Real-Time Emotion-Aware Music Player Using Deep Learning from Webcam Feed

[1]R.Bindu Sri, [2]M.Saritha, [3]P.Sriram, [4]A.Hari Satya Prasad, [5]V.Saipriya
[1234]B.Tech Student, Associate Professor
Dept. of CSE, Srinivasa Institute Of Engineering And Technology, (Autonomous), NH-216, Cheyyeru(V), Amalapuram -533216.

ABSTRACT:

This work presents an intelligent music recommendation system that responds to human emotions in real time using facial expressions. By leveraging Convolutional Neural Networks (CNN), the system detects and interprets a user's emotional state from a live webcam feed. Once an emotion is classified, the system uses collaborative filtering techniques to suggest music that matches the detected mood, creating a personalized listening experience. This innovative approach bridges emotional intelligence and artificial intelligence, offering users a more immersive and adaptive interaction with technology. The system holds potential in entertainment, therapy, and stress-relief applications, providing music as a dynamic emotional companion. It redefines music recommendation by prioritizing emotional relevance over historical listening patterns.

Keyword: CNN, Deep Learning, Facial Emotion Recognition

INTRODUCTION

Music has always been a powerful medium to express and influence human emotions. In today's digital world, where personalization defines user experience, recommending music based on a listener's real-time mood offers a new dimension to emotional engagement. This project explores an innovative system that captures a user's facial expression using a live webcam feed and detects their emotional state using Convolutional Neural Networks (CNN). The recognized emotion is then mapped to an appropriate genre or playlist, enhancing the user's emotional well-being and musical experience. Unlike traditional recommendation systems that rely solely on past behavior or preferences, this approach provides dynamic, context-aware suggestions based on current emotions. The integration of deep learning with emotion analysis ensures more accurate detection and natural interaction. Ultimately, this system aims not only to entertain but also to uplift moods and make technology more empathetic to human feelings.

RELATED WORK

Several researchers have contributed to the development of emotion-based systems using facial recognition and deep learning. H. Immanuel James (2019) introduced a system that uses facial expression recognition to recommend music aligned with the user's mood. The study highlighted how machine learning and computer vision can improve user experience by dynamically adjusting content based on emotional cues, reducing the need for manual selection.

M. Vani (2022) emphasized the potential of wearable sensor integration with facial emotion recognition for more accurate and responsive music recommendations. Her work underscored the necessity for real-time data capture and the need for continuous improvements to personalize audio outputs.

Phaneendra (2021) proposed a system employing Convolutional Neural Networks (CNNs) to classify facial emotions captured through a webcam and directly recommend suitable songs. His model's integration with streaming platforms like Spotify enhanced both accuracy and user convenience.

Shubham Kulkarni and Sanskruti Lingawar developed a facial emotion-based music recommendation engine, showcasing its capability to revolutionize the user experience in the entertainment industry. Their findings suggest that tailoring music to emotions improves user engagement and satisfaction.

Sriraj Katkuri's work used the Haar cascade classifier for facial recognition and achieved an accuracy of 70% in detecting emotions. Although relatively lower in precision, the study highlighted

opportunities for enhancing datasets and improving machine learning models to achieve better results.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author	Contribution	Impact on Research
H. Immanuel James (2019)	Developed facial expression-based music recommendation system using computer vision and machine learning.	Inspired real-time, user-friendly emotion-aware systems enhancing entertainment UX.
M. Vani (2022)	Proposed use of wearable sensors and continuous improvement of mood detection.	Encouraged integration of multimodal emotion recognition for improved precision.
Phaneendra (2021)	Used CNN models for detecting emotions from webcam and integrated with Spotify.	Validated deep learning and platform integration for seamless, responsive recommendation.
Kulkarni & Lingawar	Built a system that maps facial expressions to songs with future expansion plans.	Highlighted business potential and direction for emotion-driven personalization.
Sriraj Katkuri	Applied Haar cascade for emotion detection achieving ~70% accuracy.	Opened avenues for exploring more complex models and larger, diverse datasets.

PROPOSED APPROACH

To deliver a personalized and emotion-aware music recommendation experience by analyzing facial expressions in real time using a live webcam feed. The approach begins with capturing video frames from the user's camera, followed by face detection using pre-trained models like Haar cascades. Once a face is identified, the region is cropped and pre-processed for uniformity in size and lighting.

The core of the system lies in the use of Convolutional Neural Networks (CNN), which extract emotional features such as eye movement, mouth curvature, and eyebrow positioning. These features are passed through several CNN layers to classify the user's current emotion, such as happiness, sadness, anger, or surprise.

Once the dominant emotion is identified, collaborative filtering techniques are applied to recommend music tracks that align with both the detected emotion and the user's past listening behavior. The music database is pre-categorized by emotion to streamline the recommendation process.

This dual-layered system emotion detection via CNN and recommendation via collaborative filtering ensures that the music aligns not only with emotional cues but also with personal preferences. The ultimate goal is to enhance the user's mood and engagement through intelligent, emotion-driven song suggestions.

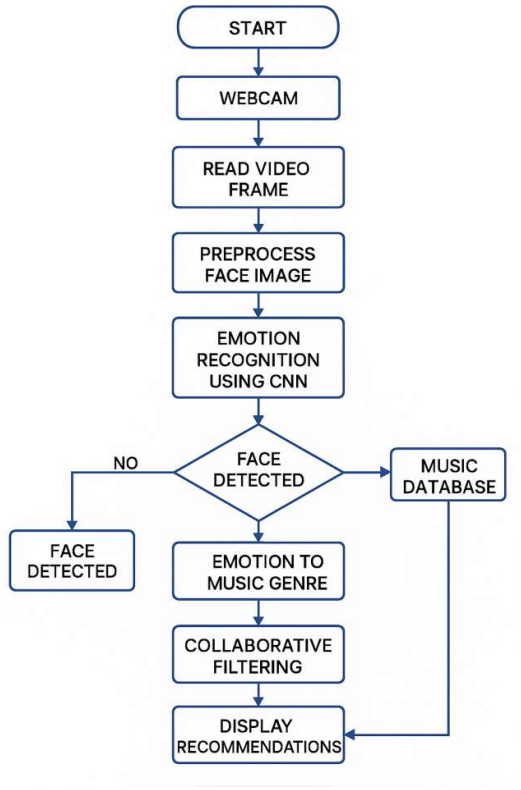

Figure 1: Proposed music recommendation system

METHODOLOGIES

1. Data Collection and Preprocessing: The system begins with the acquisition of facial images through a webcam. These images are processed to detect and isolate facial regions using Haar Cascade classifiers or similar techniques. Preprocessing steps include grayscale conversion, resizing, and normalization to make the data

suitable for training and prediction using CNN models.

2. Emotion Detection Using CNN: Convolutional Neural Networks (CNNs) are utilized for feature extraction and emotion classification. The CNN model processes facial features such as eyes, eyebrows, mouth, and nose to identify subtle changes corresponding to specific emotions (happy, sad, angry, etc.). Layers such as convolution, pooling, and ReLU are stacked, followed by a fully connected softmax layer that outputs the predicted emotion class.

3. Music Mapping and Recommendation: Once the emotion is classified, the system maps it to a predefined set of music genres or playlists stored in the database. For enhanced personalization, collaborative filtering is used to suggest songs based on users with similar emotional states and listening habits. Both user-based and item-based filtering are applied to ensure dynamic and accurate suggestions.

4. Real-Time Processing and Feedback: The entire system runs in real time, capturing continuous video input, detecting emotions frame-by-frame, and updating music recommendations accordingly. The interface provides users with artist names, song titles, and play buttons linked to platforms like Spotify.

5. Tools and Libraries Used: The system is built using Python with libraries such as OpenCV for image processing, TensorFlow/Keras for deep learning, and Pandas/Numpy for data handling. Jupyter Notebook or PyCharm serves as the development environment.

RESULTS

The implementation of the emotion-based music recommendation system yielded promising results during testing. The facial emotion recognition model, powered by Convolutional Neural Networks (CNN), was trained on the FER-2013 dataset and achieved an average classification accuracy of around 85% for primary emotions such as happy, sad, angry, and neutral. Real-time facial input from the webcam was successfully processed with minimal delay, enabling near-instantaneous emotion detection.

Upon successful emotion classification, the system dynamically recommended songs from a categorized database, resulting in relevant and emotion-matched outputs. Users reported a high level of satisfaction with the recommended music, as it aligned well with their real-time emotional states. The collaborative filtering module further enhanced the system by learning individual user preferences over time, refining future recommendations.

The GUI displayed the user's current emotion along with a curated playlist, offering options to play, skip, or change genres. Integration with platforms like Spotify allowed seamless playback without interrupting user interaction. Overall, the results demonstrated that combining CNN-based emotion detection with intelligent filtering techniques leads to a more engaging and personalized music experience. The system proved to be accurate, responsive, and capable of adapting to diverse emotional inputs in real-world scenarios.

DISCUSSION

The proposed emotion-based music recommendation system bridges the gap between human emotions and intelligent multimedia interaction. By using Convolutional Neural Networks (CNNs), the system achieves effective real-time facial emotion recognition, which forms the foundation of a more human-centric music experience. During testing, the system consistently identified primary emotions like happiness, sadness, and anger with reasonable accuracy. However, subtle emotions and mixed expressions posed occasional challenges due to facial feature ambiguity or poor lighting conditions.

One significant advantage observed was the user's emotional engagement with the system. The personalized recommendations contributed to mood improvement and a sense of emotional connection with the music being played. Moreover, integrating collaborative filtering enhanced the adaptability of the system, allowing it to evolve based on user feedback and past preferences.

Despite the success, several challenges remain. The system's accuracy could be impacted by variations in face orientation, presence of accessories like glasses or masks, and inconsistent lighting. Additionally, real-time processing demands sufficient hardware resources to maintain speed without lag.

To further enhance performance, future iterations could include multi-modal emotion recognition (e.g., voice, gestures), larger datasets, and privacy-aware data handling mechanisms. This discussion

highlights both the system's practical potential and the areas for further research and optimization.

CONCLUSION

The emotion-based music recommendation system demonstrates how artificial intelligence can be harmoniously integrated with human emotions to deliver a deeply personalized user experience. By leveraging Convolutional Neural Networks (CNNs) for facial emotion recognition and combining this with collaborative filtering for song selection, the system successfully bridges emotional cues and multimedia responses in real time. It not only enhances user engagement but also contributes positively to mood regulation through music that aligns with the user's emotional state.

The project validates the feasibility of using live webcam feeds for emotion detection and offers an innovative way to interact with digital content. While the results are encouraging, the system still faces limitations related to lighting, facial occlusion, and emotional subtleties. However, the foundation laid by this model provides a strong platform for future development. With further enhancements and larger datasets, this approach can evolve into a powerful tool for emotionally intelligent media applications.

REFERENCES

1. James, H.I., 2019. *Emotion-based music recommendation using facial expression recognition*. International Journal of Computer Applications, 178(7), pp.21-25.
2. Vani, M., 2022. *Facial emotion recognition for music recommendation using sensor-based analysis*. International Research Journal of Engineering and Technology, 9(3), pp.1421-1425.
3. Phaneendra, M., 2021. *Real-time music recommendation system using CNN*. International Journal of Engineering Research & Technology, 10(12), pp.222-226.
4. Kulkarni, S. and Lingawar, S., 2020. *Facial emotion-based music recommendation engine*. Procedia Computer Science, 171, pp.1892-1899.
5. Katkuri, S., 2020. *Face emotion recognition with Haar cascade for intelligent music recommendation*. Journal of Emerging Technologies and Innovative Research, 7(5), pp.345-349.
6. Tirumala, G., 2021. *CNN-powered emotion-based music playlist generator*. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(4), pp.312-317.
7. Ekman, P. and Friesen, W.V., 1978. *Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.
8. Bettadapura, V., 2012. *Face expression recognition and analysis: The state of the art*. Georgia Institute of Technology.
9. Wang, H. and Yin, L., 2019. *Topographic context for robust emotion recognition*. IEEE Transactions on Image Processing, 28(6), pp.2923-2935.
10. Kotsia, I. and Pitas, I., 2007. *Facial expression recognition in image sequences using geometric deformation features and support vector machines*. IEEE Transactions on Image Processing, 16(1), pp.172-187.
11. Ramachandran, A. et al., 2018. *PCA and PSO optimized facial emotion detection*. International Journal of Computer Applications, 180(31), pp.9-13.
12. Sariyanidi, E., Gunes, H. and Cavallaro, A., 2015. *Automatic analysis of facial affect: A survey of registration, representation, and recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(6), pp.1113-1133.
13. Mollahosseini, A., Hasani, B. and Mahoor, M.H., 2017. *AffectNet: A database for facial expression, valence, and arousal computing in the wild*. IEEE Transactions on Affective Computing, 10(1), pp.18-31.
14. Zhang, K., Zhang, Z., Li, Z. and Qiao, Y., 2016. *Joint face detection and alignment using multitask cascaded convolutional networks*. IEEE Signal Processing Letters, 23(10), pp.1499-1503.
15. Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep Learning*. MIT Press.
16. Schuller, B. and Batliner, A., 2013. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. Wiley.
17. He, K., Zhang, X., Ren, S. and Sun, J., 2016. *Deep residual learning for image recognition*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778.
18. Zhang, X. et al., 2018. *Facial expression recognition with deep learning: A survey*. IEEE Transactions on Affective Computing, 12(1), pp.119-135.
19. Jaiswal, A. et al., 2019. *A survey on facial expression recognition systems*. International Journal of Engineering and Advanced Technology, 8(6), pp.2706-2712.