RESEARCH ARTICLE

SAMRIDDHI Volume 16, Issue 4, 2024 Print ISSN: 2229-7111 Online ISSN: 2454-5767

Emotion-Driven Music Recommendation System

Vilas Gaikwad,[1] Rushikesh Garde[2], Nishigandha Dhage[2*], Sanskruti Patil[2], Swayam Bharsakale[2]

[1]HOD Information Technology, Trinity College of Engineering and Research, Pune, India.

[2]Information Technology BE Student, Trinity College of Engineering And Research, Pune, India.

Abstract

The "Enhanced Emotion-Driven Music Recommendation System" revolutionizes personalized music recommendations by leveraging real-time facial expression analysis. The project adopts a novel learning strategy, drawing upon established research in Convolutional Neural Networks (CNNs) for facial expression recognition. By employing a "divide-and-conquer" approach, integrating attention mechanisms, data augmentation techniques, and leveraging Haar cascade for facial detection, the system achieves remarkable precision in emotion detection. Seamlessly integrating background sound corresponding to detected emotions, the system offers a dynamic and immersive user experience.

Moving forward, the project aims to achieve even greater advancements by incorporating IoT devices to capture physiological indicators such as heart rate, thereby enriching adaptability and personalization. The "Enhanced Emotion-Driven Music Recommendation System" represents a noteworthy milestone in personalized music recommendation systems. It surpasses conventional methods by dynamically adjusting to users' emotional states in real-time.

Keywords: Music Recommendation, Emotion Analysis, User Experience, HAAR Cascade

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2024); DOI: 10.18090/samriddhi.v16i04.02

Introduction

In today's digitally driven society, where music permeates every aspect of our lives, the quest for the perfect song to match our mood remains a common challenge. While conventional music recommendation systems excel in organizing and suggesting tracks based on factors like genre or artist, they often overlook the intricate nuances of human emotion.

Our pioneering initiative, the "Emotion-Driven Music Recommendation System (EDMRS)." Recognizing the limitations of existing platforms, we set out to redefine the relationship between listeners and music by focusing on the emotional connection.

Our project aims to bridge this gap by leveraging advanced facial expression recognition technology to tailor music recommendations to the user's emotional state.

At the core of our endeavour lies a synthesis of cutting-edge technology and a deep understanding of human emotion. By analysing facial expressions, we seek to decipher the emotional landscape of the listener and curate a personalized musical experience that resonates with their feelings. This departure from conventional genre-based categorization heralds a new era in music recommendation, where emotional intelligence takes precedence over traditional metrics.

Our project not only represents a significant advancement in music technology but also underscores the potential of music to foster empathy and connection. By harnessing the power of technology to understand and respond to users' emotional needs, we envision a future where music becomes a profound tool for emotional expression and connection, enriching lives and transcending boundaries.

Corresponding Author: Nishigandha Dhage, Information Technology BE Student, Trinity College of Engineering And Research, Pune, India., e-mail: dhagenishigandha@gmail.com

How to cite this article: Gaikwad, V., Garde, R., Dhage, N., Patil, S., Bharsakale, S. (2024). Emotion-Driven Music Recommendation System. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology,* 16(4), 136-140.

Source of support: Nil

Conflict of interest: None

Problem Statement

In today's digital landscape, the vast array of music available often leaves us struggling to find songs that resonate with our current emotions. Conventional music recommendation systems lack the ability to consider our feelings, resulting in a less fulfilling experience when discovering new music. To address this issue, we propose the development of a specialized system that recommends songs based on the user's emotional state. Our objective is to enhance the enjoyment of music discovery and facilitate emotional connections with the songs we listen to.

Current research, as evidenced by studies like "CNN Learning Strategy for Recognizing Facial Expressions" and "Automated Facial Expression Recognition Framework Using Deep Learning," underscores the crucial role of facial

expression analysis in understanding human emotions. However, these studies face limitations, such as the need for further optimization in CNN architectures and the inability to address real-time analysis and dynamic adaptation to users' changing emotional states. Our project aims to overcome these challenges by creating an improved emotion-driven music recommendation system that utilizes real-time facial expression detection through webcams or laptop cameras.

The primary focus of our project is to design and implement an accurate and efficient facial expression analysis system that seamlessly integrates with the music recommendation platform. This system must not only identify users' emotional states but also continuously adapt to evolving emotions, thereby delivering more personalized and captivating song recommendations.

Literature Review

In the paper the authors introduce an algorithm called FD-CNN for facial expression recognition (FER) and evaluate its performance extensively. They use the Cohn-Kanade (CK+) dataset, containing images of eight different facial expressions, and the methodology involves preprocessing, feature extraction, and classification. The study demonstrates that FD-CNN achieves high accuracy in FER, validated through 10-fold cross-validation, confusion matrices, sensitivity, specificity, precision, recall, F1 score, and ROC curves. Results show that FD-CNN outperforms state-of-the-art FER frameworks, making it a promising tool for automated facial expression recognition across various applications (Fig. 1).[1]

The Proposed system by Dong-Hwan Lee and Jang-Hee Yoo, published in IEEE Access on July 10, 2023, presents a novel approach for improving the accuracy of facial expression recognition (FER) using Convolutional Neural Networks (CNNs). The authors address the challenge of variations in facial expressions and other factors unrelated to expressions, such as lighting and head pose. They propose a divide-and-conquer learning technique, which involves preprocessing for face detection and normalization, optimizing a ResNet-18 CNN model, aggregating comparable facial expressions, and classifying these grouped expressions. The method is evaluated on various datasets, including Tufts, RWTH, RAF, and FER2013. The results demonstrate improved accuracy compared to previous methods, particularly for similar facial expressions. The proposed strategy has the potential to enhance FER in diverse real-world scenarios.[2]

In this work the authors propose a A Facial Expression Recognition framework employing a Hierarchical Features Three-Channel Convolutional Neural Network (HFT-CNN). The model aims to address the limitations of traditional convolutional neural networks for facial expression recognition by focusing on three key regions: the entire face, eyes and eyebrows, and the mouth. They control the mean and variance of convolutional kernels to improve feature extraction, and employ multi-channel fusion technology to combine the features from these regions. Experiments on JAFFE and CK+ datasets show that HFT-CNN achieves higher recognition rates compared to traditional approaches. This innovative model offers an effective and efficient solution for improving facial expression recognition.[3]

This paper discusses the importance of facial expressions in human communication and proposes a deep learning-based method called ZFER for facial emotion recognition with zoning. The study focuses on enhancing the performance of facial expression detection using deep learning by localizing face landmarks through zoning and extracting features like the forehead. The proposed ZFER model undergoes evaluation using benchmark datasets FER 2013 and CK+, demonstrating substantial enhancements in accuracy, precision, recall, and F1 score when juxtaposed with other baseline models such as AlexNet, ResNet, and CNN. The research demonstrates the effectiveness of zoning in facial emotion recognition and suggests its application in real-world scenarios, including IoT-based face recognition systems.[4]

In the paper titled "A survey on sentiment analysis methods, applications, and challenges" by Wankhade, M., Rao, A. C. S., & Kulkarni, C., published in *Artificial Intelligence Review* (2022), the authors present an extensive overview of sentiment analysis. They discuss the increasing significance of sentiment analysis, including its tasks such as subjectivity classification, sentiment classification, opinion spam detection, implicit language detection, and aspect extraction. The paper outlines various methodologies employed in sentiment analysis, including lexicon-based, machine learning, and hybrid approaches, emphasizing the advantages and disadvantages of each. This comprehensive survey is a valuable resource for scholars and beginners in the field, providing insights into the diverse techniques used for analyzing sentiment in text data.[5-10]

Design and Implementation

Traditional music recommendation systems often fall short in capturing the complex nuances of human emotions, relying primarily on factors like genre or artist popularity. This limitation results in less accurate and personalized song suggestions, leading to a less fulfilling user experience. Additionally, while Convolutional Neural Networks have demonstrated potential in the realm of facial expression recognition, they may struggle with real-time analysis and dynamic adaptation to users' changing emotional states. However, integrating Haar cascade for facial detection can overcome these drawbacks by providing a more efficient and accurate method for real-time facial expression analysis. Haar cascade is a machine learning-based approach that detects objects in images by analyzing features at different scales. By leveraging Haar cascade, the enhanced emotion-driven music recommendation system can achieve improved accuracy and responsiveness in detecting users' emotions.

Haar cascade operates through a series of steps to detect objects in images, making it suitable for real-time

applications like facial expression analysis. The process begins with the creation of a Haar cascade classifier, which is trained on a large dataset of positive and negative images. During training, the classifier learns to identify patterns or features that distinguish between positive and negative examples. These features are represented as rectangular regions of varying sizes and positions within the image. Once the classifier is trained, it can be applied to new images for object detection. This involves sliding a window of fixed size over the image and computing a feature vector for each window. The classifier then evaluates whether each window contains the object of interest based on the learned features. If a window is classified as positive, it is considered a detection, and the corresponding region of the image is marked as containing the object.

The process is repeated at multiple scales and positions within the image to detect objects of different sizes and orientations. Finally, non-maximum suppression is applied to remove overlapping detections and refine the final set of detected objects. This step ensures that only the most relevant detections are retained, improving the accuracy and reliability of the system.

The architecture of the enhanced emotion-driven music recommendation system is crafted to seamlessly fuse real-time facial expression analysis with personalized song suggestions. The user begins by accessing the system through a registration or login process, after which they are directed to the home page. From the home page, the user is prompted to turn on their camera to enable facial expression analysis. Once the camera is activated, the user can click the play button to begin detecting their mood and playing the corresponding song in the background.

The software flow diagram outlines the process for detecting the user's emotion, including capturing facial expressions from the camera feed, preprocessing the images, and passing them through the emotion detection model. The model outputs the detected emotion, which is then used to select and play the appropriate song from the recommendation engine (Figs 1 and 2).

Output

Users will start by signing up on the registration page, where they'll create an account by providing basic details like username and password. Once registered (Fig. 3), they'll use the login page (Fig. 4) to access their account by entering

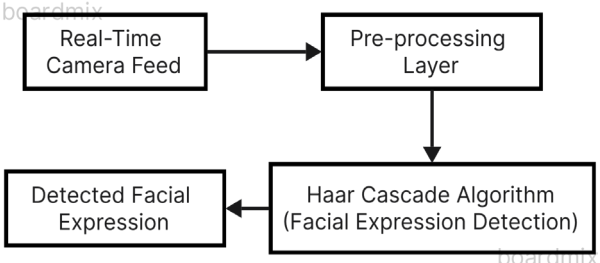

Fig. 1: Facial Expression Detection Process

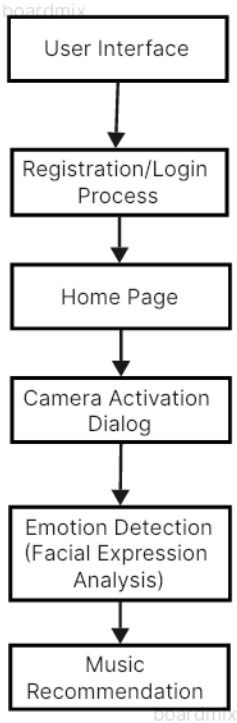

Fig. 2: System architecture

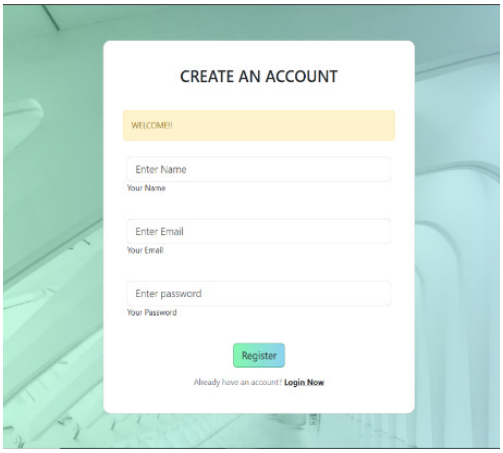

Fig. 3: Registration page

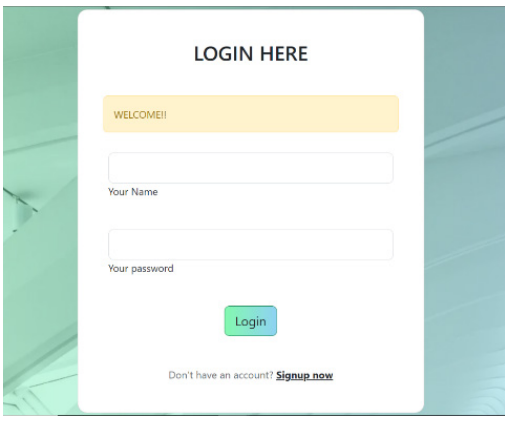

Fig. 4: Login Page

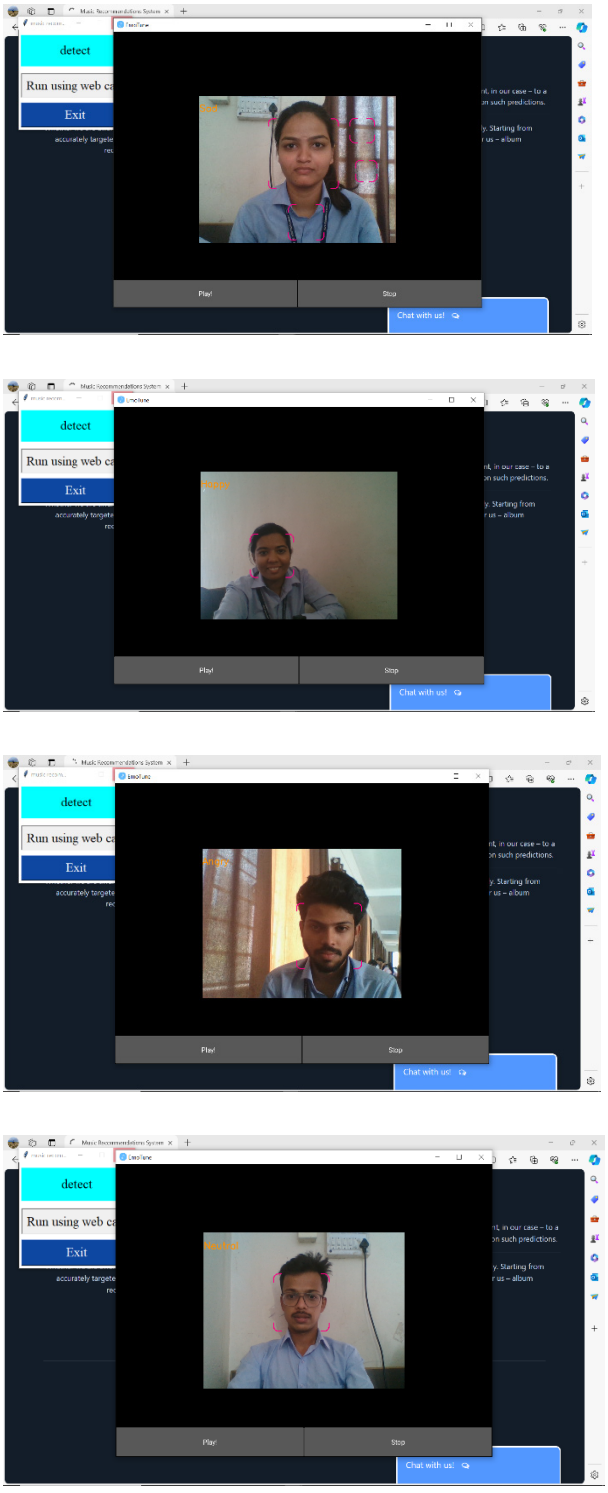

Fig. 5: Results

their username and password. After logging in, they'll be able to explore personalized music recommendations based on their emotions.These pages are designed to be easy to use, ensuring a smooth experience for users interacting with the EDMRS platform.

The attached screenshots showcase the results of facial expression recognition within the Enhanced Emotion-Driven Music Recommendation System (EDMRS), depicting four distinct emotional states: happy, sad, neutral, and angry. Each screenshot provides a visual representation of the system's ability to accurately detect and classify facial expressions in real-time.These screenshots demonstrate the system's effectiveness in capturing a diverse range of emotions, laying the foundation for personalized music recommendations aligned with users' emotional states (Fig. 5).

Future Scope

The Enhanced Emotion-Driven Music Recommendation System (EDMRS) holds significant potential for further advancement in personalized music recommendation. Looking ahead, several avenues can be explored to enhance the system's capabilities.

One area of focus is the integration of attention mechanisms to refine the granularity of emotion analysis. By capturing even subtler emotional cues, the system can provide more accurate music recommendations tailored to users' emotional states.

Furthermore, leveraging transfer learning from existing pre-trained models can expedite the system's learning process and enhance its accuracy. By incorporating knowledge from related tasks or datasets, EDMRS can adapt more quickly to new user preferences and musical contexts.

Moreover, integrating IoT devices such as smartwatches to incorporate physiological cues like heart rate presents an intriguing opportunity to enrich the emotional context for even more precise song recommendations.

Conclusion

The "Enhanced Emotion-Driven Music Recommendation System" (EDMRS) represents a paradigm shift in music discovery, offering users personalized song recommendations tailored to their emotional states. By leveraging advanced facial expression analysis and sentiment analysis, EDMRS transcends the limitations of traditional recommendation systems, providing a deeply immersive and emotionally resonant music journey. This project symbolizes the fusion of technology and art, with the potential to redefine how individuals engage with and derive joy from music. As we continue to refine and expand EDMRS, its impact on the music industry and user experience is poised to be transformative, ushering in a new era of personalized, emotionally connected music exploration.

Acknowledgement

We extend our sincere gratitude to all those who contributed to this project's inception and development. We would like to acknowledge the support and guidance of our mentors, collaborators, and research partners who shared their expertise and insights. This project would not have been possible without the dedication and effort of our team

members who worked tirelessly to bring EDMRS to life. We also express our appreciation to the research community for inspiring us with their valuable work in the fields of facial expression analysis, sentiment analysis, and music recommendation systems.

References

[1] Saeed,S., Shah, A.A., Ehsan, M.K.,Amirzada, M.R., Mahmood, A., & Mezgebo, T (2022). Automated Facial Expression Recognition Framework Using Deep Learning. IEEE Access, Journal of Healthcare Engineering,Volume 2022,ArticleID5707930. DOI:https://doi.org/10.1155/2022/5707930

[2] D. H. Lee and J. H. Yoo, "CNN Learning Strategy for Recognising Facial Expressions," in IEEE Access, vol. 11, pp. 26040-26053, 2023, doi: 10.1109/ACCESS.2023.3294099.

[3] Ying He, Yanxia Zhang, Shuxin Chen, and Yu Hu, "Facial Expression Recognition Using Hierarchical Features With Three-Channel Convolutional Neural Network," in IEEE Access, vol.11pp.79360-79372,2023,DOI: 10.1109/ACCESS.2023.3303402.

[4] Taimur Shahzad, Khalid Iqbal, Murad Ali khan, Imran, and Naeem Iqbal, "Role of Zoning in Facial Expression Using Deep Learning," IEEE Access, vol. 11,pp.1-16,2023,DOI: 10.1109/ACCESS.2023.3243850.

[5] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55, 5731–5780. DOI: https://doi.org/10.1007/s10462-022-10144-1.

[6] L. Kessous, G. Castellano, and G. Caridakis, ''Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis,'' J. Multimodal User Interfaces, vol. 3, nos. 1–2, pp. 33–48, Mar. 2010.

[7] S. Ahmad and D. H. Kim, ''A task orchestration approach for efficient mountain fire detection based on microservice and predictive analysis in IoT environment,'' J. Intell. Fuzzy Syst., vol. 40, no. 3, pp. 5681–5696, Mar. 2021.

[8] G. Muhammad, M. Alsulaiman, S. U. Amin, A. Ghoneim, and M. F. Alhamid, "A facial-expression monitoring system for improved healthcare in smart cities," IEEE Access, vol5, pp. 10871–10 881, 2017.

[9] B. A. Mercan, "Doing criminological research: affective states versus emotional reactions," 3eoretical Criminology, vol. 24, no. 2, pp. 335–352, 2020.

[10] G. Giannakakis, M. Pediaditis, D. Manousos et al., "Stress and anxiety detection using facial cues from videos," Biomedical Signal Processing and Control, vol31, pp. 89–101, 2017