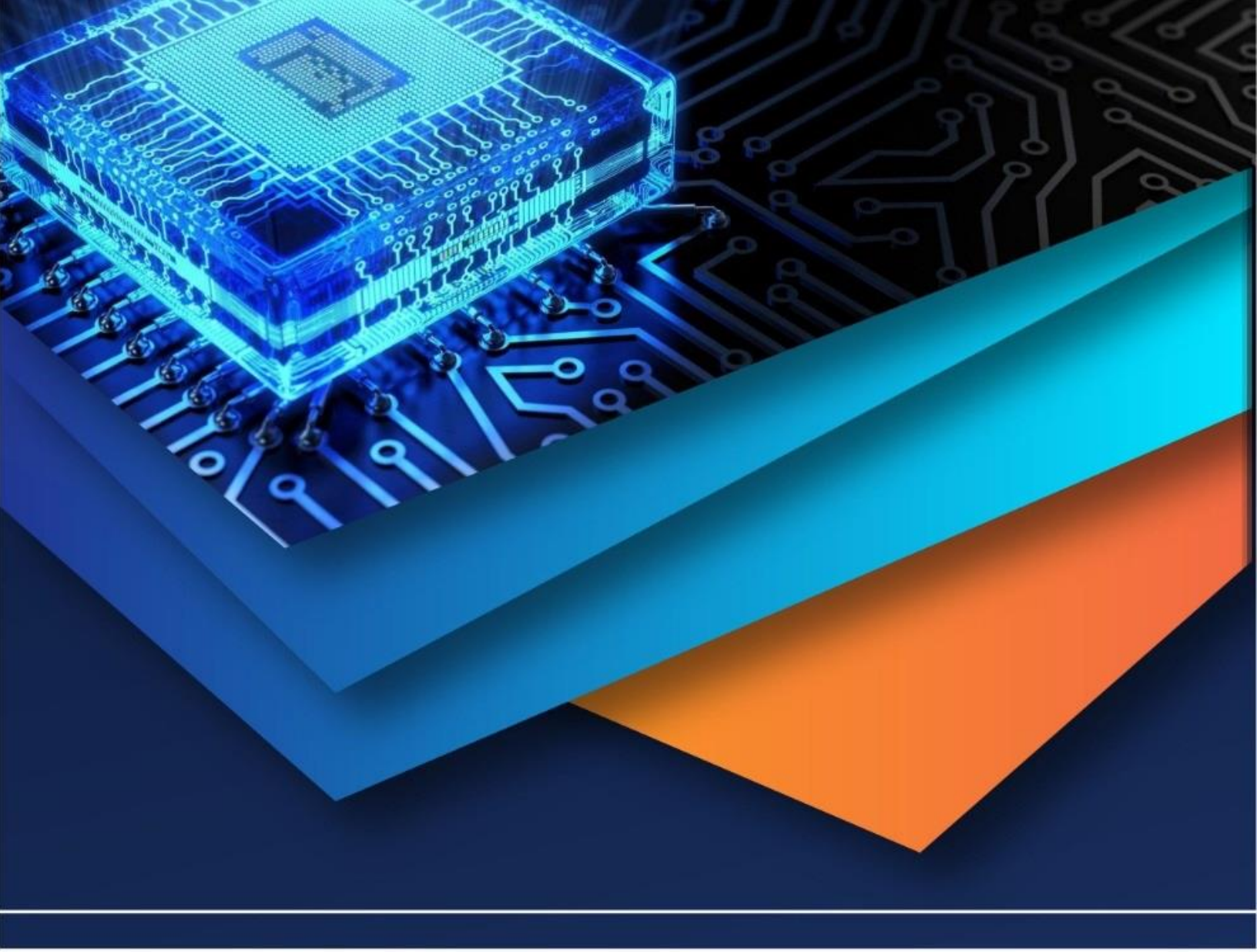
1) In CNNs may struggle to capture the nuances of complex emotions or subtle facial expressions, leading to difficulties in accurately recognizing certain emotions.
2) Mediapipe may lack extensive customization options for fine-tuning the facial expression analysis algorithms .
3) The FER dataset may lack diversity in terms of age, gender, ethnicity, and cultural backgrounds, potentially resulting in biased or limited models that do not generalize well to diverse populations.
4) CNNs often require significant computational resources, making them computationally expensive to train and deploy, especially for large datasets.
5) Mediapipe for facial detection is its sensitivity to variations in lighting conditions, poses, and occlusion.

B. *Future Scope*

1) In CNNs Enhancing the accuracy and robustness of emotion recognition algorithms, possibly through the integration of multimodal data sources such as facial expressions, voice tone, and physiological signals, can lead to more precise emotion detection.
2) Exploring the integration of multiple modalities such as facial expressions, body language, and user context (e.g., location, activity) to capture a more comprehensive understanding of user emotions and preferences for more nuanced music recommendations.
3) Emotion-based music recommendation systems hold potential in the healthcare and wellness domain for applications such as music therapy, mood regulation, and stress management.
4) Integrating real-time feedback mechanisms to continuously adjust music recommendations based on user feedback and emotional responses during music playback can provide more adaptive and personalized experiences.
5) Integrating this model into any music streaming platform could bring real time streaming service.

REFERENCES

[1] Gokul Krishnan K, Parthasarathy M, Sasidhar D and Venitha E – Emotion detection and music recommendation system using machine learning‖, International Journal of Pune and Applied Mathematics, vol. 119, pp. 1487-1498, 2018.
[2] Arto Lehtiniemi and Jukka Holm – Using animated mood pictures in music recommendation‖, 16th International Conference on Information Visualisation, 2011.
[3] 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 Solutions 2017. https://doi.org/10.1109/CSITSS.2017.8447743
[4] Kyogu Lee and Minsu Cho "Mood Classification from Musical Audio Using User Group-Deendent Model" Published by IEEE 10th International Conference on Machine Learning and Application on 2011 (978-0- 7695-4607-0/11)
[5] Renuka S. Deshmukh, Vandana Jagtap and Shilpa Paygude "Facial Emotion Recognition system Through Machine Learning" Published by ICICCS in 2017.
[6] Shlok Gilda, Hussain Zafar, Chintan Soni and Kshitija Waghrdekar "Smart Music Player Integration Facial Emotion Recognition And Music Mood Recommendation" Published by IEEE WISPNET on 2017
[7] Aurobind V. Iyer, Viral Pasad, Karan Prajapati "Emotion Based Mood Enhancing Music Recommendation" Published by IEEE International conference on RTEICT, May 2017,India. Page No: 1573-1577.
[8] Swathi Swaminathan and E. Glenn Schellenberg "Current Emotion Research In Music Psychology" Published Emotion Review Vol. 7 ,No.2 on April 2015 (189-197), ISSN 1754-0739.
[9] Prof. Pankaj Kunekar,Pranjul Agrahari, Ayush Singh Tanwar, Biju Das, "Musical Therapy using Facial Expressions"International Research Journal of Engineering and Technology. 01 Jan 2020
[10] Shambhavi Milgir, Ankita Mahadik, Vaishali Kavathekar and Prof. Vijaya Bharathi Jagan. "Mood based Music Recommendation System" International Journal of Engineering Research & Technology, 2021
[11] Mr.Akshay Choudhary,Prof.Sumeet Pate, Miss.Poonam Harane, Miss.Shreya Zunjarrao"Mood Based Music Player Using Real Time Facial Expression Extraction" International Journal for Research in Engineering Application & Management (IJREAM) 2019
[12] P. Singhal, P. Singh and A. Vidyarthi (2020) Interpretation and localization of Thorax diseases using DCNN in Chest X-Ray. Journal of Informatics Electrical and Electronics Engineering,1(1), 1, 1-7
[13] Pranjal Agrahari1, Ayush Singh TanwarMusical Therapy Using Facial Expressions, IRJET 2020
[14] M. Chen, Y. Zhang, Y. Li, M. M. Hassan, and A. Alamri, ``AIWAC: Affective interaction through wearable computing and cloud technology," IEEE Wireless Commun., vol. 22, no. 1, pp. 20_27, Feb. 2015.
[15] M. S. Hossain, G. Muhammad, B. Song, M. M. Hassan, A. Alelaiwi, and A. Alamri, ``Audio_visual emotion-aware cloud gaming framework," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 12, pp. 2105_2118, Dec. 2015.
[16] P. M. Ashok Kumar, Jeevan Babu Maddala & K. Martin Sagayam (2023) Enhanced Facial Emotion Recognition by Optimal Descriptor Selection with Neural Network, IETE Journal of Research, 69:5, 2595-2614, DOI: 10.1080/03772063.2021.1902868



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)