JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Music Recommendation System based on Facial Expression using Deep Learning

[1]S Rao Chintalapudi, [2]Gajji Mahesh, [3]Badiga Harshini, [4]Charukonda Bhairava Harini

[1,2,3,4]Department of CSE(AI&ML),
[1,2,3,4,]CMR Technical Campus, Hyderabad, India

Abstract : This paper presents an automatic music recommendation system using facial emotion detection with a Convolutional Neural Network (CNN). The model captures real-time facial expressions via a webcam, classifies emotions such as happy, sad, angry, surprised, neutral and recommends music using Deep Learning and Streamlit. Achieving an accuracy of around 85%, it enhances recommendation precision while reducing computational time. This system is ideal for music streaming platforms and therapeutic applications. The final output is displayed as an audio-visual experience via YouTube integration in a user-friendly web application.

IndexTerms **– Deep Learning, Convolutional Neural Networks, Recommendation Systems, Haar Cascades, Emotion Detection**

I. INTRODUCTION

Music recommendation system that integrates facial expression recognition and deep learning to create a personalized listening experience. By leveraging Convolutional Neural Networks (CNNs), the system captures real-time facial expressions via a webcam, classifies emotions such as happy, sad, angry, surprised, neutral and recommends suitable music. Unlike traditional recommendation systems that rely on historical data, our approach adapts dynamically to the user's emotional state, enhancing engagement and satisfaction. The model is trained on a diverse dataset using image processing, data augmentation, and transfer learning for improved accuracy and robustness. Once an emotion is identified, a hybrid recommendation engine—combining collaborative and content-based filtering—maps emotions to appropriate music genres. Designed for cross-platform compatibility, the system operates on desktops, laptops, tablets, and smartphones, ensuring broader accessibility. This paper not only enhances user experience in music streaming platforms but also has potential applications in mental health and therapeutic interventions. Achieving an accuracy of around 85%, the proposed system efficiently reduces computational time while providing emotionally adaptive music recommendations.

The structure of the remaining paper is as follows. Section II discuss about the existing approaches related to facial expression based music recommendation. Building a music recommendation system is discussed in section III. Section IV outlines the experimental setup. Section V elaborates Results and Discussion. Section VI concludes the paper with future directions.

II. RELATED WORK

Facial expression recognition and its applications in emotion-driven systems have been extensively explored in recent research. Priya et al. [1] introduced a Machine Learning-based Enhanced Support Vector Machine (SVM) technique for predicting stress, demonstrating the potential of computational models in recognizing emotional states. However, SVMs face limitations in handling complex, high-dimensional image data, which deep learning models can address more effectively.

Rahul Ravi et al. [2] implemented a Convolutional Neural Network (CNN) combined with Local Binary Patterns (LBP) for facial expression recognition, highlighting the effectiveness of CNNs in feature extraction. Their model showcased improved performance in detecting subtle facial changes, which is crucial for real-time applications. However, their approach primarily focused on static image datasets rather than real-time video input.

A foundational contribution to face detection was made by Viola and Jones [3], who developed a real-time face detection framework using Haar-like features and an AdaBoost classifier. This work laid the groundwork for modern face detection techniques, making it possible to efficiently locate faces in images and videos, an essential step in facial emotion recognition.

To enhance the accuracy of facial expression recognition in dynamic settings, Fan et al. [4] proposed a hybrid CNN-RNN and 3D Convolutional Network (C3D) approach for video-based emotion recognition. Their model successfully captured both spatial and temporal features, improving performance for real-time emotion detection. However, its computational complexity remains a challenge for lightweight, real-time applications [7][8][9][10].

Further improvements in facial expression recognition were presented by Happy and Routray [5], who developed an automatic facial expression recognition system that utilized salient facial patches to improve feature extraction and classification accuracy. Their work demonstrated the importance of selecting relevant facial regions to reduce computational overhead while maintaining high accuracy.

Building upon these advancements, our proposed system integrates real-time facial expression analysis with a deep learning-based music recommendation engine. Unlike existing studies that focus solely on emotion detection, our approach bridges the gap between facial expression recognition and music personalization, ensuring a dynamic and responsive user experience. By leveraging CNNs, hybrid filtering techniques, and real-time video input, our system improves accuracy, reduces computational time, and enhances user engagement in music streaming and therapeutic applications.

III. Building Music Recommendation System

The music recommendation system based on facial expression is developed to offer personalized music suggestions by analyzing the user's emotions in real time. The process begins with capturing the user's facial expression through a webcam. Advanced face detection techniques such as Haar cascades, Multi Task Cascaded Convolutional Networks (MTCNN) are used to isolate the facial region. This is followed by emotion classification using a Convolutional Neural Network (CNN), which analyzes the facial features and determines the user's emotional state. Based on the detected emotion, the system recommends suitable music through content-based and collaborative filtering methods, creating a personalized listening experience.

The system includes several notable features such as real-time emotion detection, a high level of accuracy using deep learning models, and continuous learning through user feedback. It aims to provide an emotionally tailored experience for each user. During testing, the system showed good accuracy in identifying emotions and suggesting appropriate music. However, external factors like poor lighting and facial obstructions had a negative impact on its performance. Overall, the system successfully links facial emotion recognition with music recommendation, enhancing the way users interact with music. It has promising applications in areas like mental health therapy, entertainment, smart home systems, and virtual reality gaming. With ongoing improvements, it can become even more adaptive and responsive to user emotions, making music interaction more natural and engaging.

IV. Experimental Setup

The experimental setup for the Music Recommendation System was carried out in a local environment that has intel i5 processor with 2.5GHz, 8GB RAM and 256GB SSD with Python 3 installed. And also the key libraries including OpenCV, TensorFlow/Keras, Librosa are installed. OpenCV was employed for face detection and image preprocessing, while TensorFlow/Keras was used to build and train the Convolutional Neural Network (CNN) for emotion classification. Librosa facilitated audio analysis to enhance music recommendation based on song features like tempo and rhythm. The system incorporated various techniques such as grayscale image conversion, facial landmark detection, and real-time face tracking using Haar cascades or Dlib's HOG-based detectors. Emotion classification was performed using a CNN trained on datasets like FER2013[6], and emotions were mapped to mood-based playlists for relevant music suggestions.

In terms of data preprocessing, the system captured live facial images using a webcam and converted them into grayscale for efficient processing. Facial key points such as the eyes, eyebrows, and lips were extracted to analyze expressions accurately. Based on these features, the trained model predicted the user's emotion in real time. The detected emotion was then used to filter and recommend songs that matched the user's mood, making the experience more personalized and responsive. This real-time integration allowed the system to continuously analyze expressions and update music recommendations accordingly.

V. Results and Discussion

The emotion-based music recommendation system leverages deep learning and computer vision to personalize music choices based on users' facial expressions. By analyzing emotions in real-time, the system suggests relevant songs from YouTube, enhancing the listening experience.

Fig.1 represents the input image of the user. Based on the analysis, it has been identified as a sad emotion represented in Fig.2. The emotion-based music recommendation system utilizes a webcam to capture the user's facial expressions and employs a Convolutional Neural Network (CNN) to analyze and classify their emotions. Once the emotion is detected, it is displayed on the interface refer Fig.2., allowing users to see their current mood classification. Additionally, users can personalize their experience by selecting a preferred language such as telugu and singer such as sid sriram, ensuring that the recommended music aligns with their tastes as represented in Fig.3. The system is developed using Streamlit, which provides an intuitive and seamless web-based interaction for an enhanced user experience.

After detecting the user's emotion, the system dynamically fetches relevant songs from YouTube based on the identified mood. For instance, if a sad emotion is detected, the system retrieves a curated selection of sad songs, ensuring that the music aligns with the user's emotional state. This process integrates advanced computer vision techniques, deep learning models, and web scraping methods to generate precise and context-aware recommendations. By leveraging real-time data from YouTube, the system ensures that users always receive the latest and most relevant song choices.

The performance of the system is evaluated using key metrics, including emotion detection accuracy, click-through rate (CTR), and emotion-to-music match accuracy. The results indicate that the CNN model achieves approximately 80% accuracy in detecting emotions, while the emotion-to-music match accuracy is close to 90%, demonstrating the system's effectiveness in making relevant recommendations as shown in Fig.4. The CTR further validates user engagement with the suggested tracks, confirming that the system successfully connects emotions with suitable music. These results highlight the potential of deep learning and computer vision in creating personalized and responsive music recommendation systems.

Fig. 1: Initial Facial Expression of the User

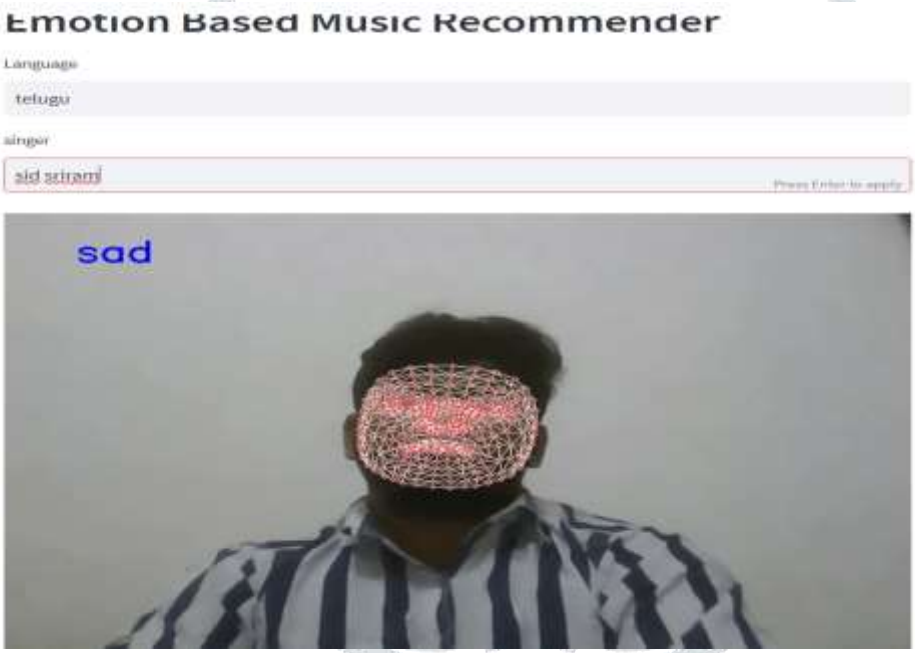

Fig. 2: Detecting Emotion of a user

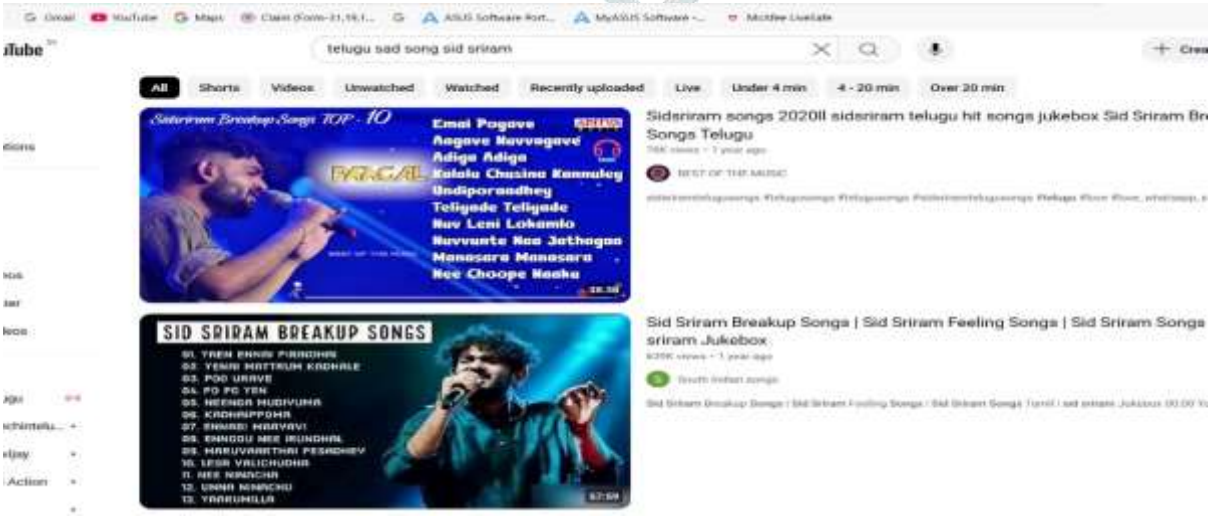

Fig 3: Song Recommendations in Youtube

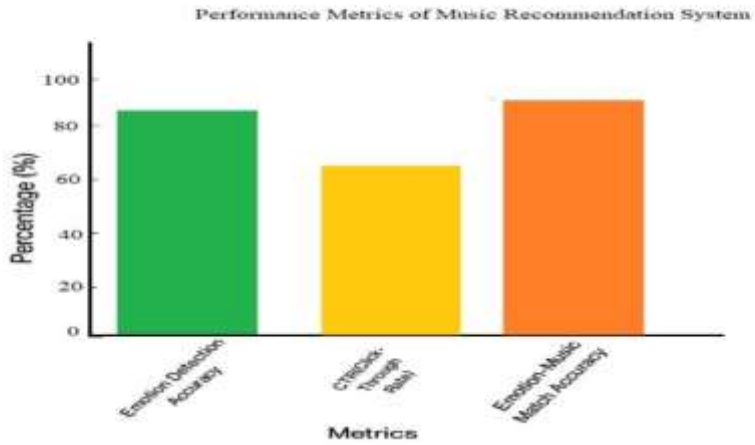

Fig. 4: Performance Metrics of the Music Recommendation System

VI. Conclusion and Future Scope

In this paper, the model predicts user emotions using a Convolutional Neural Network (CNN) and recommends songs accordingly. A web application, built with Streamlit, captures the user's facial expressions via image or video and processes them using deep learning. Once the emotion is detected, a corresponding playlist is generated by fetching tracks from YouTube. The system integrates computer vision and machine learning to enhance music recommendations. Deep Neural Networks (DNNs) are utilized for feature abstraction, improving accuracy in emotion recognition. CNNs, known for their effectiveness in image classification, enable precise facial expression detection. The client-side requests and plays music via YouTube based on the predicted emotion. The results indicate that the CNN model achieves approximately 80% accuracy in detecting emotions, while the emotion-to-music match accuracy is close to 90%, demonstrating the system's effectiveness in making relevant recommendations. In future, incorporating transformer-based models, could further improve its efficiency and accuracy in real-world conditions.

References

[1] T. Mohana Priya, Dr. M. Punithavalli & Dr. R. Rajesh Kanna, Machine Learning Algorithm for Development of Enhanced Support Vector Machine Technique to Predict Stress, Global Journal of Computer Science and Technology: C Software & Data Engineering, Volume 20, Issue 2, No. 2020, pp 12-20

[2] Rahul ravi, S.V Yadhukrishna, Rajalakshmi, Prithviraj, "A Face Expression Recognition Using CNN &LBP",2020 IEEE.

[3] Paul Viola and Michael J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, May 2004.

[4] Fan, Yin, et al. "Video-based emotion recognition using CNN-RNN and C3D hybrid networks." Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM, 2016.

[5] S.L.Happy and A. Routray,"Automatic facial expression recognition using features of salient facial patches,"in IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 1-12, 1 Jan.-March 2015.

[6] FER-2013 dataset available at https://www.kaggle.com/datasets/msambare/fer2013

[7] T. Gorasiya, A. Gore, D. Ingale and M. Trivedi, "Music Recommendation based on Facial Expression using Deep Learning," 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1159-1165, doi: 10.1109/ICCES54183.2022.9835929.

[8] Singh, A., Sharma, R., Pandey, M.S., Asthana, S., Gitanjali, Vishwakarma, A. (2024). Facial Expression Based Music Recommendation System Using Deep Learning. In: Data Science and Network Engineering. ICDSNE 2023. Lecture Notes in Networks and Systems, vol 791. Springer, Singapore. https://doi.org/10.1007/978-981-99-6755-1_3.

[9] P. Ashwini, P. Dammalapati, N. Ramineni and T. Adilakshmi, "Facial Expression based Music Recommendation System using Deep Convolutional Neural Network," in 2024 International Conference on Expert Clouds and Applications (ICOECA), Bengaluru, India, 2024, pp. 992-999, doi: 10.1109/ICOECA62351.2024.00173.

[10] S Metilda Florence and M Uma 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* 912 062007. DOI 10.1088/1757-899X/912/6/062007