Facial Emotion Recognition for Music Personalization: A Comprehensive Review

Nagamalla Akshitha[1] ,Sri. P . Sreenivasa Rao[2]

[1]PGscholor, Department of CSE, JNTUH College of Engineering Jagtial . Nachupally (Kondagattu),

[2]Associate professor, Department of CSE, JNTUH College of Engineering Jagtial , Nachupally,(Kondagattu),

ABSTRACT

In the evolving field of affective computing, integrating emotion recognition with content personalization has paved the way for intelligent multimedia systems. Facial Emotion Recognition (FER), powered by deep learning, enables real-time assessment of human emotional states through facial expressions, making it highly effective for applications like personalized music recommendation. This review explores the development and implementation of Fully Connected Neural Network (FCNN)-based frameworks designed to generate music playlists in real time based on detected emotions.Recent studies demonstrate that deep learning models, especially Convolutional Neural Networks (CNNs) combined with FCNN layers, outperform traditional machine learning techniques such as Support Vector Machines (SVM) and PCA in emotion classification accuracy and processing speed. Lightweight models have been emphasized for real-time deployment, with some architectures achieving over 90% accuracy while maintaining low computational overhead. The combination of datasets like FER2013 and CK+48 has further enhanced model generalization.Moreover, emotion-based music recommendation systems have shown promise in improving user mood, engagement, and mental well-being. FCNN layers, as part of the decision-making module in such systems, map detected emotions to curated music tracks effectively by processing the abstract features extracted by convolutional layers.This paper provides a critical overview of the current state-of-the-art in emotion-driven music recommendation, focusing on the architecture, performance, and real-time capabilities of FCNN-based systems. Challenges such as emotion ambiguity, dataset bias, and personalization are discussed, alongside future directions involving multimodal emotion input and deployment on edge devices.

Keywords: *Facial Emotion Recognition (FER), Fully Connected Neural Network (FCNN),Real-Time Emotion Detection, Music Recommendation System.Deep Learning.Convolutional Neural Networks (CNN)*

1. Introduction

In the realm of human-computer interaction, the ability of machines to perceive and interpret human emotions has become a cornerstone of affective computing. Among the various modalities for emotion recognition, **Facial Emotion Recognition (FER)** has garnered significant attention due to its non-intrusive nature and high expressiveness of facial cues. FER systems use visual information to classify emotional states such as happiness, sadness, anger, fear, and surprise. The effectiveness of FER has led to its integration into diverse applications, including mental health monitoring, intelligent gaming, and multimedia personalization.Recent advancements in **deep learning**, particularly **Convolutional Neural Networks (CNNs)**, have propelled the performance of FER systems. For example, Mehrotra et al. [1] proposed a CNN-based approach that emphasized low computational cost while maintaining robust recognition accuracy. Their model demonstrated the potential of streamlined CNNs in real-time applications by significantly reducing training time. Similarly, Mishra et al. [2] explored multiple deep learning frameworks for FER and highlighted that CNNs could extract subtle features from facial images, leading to improved recognition rates even under varying lighting and pose conditions.The fusion of FER with **personalized content delivery**, such as music, presents a promising direction for emotionally adaptive systems. Music is widely recognized for its impact on emotional regulation and mental well-being. However, manually selecting music that aligns with a user's current mood can be cumbersome and ineffective, especially during emotional distress. This gap has motivated researchers to explore **emotion-aware music recommendation systems**, wherein the detected emotional state drives the selection of mood-congruent music tracks.Bhoomika et al. [3]

conducted a comparative study of pre-trained deep learning models such as VGG16 and DenseNet121 for FER and found that these models, while accurate, are computationally intensive and less suitable for real-time or embedded scenarios. To address these challenges, lightweight models such as **MobileNetV2** have been proposed. For instance, Gadagkar et al. [4] developed a CNN-based emotion recognition module integrated with MobileNetV2 that performed real-time music recommendations on mobile platforms. While effective, such models still face limitations when balancing speed, accuracy, and resource usage.

On the other hand, hybrid approaches combining traditional machine learning with feature extraction methods have also demonstrated promise. Seshaayini et al. [5] proposed a system that employed **Principal Component Analysis (PCA)** for dimensionality reduction, followed by a **Support Vector Machine (SVM)** classifier, achieving 100% accuracy in emotion-based music selection under controlled conditions. While impressive, such accuracy is often constrained by dataset quality and lack of real-time adaptability.To overcome the limitations of heavy models and to ensure real-time responsiveness, this paper introduces **Emotify**, a lightweight, real-time facial emotion recognition system integrated with a music player. The proposed system is based on a custom-designed CNN trained on a hybrid of **FER2013** and **CK+48** datasets. It detects six primary emotions—*Happy, Sad, Anger, Fear, Surprise*, and *Neutral*—and automatically recommends and plays music using the **YouTube Data API**.The novelty of this work lies in its balance of **accuracy, speed, and deployment efficiency**, offering a personalized musical experience driven by the user's emotional state. The system supports three modes of interaction: real-time webcam capture, local image upload, and manual search, making it adaptable to various user preferences and environments.

II.Literature Survey

Facial Emotion Recognition (FER) has emerged as a key domain within affective computing and human-computer interaction, enabling systems to identify and respond to human emotional states based on facial expressions. Recent research has significantly advanced this field, particularly through the adoption of deep learning techniques such as Convolutional Neural Networks (CNNs).

A CNN-based model trained on FER2013 and AffectNet datasets achieved 71.61% accuracy while effectively minimizing training time, illustrating the potential of lightweight architectures for facial emotion detection [1]. In another study, VGG16 and VGG19 architectures were employed, with VGG19 achieving an accuracy of 92.5%, thereby demonstrating the capability of deeper networks in extracting fine-grained facial features [2]. DenseNet121 also exhibited strong performance, surpassing ensemble models with an accuracy of 86% [3].

MobileNetV2, known for its efficiency in mobile environments, was used alongside a CNN-based emotion recognition module to deliver real-time music recommendations, showing the potential for FER in resource-constrained settings [4]. Another approach integrated PCA for dimensionality reduction and used an SVM with a polynomial kernel, attaining 100% accuracy in mood-based music selection, highlighting the effectiveness of hybrid models in specific tasks [5].

Comprehensive reviews in the field of FER emphasized the growing need for systems to recognize not only basic emotions but also complex affective states. These studies advocated the use of multimodal approaches and attention mechanisms to improve the reliability of emotion recognition in dynamic environments [6][7]. For example, a CNN model reported 92% training accuracy, 88% validation accuracy, and 85% test accuracy after 20 epochs, showcasing good generalization performance [8]. In the educational sector, FER methods have been explored for student emotion analysis, distinguishing between manual annotation and machine learning-based techniques [9].

In the context of music recommendation, several studies have merged emotion detection with automated playlist generation. One such system achieved 66% accuracy, with users expressing satisfaction with emotion-based playlists [10]. Another approach used real-time video analysis combined with musical features like tempo and pitch for emotion-to-music mapping [11]. Systems developed using OpenCV, TensorFlow, and Tkinter classified facial expressions and suggested songs accordingly, operating efficiently in real-time environments [12].

A Streamlit-based web application utilized webcam inputs for FER, matching detected moods with appropriate music selections [13]. In another study, CNN models trained on the FER dataset recognized emotions such as happiness, fear, sadness, and surprise with accuracies of 96%, 97%, 93%, and 94% respectively [14]. Further, systems using CNN architectures for facial emotion detection demonstrated improved performance in both accuracy and time efficiency for music recommendation tasks [15].

Comparative studies revealed that CNNs outperform traditional models like SVMs, with reported accuracies of 81% and 77%, respectively [16]. A 6-layer CNN with max pooling achieved 83% accuracy for detecting emotions such as happy, sad, and neutral [17]. Other real-time systems based on convolutional and fully connected layers aimed to enhance user satisfaction through personalized music recommendations [18].

Multiple classifiers have also been assessed for music emotion recognition. Random Forest achieved the highest success rate at 75%, followed by SVM (63%), Decision Tree (60%), and Naive Bayes (50%), indicating that machine learning classifiers can still be relevant with proper feature selection [19].

More recent studies explored advanced FER systems, such as those achieving 93.32% accuracy using novel CNN architectures [20]. Face detection techniques like MTCNN and embedding-based models like FaceNet have been used to improve emotion prediction through enhanced facial feature extraction [21]. Additionally, chatbot-based systems have been proposed to assess user emotions via conversation and generate mood-specific playlists [22].

Transfer learning approaches using VGG19 and datasets like CK+, JAFFE, and FER2013 have achieved impressive accuracies—99% on CK+, 93% on JAFFE, and 65% on FER2013—highlighting the importance of dataset diversity in FER [23]. Hybrid models like FERConvNet_HDM, which combine denoising methods with CNNs, have achieved 85% accuracy on FER2013 and 95% on low-resolution datasets, illustrating advancements in handling image noise and resolution limitations [24][25].

TABLE 1: Summery Table

Ref No	**Model/Technique**	**Dataset**	**Key Outcome**	**Application Area**
1	CNN	FER2013	Low computation cost, 71.61% accuracy	Real-time FER
2	CNN	Unspecified	Deep learning for robust FER	Emotion classification
3	VGG16, VGG19, DenseNet121	Unspecified	VGG19 achieved 92.5% accuracy	Model comparison
4	MobileNet V2 + CNN	Unspecified	Real-time mobile emotion recognition	Music recommendation
5	PCA + SVM	Unspecified	100% accuracy in music selection	Hybrid approach
6	CNNs	Review	Advocated multimodal FER	DL in FER (survey)
7	ResNet50, Inception	Unspecified	Attention mechanisms for dynamic expression recognition	Dynamic FER
8	CNN	FER2013	Training: 92%, Validation: 88%, Test: 85%	Generalized CNN model
9	ML methods	Educational FER	Manual vs ML-based emotion annotation	Academic context
10	CNN	Unspecified	66% accuracy; users satisfied with playlist	FER-based playlist generation
11	CNN	Unspecified	Used tempo, pitch, and	Emotion-to-music mapping

			volume for mapping	
12	CNN + OpenCV	Unspecified	Live FER + music suggestion	Real-time interface
13	CNN	Unspecified	Streamlit GUI; webcam capture	Web-based FER system
14	CNN	FER	Happy: 96%, Fear: 97%, Sad: 93%, Surprise: 94%	High accuracy FER model
15	CNN	Unspecified	Fast, optimized emotion-based music recommendation	Efficient FER+music system
16	CNN vs. SVM	FER	CNN: 81%, SVM: 77%	Model performance comparison
17	6-layer CNN	Unspecified	83% accuracy for happy/sad/neutral	FER-based music player
18	CNN	Unspecified	Improved user satisfaction in real-time setup	Emotion-aware music player
19	RF, SVM, DT, NB	Unspecified	RF: 75%, SVM: 63%, DT: 60%, NB: 50%	ML classifier comparison
20	CNN	Unspecified	93.32% accuracy for music player	DL-based recommendation
21	MTCNN + FaceNet	Unspecified	Facial embedding enhanced accuracy	FER + music recommendation
22	Chatbot	Conversational data	Playlist generation via emotion-based dialog	NLP-based emotion detection
23	VGG19 (Transfer Learning)	CK+, JAFFE, FER2013	99% (CK+), 93% (JAFFE), 65% (FER2013)	Sentiment from images
24	FERConvNet_HDM + denoising	FER2013, LRFE	85% (FER2013), 95% (LRFE)	Low-res FER improvement
25	Hybrid CNN + filtering	FER2013	85% accuracy	Noise-robust FER

III.Conclusion

The integration of facial emotion recognition (FER) with intelligent music recommendation systems represents a significant advancement in the field of affective computing and human-centered AI. This review analyzed a wide range of methodologies—from traditional machine learning classifiers like SVM and PCA to advanced deep learning architectures such as VGG19, MobileNetV2, and hybrid CNN models. The studies reviewed consistently demonstrate that deep learning, particularly convolutional neural networks followed by fully connected layers (FCNN), provides superior accuracy, robustness, and generalization capabilities for real-time emotion recognition tasks.

While complex models such as VGG19 and DenseNet121 achieve high accuracy, they often demand substantial computational resources, making them less suitable for real-time or mobile applications. On the other hand, lightweight and optimized models—such as MobileNetV2-based frameworks or custom shallow CNN+FCNN combinations—have emerged as promising solutions, offering a balance between accuracy and efficiency.

Emotion-based music recommendation systems, which map recognized emotions to curated playlists, have shown great potential in enhancing user satisfaction, mood regulation, and mental well-being. However, key challenges remain, including emotion ambiguity, personalization limitations, and reliance on single-modal inputs.

To address these issues, future research should focus on multimodal emotion recognition (e.g., combining facial cues with voice or physiological signals), context-aware playlist generation, and deployment on edge devices for real-world accessibility. With continued advancements in FCNN-based architectures and emotional intelligence modeling, such systems are well-positioned to become integral components of personalized digital experiences.

IV.REFERENCES:

[1] M. Mehrotra, K. P. Singh and Y. B. Singh, "Facial Emotion Recognition and Detection Using Convolutional Neural Networks with Low Computation Cost," *2024 2nd Int. Conf. on Disruptive Technologies (ICDT)*, Greater Noida, India, 2024, pp. 1349–1354, doi: 10.1109/ICDT61202.2024.10489678.

[2] P. Mishra, A. S. Verma, P. Chaudhary and A. Dutta, "Emotion Recognition from Facial Expression Using Deep Learning Techniques," *2024 IEEE 9th Int. Conf. for Convergence in Technology (I2CT)*, Pune, India, 2024, pp. 1–6, doi: 10.1109/I2CT61223.2024.10543313.

[3] G. Bhoomika, V. D. Pujitha, M. Sindusha, C. S. Rekha and B. Suvarna, "Facial Emotion Recognition: A Comparative Study of Pre-trained Deep Learning Models," *2024 IEEE 3rd World Conf. on Applied Intelligence and Computing (AIC)*, Gwalior, India, 2024, pp. 377–382, doi: 10.1109/AIC61668.2024.10730808.

[4] A. V. Gadagkar, S. S, S. Begum and A. S. M, "Emotion Recognition and Music Recommendation System based on Facial Expression," *2024 2nd Int. Conf. on Advances in Information Technology (ICAIT)*, Chikkamagaluru, India, 2024, pp. 1–6, doi: 10.1109/ICAIT61638.2024.10690441.

[5] K. Seshaayini, S. L. Srinithya, P. Verma, A. Visalatchi and N. Neelima, "Emotion Recognition Based Music Player," *2023 5th Int. Conf. on Electrical, Computer and Communication Technologies (ICECCT)*, Erode, India, 2023, pp. 01–05, doi: 10.1109/ICECCT56650.2023.10179716.

[6] S. Deshmukh et al., "Advances in Facial Emotion Recognition: Deep Learning Approaches and Future Prospects," *2024 MITADT School of Computing Int. Conf. (MITADTSoCiCon)*, Pune, India, 2024, pp. 1–3, doi: 10.1109/MITADTSoCiCon60330.2024.10574908.

[7] C. M. Prashanth et al., "Facial Emotion Detection: A Comprehensive Survey," *2024 Int. Conf. on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)*, Coimbatore, India, 2024, pp. 80–86, doi: 10.1109/ICC-ROBINS60238.2024.10533999.

[8] Uneza, D. Gupta and S. Saini, "Facial Expression Analysis: Unveiling the Emotions Through Computer Vision," *2024 11th Int. Conf. on Reliability, Infocom Technologies and Optimization (ICRITO)*, Noida, India, 2024, pp. 1–5, doi: 10.1109/ICRITO61523.2024.10522418.

[9] B. Fang, X. Li, G. Han and J. He, "Facial Expression Recognition in Educational Research From the Perspective of Machine Learning: A Systematic Review," *IEEE Access*, vol. 11, pp. 112060–112074, 2023, doi: 10.1109/ACCESS.2023.3322454.

[10] R. Shekhawat et al., "Syncing Music With Emotions Using Computer Vision," *2024 7th Int. Conf. on Circuit Power and Computing Technologies (ICCPCT)*, Kollam, India, 2024, pp. 522–527, doi: 10.1109/ICCPCT61902.2024.10672743.

[11] M. S. Guthula and M. Bordoloi, "Music Recommendation System Using Facial Detection Based Emotion Analysis," *2024 Int. Conf. on Emerging Techniques in Computational Intelligence (ICETCI)*, Hyderabad, India, 2024, pp. 296–301, doi: 10.1109/ICETCI62771.2024.10704201.

[12] S. Saranya et al., "Emotion Based Music Recommendation System," *2024 Int. Conf. on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC)*, Coimbatore, India, 2024, pp. 255–260, doi: 10.1109/ICSSEECC61126.2024.10649404.

[13] M. M. Joseph et al., "Emotion Based Music Recommendation System," *2023 Int. Conf. on Computational Intelligence and Knowledge Economy (ICCIKE)*, Dubai, UAE, 2023, pp. 505–510, doi: 10.1109/ICCIKE58312.2023.10131874.

[14] J. S. Joel et al., "Emotion Based Music Recommendation System using Deep Learning Model," *2023 Int. Conf. on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2023, pp. 227–232, doi: 10.1109/ICICT57646.2023.10134389.

[15] M. K. Singh, P. Singh and A. Sharma, "Facial Emotion Based Automatic Music Recommender System," *2023 Int. Conf. on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*, Ghaziabad, India, 2023, pp. 699–703, doi: 10.1109/ICSEIET58677.2023.10303573.

[16] P. Sudhakaran, P. K. Nair and A. Suraj, "Music Recommendation using Emotion Recognition," *2022 IEEE 2nd Mysore Sub Section Int. Conf. (MysuruCon)*, Mysuru, India, 2022, pp. 1–7, doi: 10.1109/MysuruCon55714.2022.9972635.

[17] R. Arya, C. Bhatt and M. Mittal, "Music Player Based on Emotion Detection Using CNN," *2022 IEEE North Karnataka Subsection Flagship Int. Conf. (NKCon)*, Vijaypur, India, 2022, pp. 1–5, doi: 10.1109/NKCon56289.2022.10126761.

[18] B. Zhai, B. Tang and S. Cao, "Music Recommendation System Based on Real-Time Emotion Analysis," *2022 Int. Conf. on Culture-Oriented Science and Technology (CoST)*, Lanzhou, China, 2022, pp. 334–338, doi: 10.1109/CoST57098.2022.00075.

[19] D. Unni, A. M. D'Cunha and D. G, "A Technique to Detect Music Emotions Based on Machine Learning Classifiers," *2022 2nd Int. Conf. on Interdisciplinary Cyber Physical Systems (ICPS)*, Chennai, India, 2022, pp. 136–140, doi: 10.1109/ICPS55917.2022.00033.

[20] P. Vijayeeta and P. Pattnayak, "A Deep Learning Approach for Emotion Based Music Player," *2022 OITS Int. Conf. on Information Technology (OCIT)*, Bhubaneswar, India, 2022, pp. 278–282, doi: 10.1109/OCIT56763.2022.00060.

[21] O. Ghosh et al., "Music Recommendation System based on Emotion Detection using Image Processing and Deep Networks," *2022 2nd Int. Conf. on Intelligent Technologies (CONIT)*, Hubli, India, 2022, pp. 1–5, doi: 10.1109/CONIT55038.2022.9847888.

[22] A. Nair et al., "Emotion Based Music Playlist Recommendation System using Interactive Chatbot," *2021 6th Int. Conf. on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2021, pp. 1767–1772, doi: 10.1109/ICCES51350.2021.9489138.

[23] G. Meena, K. K. Mohbey, A. Indian and S. Kumar, "Sentiment Analysis from Images using VGG19 based Transfer Learning Approach," *Proc. of Int. Conf. on Industry Sciences and Computer Science Innovation*, Ajmer, India, *Procedia Computer Science*, vol. 204, pp. 411–418, 2022, doi: 10.1016/j.procs.2022.08.050.

[24] P. N. R. Bodavarapu and P. V. V. S. Srinivas, "Facial Expression Recognition for Low Resolution Images using CNN and Denoising Techniques," *Indian Journal of Science and Technology*, vol. 14, no. 12, pp. 971–983, 2021, doi: 10.17485/IJST/v14i12.14.

[25] P. N. R. Bodavarapu and P. V. V. S. Srinivas, "An Optimized Neural Network Model for Facial Expression Recognition over Traditional Deep Neural Networks," *Int. Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 7, 2021, doi: 10.14569/IJACSA.2021.0120751