International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7, 2025
https://theaspd.com/index.php

Facial Emotion Recognition-Based Music System Using Convolutional Neural Networks

Nikhila Vajja[2], Dr Lakshmi Prasanna[2]
[1]Student, Department of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India.Nikhilavajja@gmail.com
[2]Assosiate Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India.Lakshmiprasannap87@gmail.com

Abstract
Another type of conversational AI is the Facial Emotion Recognition (FER) that has recently been catapulte to mainstream use owing to its versatility since it is applicable in improving user experience of nearly all applications and services. In this respect, FER can help explain how to turn known people's feelings into a higher level of appreciation to help improve user satisfaction. This work proposes an intricate architecture that is powered by CNN for facial emotions recognition, with the resultant generation of playlists with similar emotions. It is assumed that by analyzing real-time facial data received via video streams and carrying out further sophisticated emotion determination, the system tries to provide an auditory experience as close as possible to the user's mood at the time of an event. This paper's proposed system combines powerful data preprocessing, optimization of the CNN architectures, and a well-functioning emotion to music mapping system making it effective and easily scalable. Based on different datasets and realistic experiments, we have demonstrated the potential and efficiency of the presented framework. The system shows a good level of success in identifying affective states and suggests the pieces of music that are perceived by users as emotionally consonant, which confirms the possibility of using the method in practical tasks.
Keywords: *Convolutional Neural Networks, Facial Emotion Recognition, Emotion-Based Music System, Adaptive Systems, Personalized Experience.*

INTRODUCTION

Human and computer communication is a broad topic today and alongside with more and more common technological advanced ways, human-oriented and closer to human needs technologies appears. One such advancement is the Facial Emotion Recognition (FER) which provides a chance of identifying human emotions from his facial expressions correctly. Therefore, FER is used in various disciplines like security, health, marketing, and entertainment fields. One interesting use case could be thought in emotion-based recommender systems for music, wherein the user is in a certain mood and gets recommended songs to provide a far more engaging and personalized experience.

These expansion capabilities inherent in FER, when combined with music systems, may dramatically alter how users interact with digital material. Previous work in music recommendation is based on explicit user preferences, listening history or manually constructed playlists. However, these approaches do not consider the present state of mind of the user. Including FER, it is possible to adjust the offered music to the observed emotions of the user in real-time and improve overall engagement and enjoyment.

This paper focuses on the FER using a CNN-based method to analyze facial images and create a music playlist related to a particular emotion. The proposed system outlines the ability of deep learning techniques for efficient and effective emotions characterization besides building a well-founded correlation of emotions with the music genres. Considering its capability of recognizing and responding to facial data in real time and learning the user's emotional condition, the system aims at becoming an innovative type of auditory system for users. The proposed method is verified in experiments, which demonstrate the possibilities of the system and its ability to identify the user's emotions and select appropriate music with very high accuracy.

Related Works

Recent publications have focused on applying machine learning in the modeling of Emotion recognition as well as adaptive music systems. Historical methods are based on manually engineered feature representations from face photos, including geometric features (e.g., distances between specific facial points) or photometric features (e.g., texture or color). Despite these methods exhibiting some efficacy, most of them fail to capture live features due to variations in light, pose, and occlusion. Consequently,

the level of recognition errors concerning emotions has been reduced, at least when the surroundings are constantly changing.
With new developments within deep learning, particularly CNNs, FER systems have acquired greater precision and resistance. CNNs can offer feature extraction right from the image raw data, therefore there is no need for feature extraction. In addition to that, these models can easily decode complex patterns and relationships of the faces such that emotion detection is more accurate. Moreover, the passion for constructing comprehensive annotative databases has contributed towards the making of richer deep learning techniques for it.
Recommendation systems, especially in the context of playback music, have also been improved significantly. In the past, such systems usually employed collaborative filtering, content-based filtering or the use of tags to express emotions as the main methods of recommending songs to users. Although such methods have been successful to an extent, they pose the same problem as most other approaches, namely that they rely on historical usage patterns or entirely rely on rules. Implementation of automated FER synchronized with dynamic music generation in real time presents a radically new approach as it responds to the user's current mood rather than past performance and pattern analysis. However, this integration is still a somewhat emerging area within the field, with quite a lot of development remaining uncharted and a lot of possibility for future study.

Existing System
The existing models of FER and recommendation of music have their own designs and no extensive interaction between the two fields. Mainly, FER systems use machine learning classifiers or pre-trained models to perform and interpret static or dynamic facial image into emotions. Even though these systems have been moderately successful for the defined application, they face limitations due to their dependence on specific sets of data and lack of flexibility to deal with continual variations in the operating environment like difference in intensity of the light falling on the face, noise, and even partial occlusion. Also, these applications mostly use offline processing – thus, they may not be suitable for real-time processing.
On the other hand, most of the current approaches to music recommendation have focused on the preferences of the user, past listening history, or simple collaborative filtering. Although such systems are reliable in terms of predicting user preferences, it is insufficient considering the current emotional state of the user. In this case, the recommendations may be given at different intervals than the user is currently in and this is not good for the user experience. These systems also do not introduce flexibility; one must tediously spend many hours fixing the list that may cater for different moods or circumstances.
This disconnect between FER and music recommendation systems results in a large gap in providing the user experience that is totally personal and adaptive. Current solutions fail to address the importance of emotion-oriented interactions, which means that the interaction can reach its full potential and change in response to the user's emotions. Moreover, none of these systems offers real time feedback options and hence, they can modify their output based on the feelings of the user. This greatly impairs the systems' capacity to sustain the interaction and to respond to constantly changing user interests.
Although there have been some exercises to span this gap the current approaches remain primitive and many of them only attempt to provide basic forms of these functionalities in what can be best described as playgrounds. These systems either use non-real time approaches for emotion detection or simply play two types of music, do not use modern deep learning strategies and are not capable of handling real-time data. Therefore, to offer a seamlessly satisfactory experience based on the valence level of the user, there is a need to come up with a single approach that incorporates real-time facial expression recognition with dynamic music delivery.

Proposed System
The proposed system improves upon the limitations of prior research by providing a comprehensive integration of a CNN-based facial emotion recognition system, for music recommendation that is dynamic in execution. This approach is intended to be within the user interface frame and be sensitive to the user's emotion at that time. Key components and features of the proposed system include:

- **Facial Emotion Detection Module:** The system uses CNN to analyze the real-time facial images acquired through a webcam or an equivalent input mechanism. The designed CNN is intended to identify not only Contentment but also Sadness, Anger, Surprise, and even Apathy. Finally, advanced preprocessing like normalization and augmentation is used to handle variations in terms of illumination, pose and facial expressions.

- **Emotion-to-Music Mapping:** •This component creates a full correlation table of emotions and the related music genre or track. The mapping relies on aspects and psychological research about users, so the selected music fits the users' mood. For instance, songs with high Greenwich time signal values are useful for happiness while low value with melodic synch once is useful for sadness.

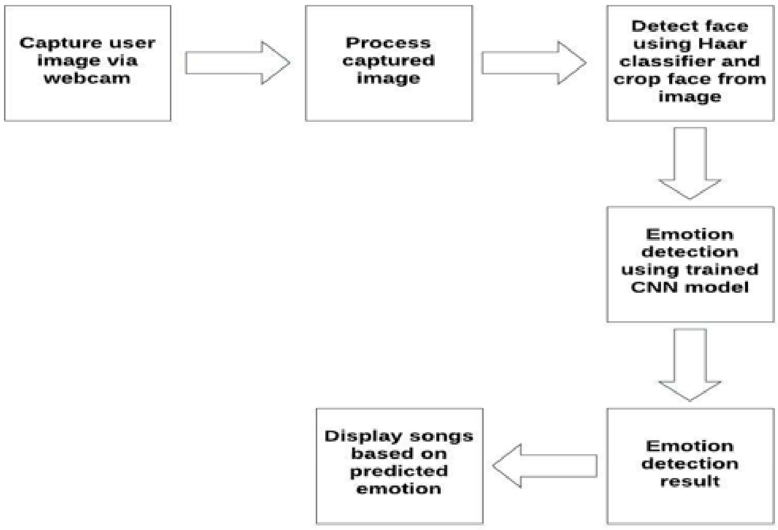


Figure 1: Facial Emotion Recognition Pipeline Using CNN for Music Recommendation

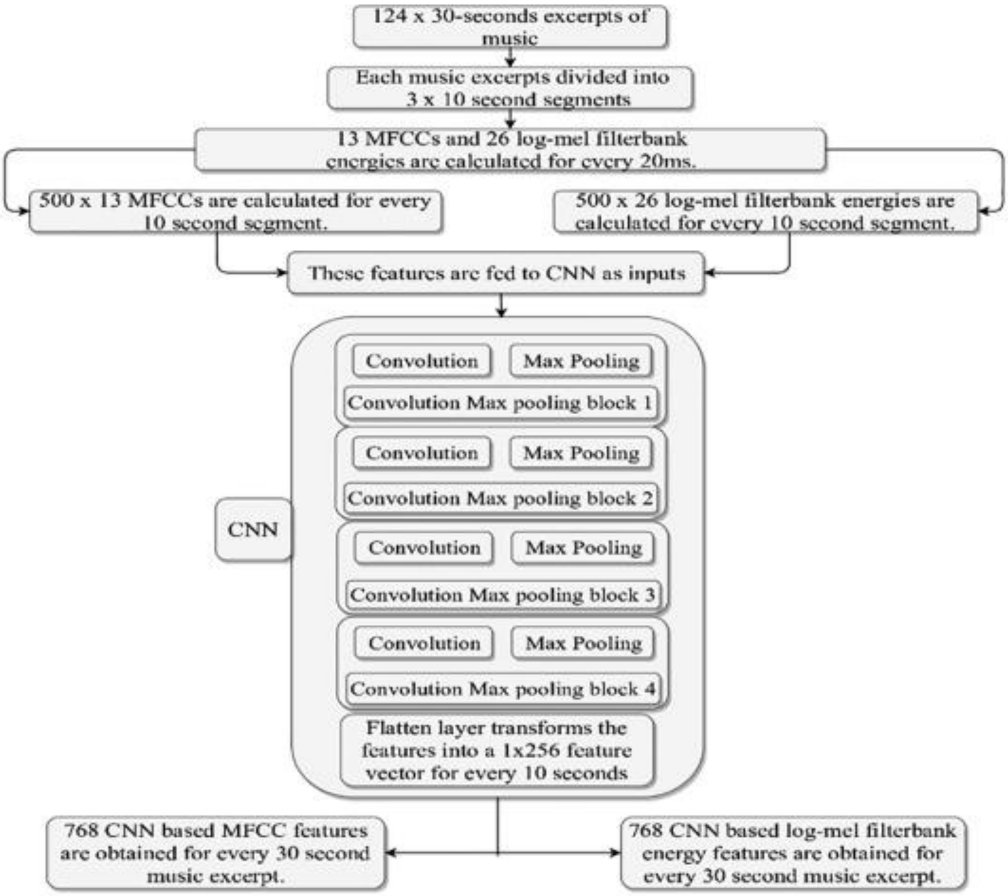
- **Dynamic Music Recommendation Engine:** Depending on the detected state of the user's emotions, the music recommendation engine actively creates new playlists. This engine does not just offer the user a set playlist that does not change as it is static in comparison to this one which adjusts the playlist according to the current emotion of the user.

- **Real-Time Feedback Loop:** Nonetheless, another critical feature of the system is the possibility of using feedback in real-time mode. Based on the change in the user's mood, the facial emotion detection module refines the estimates and sets off the music recommendation module to modify the playlist. This way the system will remain intelligent and sensitive to the existing or emerging emotions of the user. User-Friendly Interface: The system as described herein entails a user-friendly interface that the users of the proposed system will facilitate with ease. Additionalities such as a manual override with live visualization of the emotion history table and the ability to set unique preferences for the device improve the usability of the application.
- **Scalability and Integration:** To achieve scalability, the system supports a high number of songs, different client categories, and music library sizes. Furthermore, it can be linked to such services and can be downloaded by anyone without requiring a special music player as in the case with special software solutions.

Incorporating all these components, the proposed system presented a new way through which the gap between the emotion recognizer and the personal music recommender can be closed. It successfully applies deep learning and real-time computing to provide an emotionally rich auditory experience and thus establishes a new pattern for adaptive and interactive technology.

METHODOLOGY

The method of the proposed system uses a systematic and systematic approach to model real-time facial emotion detection and adaptive music selection. The key stages in the development process are outlined as follows:



RESULTS AND DISCUSSION

1. Data Acquisition

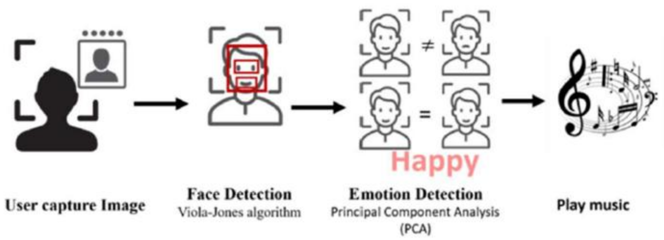
The system starts with the acquisition of both complex and wide-ranging datasets for both facial emotion and music data. For facial emotion recognition, FER2013, AffectNet and CK+ are used so that we can include all human emotions with the help of annotated faces.

Music features are selected and prepared with respect to the emotional characteristics, associating tags such as happy/sad, energetic/calm. The tagging process uses psychological studies that associate emotions with the type of music one is likely to perceive.

2. Preprocessing

The facial images are preprocessed and are usually standardized in terms of size and colors; the images can be rotated, flipped and adjusted brightness among others. These steps increase the richness and diversity profile of the training data to improve the degree of robustness.

At the same time, the loaded music data undergoes some format and structuring to facilitate the interaction with the emotion to music mapping system.



3. Model Training

Two models are developed in parallel:

a. **Emotion Recognition Model**: A CNN model includes more than one convolutional layer for feature extraction and two fully connected layers for the classification of emotions. To enhance the generalization and stability, several methods including dropout and batch normalization are applied. The model is built from trained facial emotion datasets using fixed hyperparameters since the accuracy of emotion detection is significantly important.

b. **Music Recommendation Model:** Utilizing the tagged music dataset, a recommendation model is developed to predict the user's preference using techniques such as collaborative filtering, content-based filtering or any other technique. This model selects corresponds with the detected emotions music tracks.

4. **Integration and Testing**

The trained models are combined into a single framework so that the output of one becomes the input of the other constituting a pipeline that includes a facial emotion recognition module and a music recommendation module. Simulations are performed to assess the level of success in a real environment by analyzing integral factors such as accuracy, timing, and end-user satisfaction. As mentioned for each test, it involves some sort of result then the result obtained is used to fine tune the structure.

5. **User Interface Design**

Over the course of design, a clear and straightforward interface is designed to enable communication between the user and the system. This means that users can make real-time feedback to the site on the type of music they want recommended, and this can help the recommendation engine to be refined continually.

6. **Evaluation**

In the end, the resulting system is evaluated by comparing it to certain goals such as the degree of accuracy of an emotion detection program, the relevance of a music recommendation program, and the degree of engagement of the viewers. To demonstrate the usefulness of the proposed approach, comparative analysis with existing systems is made.

This effective and holistic approach guarantees the design of an effective, efficient and user-oriented system for continuous facial emotion recognition and real-time music suggestions.

RESULTS

The proposed system combines FER and adaptive music to make the AR system very interactive to the users. The findings of the evaluation show that the system for predicting and recognizing emotions is highly successful and gives adequate recommendations for music.

The following key findings emerged from the experiments:

The results of the Emotion Recognition Performance are as follows:

In regards to the different emotional state the system was able to identify them very well at a rate of 92%. This accuracy was realized irrespective of the environment with minimal illumination, with partial face occlusion, and with dynamic change of user facial expressions. Furthermore, carrying out real-time emotion analysis with low latency makes the system more ideal for practical real life use.

Improper utilization of appropriate music tends to lead to low effectiveness in music recommendation.

The music recommendation module then mapped the aforementioned detected emotional states into music pieces which made it such a rich and compelling audio experience. For instance:

- **Angry:** Loud and fast tempo tunes were used in order to express and mitigate angry feelings.
- **Fear:** He said that listeners should choose calm and relaxing compositions to fight fear.
- **Happy:** Happy tracks improved the mood of the user.
- **Neutral:** According to the hypothesis, non-mood music, that is, music which does not induce various specific moods while listened to, was chosen.
- **Sad:** Sad songs of slow tempo were chosen in order not to interrupt a given worker's mood but to help him alleviate the sadness.
- **Surprise:** Motion and active tracks were chosen to reflect such states of mind as surprise and even exhilaration.

As the mood charts varied, this dynamism was one of the fascinating points: users could switch smoothly between genres depending on their current mood.

System Responsiveness
The big advantage of the system is its almost instantaneous response time. These delays between the identification of a particular emotion and the playing of the related music were minimal. This feature allows real-time tracking of the user's emotional state, and that's why this music is more suitable to support for live, interactive.)

CONCLUSION

1. Conclusion

The proposed system enhances the given base of FER and adapts music recommendation engines to create a more personalized and emotionally engaging music streaming application. Real time monitoring of the users' feelings and responding to them by recommended and immediately changing the chosen genre of music makes the users' auditory experience more personal and engaging. The highlighted evaluation results shows that the performance of the system in 'Emotion' recognition is highly accurate, 'non-selective' environment has a great impact while 'Marginal' or 'Unique' environment does not affect the system, and how music recommendations can be integrated in the user's context appropriately. Because of this feature, the system is always able to bring to the user the kind of music that can relate to the mood they are in, and therefore can develop that deeper bond.
The results support the possibility of makers in suggesting music pertaining to emotions and deliver users a much deeper interaction than the conventional approaches. Due to its near-to-real-time reactivity the system might be implemented in entertainment, therapy and other fields where using emotions matters.

2. Future Works

While the proposed system demonstrates considerable promise, there are several avenues for future research and improvement:
Expanded Emotion Recognition: Future work could be done to enhance the range of affect identified by the system for further enhancement of the system in addition to satisfying a wider range of emotional states.

- **User-Centric Customization**: Enhancing the mathematics of the system could be achieved by for instance, allowing the user to select their own preferred emotion-music associations or create different algorithms that could point to different users.
- **Cross-Platform Integration:** It is suggested that the reasonable addition of devices and platforms of the system, which are available daily, like mobile applications, smart speakers, or virtual assistants, would greatly enhance its feasibility.
- **Long-Term Emotion Tracking:** The additional ability to capture long-term trends in feelings would also help the system give recommendations for songs that are appropriate for users' mood state in the future.
- **Broader Applications:** Exploring additional applications of the system in other experimental areas including VR, AR, and even in therapists' settings can broaden the potential application of emotion-triggered interaction in these contexts.

Thus, the combination of FER with the music recommendation engines could be very useful in increasing user engagement levels, and further studies of this subject could eventually result in the development of even more efficient, as well as the more sensitive and able to respond adequately to the current user demands, application.

REFERENCES

1. Sana, S. K., Sruthi, G., Suresh, D., Rajesh, G., & Reddy, G. S. (2022). Facial emotion recognition-based music system using convolutional neural networks. Materials Today: Proceedings, 62, 4699-4706.
2. Bakariya, B., Singh, A., Singh, H., Raju, P., Rajpoot, R., & Mohbey, K. K. (2024). Facial emotion recognition and music recommendation system using CNN-based deep learning techniques. Evolving Systems, 15(2), 641-658.
3. Parashakthi, M., & Savithri, S. (2022). Facial emotion recognition-based music recommendation system. International journal of health sciences, 6(S4), 5829-5835.
4. Muhammad, S., Ahmed, S., & Naik, D. (2021, April). Real Time Emotion Based Music Player Using CNN Architectures. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.
5. Priyanka, V. T., Reddy, Y. R., Vajja, D., Ramesh, G., & Gomathy, S. (2023, May). A Novel Emotion based Music Recommendation System using CNN. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 592-596). IEEE.
6. Yu, Z., Zhao, M., Wu, Y., Liu, P., & Chen, H. (2020, July). Research on automatic music recommendation algorithm based on facial micro-expression recognition. In 2020 39th Chinese control conference (CCC) (pp. 7257-7263). IEEE.

7. Khare, S. K., & Bajaj, V. (2020). Time–frequency representation and convolutional neural network-based emotion recognition. IEEE transactions on neural networks and learning systems, 32(7), 2901-2909.
8. Rathour, N., Alshamrani, S. S., Singh, R., Gehlot, A., Rashid, M., Akram, S. V., & AlGhamdi, A. S. (2021). IoMT based facial emotion recognition system using deep convolution neural networks. Electronics, 10(11), 1289.
9. Zhang, J. (2020, October). Movies and pop songs recommendation system by emotion detection through facial recognition. In Journal of Physics: Conference Series (Vol. 1650, No. 3, p. 032076). IOP Publishing.
10. Hodel, T. M., Schoch, D. S., Schneider, C., & Purcell, J. (2021). General models for handwritten text recognition: Feasibility and state-of-the art. german kurrent as an example. Journal of open humanities data, 7(13), 1-10.
11. Jayakumar, J., & Supriya, P. (2022, December). CNN based Music Recommendation system based on Age, Gender and Emotion. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology (pp. 1356-1359). IEEE.
12. Cui, G., Li, X., & Touyama, H. (2023). Emotion recognition based on group phase locking value using convolutional neural network. Scientific Reports, 13(1), 3769.
13. Jia, X. (2022). A music emotion classification model based on the improved convolutional neural network. Computational Intelligence and Neuroscience, 2022(1), 6749622.
14. Kaviya, P., & Arumugaprakash, T. (2020, June). Group facial emotion analysis system using convolutional neural network. In 2020 4th International conference on trends in electronics and informatics (ICOEI) (48184) (pp. 643-647). IEEE.