Moodify : Adaptive Music Selection by Facial Expression

ABHISHEK RAJ C K , ADARSHA K, BHUVAN T J R, ASHISH GOWDA G C, Dr.CHANDRASHEKAR D K

Dr.Chandrashekar D K, Associate professor, SJB Institute of Technology, Banglore,India

Abhishek Raj C K, Dept. Of CSE(DS), SJB Institute of Technology, Banglore,India

Adarsha K, Dept. Of CSE(DS), SJB Institute of Technology, Banglore,India

Bhuvan T J R, Dept. Of CSE(DS), SJB Institute of Technology, Banglore,India

Ashish Gowda G S, Dept. Of CSE(DS), SJB Institute of Technology, Banglore,India

---***---

Abstract : Moodify is an emotionally brave music selection system that personalizes audio experiences by detecting users' facial expressions in real-time. Bridging affective computing and music information retrieval, the system leverages computer moves and deep learning techniques—primarily Convolutional Neural Networks (CNNs)—to classify emotional states such as happiness, sadness, anger, and calmness. Based on these inputs, Moodify curates adaptive playlists aligned with or intended to shift the user's mood, creating a dynamic and empathetic listening environment. The system integrates with platforms like Spotify via API, supports real-time emotion monitoring, and adapts music accordingly. Developed using a modular architecture of ReactJS, Flask, and TensorFlow, Moodify emphasizes privacy, scalability, and cross-device compatibility. Beyond entertainment, its applications span mental health, education, elderly care, and automotive systems. By transforming passive listening into an emotionally aware interaction, Moodify exemplifies the future of human-centric, affective technology in digital media experiences

***Key Words*:** Facial Emotion Recognition (FER),Convolutional Neural Networks (CNNs),Adaptive Music Recommendation, Affective Computing

1. INTRODUCTION:

Adaptive music choice means changing the music a person hears depending on their current situation, feelings, or likes. Unlike normal music apps that play songs based only on fixed things—like your favorite genres, saved playlists, or what you've listened to before—adaptive systems try to be more lively and responsive. They adjust the music in the moment by noticing signals such as your actions, mood, expressions, the time of day, or even where you are.

The most exciting directions in adaptive music systems is emotion-based recommendations, where the platform senses how a listener feels and then chooses music to match that state. To detect emotions, the system may rely on sign like facial expressions, tone of voice, or even body signals such as heart rate and skin responses. Advanced methods, especially deep learning design like Convolution Neural Networks (CNNs), are widely used forrecognizing facial emotions and can accurately classify feelings such as joy, sadness, anger, fear, or neutrality. Once these emotional states are identified, they are connected to different musical elements—like rhythm, tempo, positivity (valence), or lyrics—so the system can play songs that either reflect the listener's mood or gently guide it in a new direction.

Real-time adaptability is a key feature of modern adaptive music systems. Emotions are not static—they fluctuate frequently due to internal and external stimuli. Systems like *Moodify* monitor facial expressions continuously and update playlists dynamically based on emotional transitions. For instance, if a user initially appears stressed but gradually relaxes, the music selection may shift from calming ambient tracks to more upbeat and energizing songs, thereby creating an emotionally synchronized experience.

The core technology behind this system is Facial Emotion Recognition (FER), a branch of computer vision and machine learning systems analyze facial features— such as eye shape, mouth curvature, eyebrow movement, and facial muscle contractions—to classify emotions like joyness, sadness, anger, fear, surprise, disgust, and neutrality. These systems often employ Convolution Neural Networks (CNN s) trained on large datasets such as FER-2013, AffectNet, or CK+. Real-time image processing is facilitated by tools like OpenCV, Dlib, and MediaPipe, which help detect and extract facial landmarks from webcam or mobile camera feeds.

2. Literature Review

Shanthakumari et.al.,[1], present a novel approach to improving music recommendations by applying deep learning techniques to detect user emotions from facial expressions. The goal of the system is to boost listener satisfaction by aligning music genres with the individual's emotional state. To achieve this, the researchers use a Convolution Neural Network (CNN) for emotion recognition, which effectively analyzes facial features to identify emotions such as happiness, sadness, and neutrality. Once these emotions are classified, they are linked to specific Spotify music genres, allowing the system to deliver highly personalized suggestions.

Shlok Gilda et.al.,[2], introduce an emotion-driven music recommendation system designed to enrich the listening experience through adaptive playlist creation. This system, called EMP, is built around three core components: the Emotion Module, the Music Classification Module, and the selection Module. The Emotion Module employs a Convolution Neural Network (CNN) to analyze facial expressions and classify emotions into four categories—happy, sad, angry, and neutral. Using the FER2013 dataset, which contains over 26,000 grayscale facial images, the model achieves an impressive accuracy rate of 90.23%.

Kevin Patel et.al.,[3], present a system that creates personalized song playlists by detecting a user's facial expression and examining song mood. The system aims to improve user experience by suggesting music that suits their emotional state and listening preferences. The proposed system includes four key phases: developing a facial emotion detection model, collecting user listening history through the Spotify API, classifying song moods, and creating customized playlists.

Sonika Malik et.al.,[4], a system that recommends songs by analyzing a user's facial expressions. The goal is to enhance user experience by automatically playing songs that match their emotional state, reducing the manual effort of selecting music. The system uses a *seven-layer Convolution Neural Network (CNN)* for emotion detection, which effectively identifies seven emotions:*happy, **angry, **surprised, **neutral, **fear,**disgust, and **sad*.

Anukriti Dureha et.al.,[5], put front a new algorithm that creates personalized music playlists by analyzing a user's facial expressions. The main aim of their work is to improve accuracy, minimize processing time, and avoid reliance on extra devices such as EEG sensors. The system combines two modules—Facial Expression Recognition (FER) and Audio Emotion Recognition (AER)—which are connected through a metadata file containing song information along with corresponding emotional labels.

Preema J. S. et.al.,[6], suggest a system that automatically curates music playlists by interpreting the user's facial expressions. The purpose of this method is to minimize the time and effort required to manually organize large collections of songs, allowing the system to show tracks that align with the listener's current mood.

S. Alhagry et.al., in [7] propose an EEG-based emotion recognition method utilizing Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNNs). Emotion, being vital for human communication and interaction, is difficult to capture through traditional means such as facial expressions or speech, which can be faked or ambiguous. EEG signals, reflecting electrical brain activity, offer a more reliable physiological source for emotion recognition.

D. Su et.al.,[8], present Adaptive Music for Affect Improvement (AMAI), a system designed to generate and play music with the goal of enhancing the listener's emotional well-being through dynamic audio feedback. Drawing on techniques from game music and emotion-aware computing, AMAI modifies musical features— such as tempo, harmony, and instrumentation—in real time, guided by the user's emotional valence detected through facial expression analysis.

R. Nand et.al.,[9], propose a Mood Music selection System that leverages deep learning–based emotion detection techniques to deliver personalized song suggestions aligned with the listener's present emotional state. In contrast to conventional systems that rely mainly on genre preferences or listening history, this model incorporates real-time emotion recognition using signals such as facial expressions, speech characteristics, and voice tone.

H. Kabani et.al.,in [10], introduce an Emotion-Based Music Player that plays user-specific playlists real-time based on the facial expressions of the user. Overcoming the shortcomings of classical playlist management, this approach combines audio feature extraction with facial emotion founder to provide automatic song selection.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

3. DATASET:

The Facial Expression Recognition (FER-2013) dataset from Kaggle is one of the most widely used resources in affective computing, particularly for developing and evaluating facial emotion recognition models. Originally released as part of a Kaggle competition held alongside the International Conference on Machine Learning (ICML) in 2013, it was created to advance research in deep learning and computer vision. Over time, FER-2013 has become a benchmark dataset for real-time emotion detection, supporting applications such as emotion-aware user interfaces, adaptive music recommendation systems, and mental health monitoring technologies.

The FER-2013 dataset consists of 35,887 grayscale facial images, each with a resolution of 48×48 pixels. These images capture close-up views of human faces displaying different expressions. Every image is annotated with one of seven fundamental emotions—Angry, Disgust, Fear, joy, Sad, Surprise, and Neutral—based on Paul Ekman's theory of universal emotions. The dataset is saved in CSV format, where each row contains the flattened pixel values of an image along with its corresponding emotion label.

The FER-2013 dataset is split into three subsets: training, public test, and private test. The training portion contains 28,709 images, which accounts for roughly 80% of the entire dataset. The remaining 20% is divided equally between public and private test sets, with 3,589 images in each. This structure enables researchers to both train models and evaluate them in a consistent manner, making it easier to compare different approaches. Typically, the public test set is used for validating during model development, while the private test set is reserved for final performance assessment.

One of the key reasons behind the popularity of FER-2013 is its ease of use and accessibility. It provides a straightforward resource for training deep learning desings, mainly Convolution Neural Networks (CNNs), to classify facial expressions. While the pictures are relatively low in resolution, the dataset is still highly effective for real-time applications and has been widely taken in both research prototypes and commercial solutions. Its compact size further makes it suitable for developing and testing lightweight models that can operate effectively on mobile or embedded devices.

Despite its usefulness, FER-2013 comes with several limitations. Since the images are grayscale and low in resolution, models trained on the dataset may struggle to capture fine details needed to recognize subtle or complex emotions. Another drawback is class imbalance— emotions such as Disgust and Fear appear less frequently, which can bias the model's performance. The dataset also lacks demographic variety and does not provide metadata like age, gender, or ethnicity, making it harder to generalize results across diverse populations. Moreover, all images are static, meaning that temporal dynamics and micro-expressions—crucial factors in real-world emotion recognition—are not represented.

Dataset	Name & No. of images in each emotion						
	Happy	Sad	Angry	Disgust	Sad	Surprise	Neutral
Fer2013	8989	6077	4953	547	6077	4002	6198

Fig 3. Sample images of each expression in Fer2013 dataset

Figure 1. Sample images of each expression in FER2013 dataset

4. KAGGLE-FER2013 Image Characteristics and Challenges:

The FER-2013 (Facial Expression Recognition 2013) dataset is a popular reference in emotion recognition research. The image gives a clear visual summary of its makeup and some sample images. The dataset is divided into seven main emotional categories: joy, Sad, Anger, Disgustingly, Fear, Surprised, and Neutral. The table shows how images are spread across these emotions, while the sample images provide a look at the facial expressions that were captured.

From the image, it's evident that the dataset is ncedimbala. The "Happy" class has the highest number of samples (8989 images), followed by "Neutral" (6198) and "Sad" (6077). In contrast, emotions like "Disgust" (547 images) and "Fear" (fewest displayed images) are significantly underrepresented. This uneven distribution can create challenges to train machine learning models, as models tend to become biased toward the majority classes, reducing classification accuracy for minority emotions like "Disgust" or "Fear." Addressing this imbalance often requires data augmentation or reweighting techniques during model training.

The sample images shown in the figure also reveal another challenge: variation in facial angle, lighting, and expression clarity. Some images are frontal and clear, such as those labeled "Happy" and "Surprise", while others—such as "Neutral" and "Angry"—appear to include more ambiguous or less expressive faces. This diversity introduces natural variability, which is good for generalization but increases the difficulty of accurate classification, especially for models not robust to pose and lighting variation.

Additionally, SAR images are the representation of backscatter intensity of objects, without color, shadow, or texture information that CNNs normally learn from in optical data. These constraints enhance the reliance on strong preprocessing, feature extraction, and data augmentation methods.

There's also a practical limitation in that FER-2013 consists of static images, not video sequences. Emotions in real life are often best interpreted dynamically— through the change in facial expressions over time. Since the FER-2013 dataset lacks temporal data, it cannot support studies on dynamic or transitional emotions, nor can it aid in detecting micro-expressions that occur in short durations.

In summary, the FER-2013 dataset, as illustrated in the image, provides a compact and valuable resource for developing and benchmarking facial emotion recognition models. Its main strengths lie in accessibility, standardized format, and diverse facial representations. However, key challenges—including class imbalance, low resolution, lack of demographic diversity, and absence of temporal context—must be acknowledged and addressed in future work to ensure fair, robust, and real- world-applicable emotion recognition systems.

5. Methodology:

Based on the referenced report titled *"Moodify: Adaptive Music Selection Using Facial Expressions"*, the methodology of the project can be described as a comprehensive, multi-layered system designed to create an emotionally intelligent music recommendation experience. At its core, Moodify integrates expression recognition with adaptive music recommendation, enabling real-time emotional feedback to guide personalized playlist generation.

The process begins with data collection, where two primary inputs are gathered: facial expressions via a webcam and historical music preferences through integration with music platforms like Spotify. Facial data is processed using a pre-trained Convolutional Neural Network (CNN) model, such as MobileNet, trained on datasets like FER-2013, to detect key emotional states— happiness, sadness, anger, calmness, and neutrality. Simultaneously, a user's music history is retrieved via APIs and enriched with metadata (tempo, energy, valence, lyrics sentiment), forming a user-specific emotional profile.

1. Data Collection:

In the Moodify system, data collection is a critical foundation that enables accurate emotion detection and effective music recommendation. The process involves gathering two primary types of data: facial expression inputs and user music listening history. Facial data is collected in real time through a webcam or mobile camera, capturing frames of the user's face during interaction.

2. Data Preprocessing:

In the Moodify system, data preprocessing is essential for recognizing emotions accurately and recommending music effectively. For facial expression analysis, preprocessing starts with capturing real-time images through a webcam. Then, face detection occurs using algorithm like Haar Cascade or Dlib. Song metadata from platforms like Spotify is cleaned and matched to emotional labels. These processioning steps ensure that both facial inputs and music data are standardized. This standardization allows for consistent, accurate, and real- time performance throughout the system.

3. Extraction Using CNNs

In the Moodify system, extraction using Convolution Neural Networks (CNNs) is nessesary for detecting facial emotions. CNNs are deep machine learning desings that focus on processing image daatasets. They automatically identify and learn patterns.

4. Model Training

The training of the Moodify's facial recognition system used Convolution Neural Networks (CNN s). These networks were trained on datasets like FER-2013, which contains thousands of labeled grayscale facial images showing various emotions. The model was optimized with architectures like MobileNet to ensure it ran efficiently. It learned to classify expressions into main emotional states such as happy, sad, angry, and neutral. Data augmentation techniques helped address class imbalance and improve how well the model performs with new data. The training focused on achieving high accuracy while keeping the system lightweight for real- time use. After training, the model was integrated into Moodify's system to provide continuous emotion detection through webcam input.

5. Techniques and Technologies:

Figure 2. Architecture diagram

Moodify employs a combination of advanced techniques to deliver emotionally adaptive music recommendations. It utilizes Convolutional Neural Networks (CNNs), specifically models like MobileNet, for real-time facial emotion recognition based on datasets such as FER-2013. For music analysis, the system leverages audio feature extraction tools like LibROSA to assess song characteristics such as tempo, energy, and valence. Sentiment analysis of lyrics is performed using natural language operation tools. The recommendation engine uses collaborative and content-based filtering enhanced by machine learning to adapt to user preferences over time. Real-time feedback loops and adaptive playlist logic ensure dynamic responsiveness to emotional shifts.

6. Proposed System:

The system proposed, Moodify, is a landmark in the area of music recommendation through combining real-time facial expression analysis with adaptive generation of music playlists It has been developed to personalize and enrich the listening experience of the user through interpretation of emotional signals from facial expressions and their conversion into appropriate musical choices.

The main function of Moodify relies on facial expression recognition (FER) technology, which is the main way to detect emotions. A webcam or smartphone camera captures the user's face. Then, frames are analyzed using a deep learning designs, usually a Convolution Neural Network (CNN) that is trained on datasets like FER-2013.

This model identifies key emotional states like happiness, sadness, anger, calmness, and neutrality by analyzing facial features and micro-expressions. The identified emotion is then classified and mapped to a predefined emotional category used by the system for music selection. This approach ensures that music is chosen based not only on historical behavior but also on the user's current, real-time mood.

This model identifies key emotional states like happiness, sadness, anger, calmness, and neutrality by analyzing facial features and micro-expressions. This approach ensures that music is chosen based not only on historical behavior but also on the user's current, real-time mood.

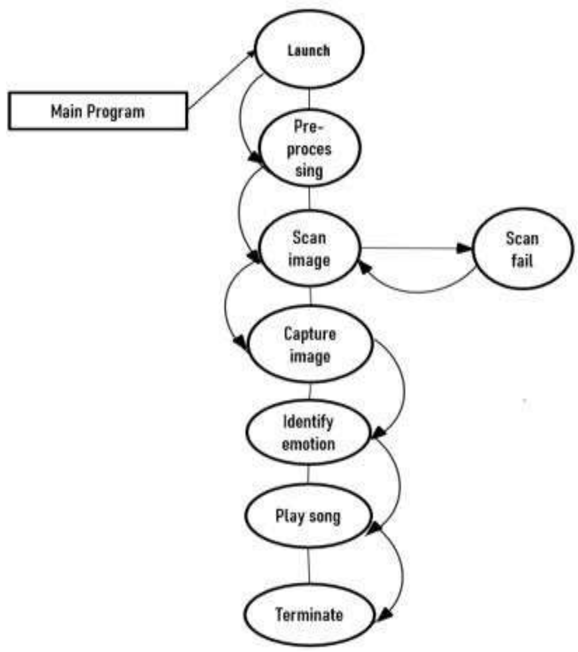

Figure 3. Flow of data

6. Conclusion:

In conclusion, Moodify represents a significant step toward emotionally intelligent technology by seamlessly integrating real-time facial emotion recognition with adaptive music recommendation. By analyzing users' facial expressions and aligning them with curated playlists, the system personalizes the listening experience while promoting emotional well-being. Its modular design, machine learning integration, and ethical data handling make it scalable, responsive, and user-friendly. Moodify not only enhances music enjoyment but also holds transformative potential in fields like mental health, education, and elder care. Ultimately, it redefines human-computer interaction by enabling technology to respond empathetically to human emotions through the universal medium of music.

REFERENCES

[1] Shantha Kumari, Nalini, Vinothkumar, Venkataprassanna, Nikilesh, Nithin Pranav wrote “Spotify Genre Recommendation Based On User Emotion Using Deep Learning” in 2022 at the Fifth Internationalized Conference on Computational Intelligency and Communication Technologies (CCICT).

[2] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar, “Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation”, This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2017 conference.

[3] Kevin Patel, Rajeev Kumar Gupta,upon Facial Expression and Song Mood,” 2021 International Conference of Artificially Intelligent and Machine Visions (AIMV)..

[4] Sonika Malik, “Music selection based on Facial Emotion Detection”, Journal of Computer Application (0975 –
8887) Volume 185 – No. 1, April 2023.

[5] Anukriti Dureha, “An Accurate Algorithm for
Generating a Music Playlist based on Facial Expressions”, Journal of Computer Application (0975 – 8887) Volume 100– No.9, August 2014.

Journal of scientific Research & Technology (IJERT) ISSN: 2278-0181 ICRTT - 2018 Conference Proceedings.

[7] Pampalk, E. and M. Gasser, “An Implementation of a Simple Playlist Generator upon Audio Similarity Measures and User Feedback” in proceeding of seventh International Conference on Music Information Retrieval, 2006, pp. 1-3.

[8] Adiyansjah, Alexander A S Gunawan, Derwin Suhartono, “Music Recommender System Based on Genre using Convolutional Recurrent Neural Networks”, Procedia Computer Science, vol. 157, 2019, pp. 99-109.

[9] Chen, HC., Chen, A.L.P. “A Music Recommendation Structure Based on Music and User Grouping”, Journal of Intelligent Information Systems,vol. 24,2005, pp.113– 132.

[10] Schedl Markus , Zamani Hamed , Chen Ching-Wei , Deldjoo Yashar , Elahi Mehdi,“Current Challenges and methods in Music selection structure Research”,International Journal of scientific Information Retrieval,vol. 7(2) ,2018, pp.95-116.

[11] Salma Alhagry, Aly Aly Fahmy, Reda A. El-Khorib,” Emotion Recognition based on EEG using LSTM Recurrent Neural Network”, (IJACSA) International author of Advance Computer Science and Applications, Vol. 8, No. 10, 2017.

[12] David Su, Rosalind Picard,Yan Liu,” AMAI: Adaptive songs for Affect Improvement”.

[13] Rugwed Nand1, Devkinandan Jagtap2, Vedant Achole3, Atharva Jadhav4, Prof. Prranjali Jadhav5,” Mood Music Recommendation Using Emotion Detection”, Journal for Research in Applied Science & Engineering Technology (IJRASET).

[14] Hafeez Kabani1, Sharik Khan2, Omar Khan3, Shabana Tadvi4,” Emotion Based Music Player”, International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.

[15] Rakhitha Wimalaweera 1,1 and Lakna Gammedda 2,” A

[16] Zhengyu ,Hongxu Ma, Cali Chen ," Generate music with customized music style (Music Generation)", Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.) .

[17] Lokesh Devathati1, Chekuri Bhargava Surendra2, Chekka Madhava Rao3, Marri Bhargava Ranga Swamy4 , Masood Khan5," Facial Expression-Driven Song Suggestion System".

[18] Spiros Ioannou1, Loic Kessous2, George Caridakis1, Kostas Karpouzis1, Vered Aharonson2 and Stefanos Kollias1," Adaptive On-Line Neural Network Retraining for Real Life Multimodal Emotion Recognition", School of Electrical and Computer Engineering, National Technical University of Athens.

[19] Steven R. Livingstone1,3, Ralf Mühlberger1, Andrew R. Brown2,3 and Andrew Loch," Controlling musical emotionality: an affective computational architecture for inluencing musical emotions".

[20] Emery Schubert," EMOTIONFACE: PROTOTYPE FACIAL EXPRESSION DISPLAY OF EMOTION IN MUSIC", Proceedings of ICAD 04-Tenth Meeting of the Conference on Auditory Display, Sydney, Australia, July 6-9, 2004.

[21] Dr. M.K. Jayanthi Kannan, Anirudh Kanwar, Harsh Chaturvedi, Aditya R Patil, Abhimaan Yadav," AI-Driven Vibebox: Adaptive Music Streaming Personalized Based on Emotion", International Journal of Advance Research, Ideas and Innovations in Technology.

[22] Prof. Vandana Navale1, Sanjana V. Phand2, Pranav K. Patil3, Dhananjay V. Rangat4, Rohan J. Sathe5,"FeelFlow: Music Recommendation Structure Based on your Facial Expression", Journal of Research in Applied Science & Engineering .

[23] Shavak Chauhan 1, Rajdeep Mangrola 2, D.Viji 3," Analysis of Intelligent movie recommender system from facial expression", Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021).

[24]ELLEN DISSANAYAKE School of Music University of Washington," If music is the food of love, what about survival and reproductive success?".

[25]Preema J. S, Rajashree, Sahana M, Savitri H, Shruthi S. J," Review on Facial Expression Based Music Player", International Journal of Engineering Research & Technology (IJERT).

[26]Pranjul Agrahari1, Ayush Singh Tanwar2, Biju Das3, Prof. Pankaj Kunekar4," Musical Therapy using Facial Expressions", International Research Journal of Engineering and Technology (IRJET).

[27]Ian R. O'Keeffe," Soundtrack Localisation: Culturally Adaptive Music Content for Computer Games", Localisation Research Centre, Centre for Next Generation University of Limerick, Co. Limerick, Ireland.

[28] Patrick E. Savagea , Psyche Louib, Bronwyn Tarrc, Adena Schachnerd, Luke Glowackie, Steven Mithenf and W. Tecumseh Fitchg," Music as a coevolved system for social bonding".

[29] Jeevan Babu Maddala1, S. Pavithra2, T. Vijay Kumar Reddy3, V. Mounika4, T. Maidhili5," Music selection Structure on Facial movements by using CNN", Journal international Research in Applied Science & Engineering .

[30] Ciprian A. Corneanu, Marc Oliu, Jeffrey F. Cohn, Sergio Escalera," Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect related Applications", IEEE Trans Pattern Anal Mach Intell. 2016 August .