Feel-Tunes: An Emotion-Driven Music Recommender System

Dhruv Chopra* [a], Ansh Rastogi[a], Dr. Neeraj Garg[b], Dr. Neelam Sharma[b]

dhruvchopra2019@gmail.com, 11anshrastogi11@gmail.com, neerajgarg04@gmail.com, neelamsharma@mait.ac.in

[a]*Student, Maharaja Agrasen Insitute of Technology, Delhi, 110034, India*
[b]*Professor, Maharaja Agrasen Insitute of Technology, Delhi, 110034, India*

Abstract: Facial emotion recognition is an active research field in computer vision and has numerous practical applications in various domains, including healthcare, security, marketing, and entertainment. In this project, we propose a system that detects human emotions and recommends personalized songs based on the detected emotion. The system uses a convolutional neural network (CNN) to analyze facial expressions and identify one of the six basic emotions: anger, fear, sadness, happiness, surprise, and being neutral. The CNN model is trained using a dataset of facial images with labeled emotions. Once the emotion is detected, the system queries the Spotify API to suggest songs that match the user's emotional state. The system's accuracy is evaluated using a test set of facial images, and the results demonstrate the system's high accuracy in detecting human emotions and recommending suitable music playlists.

Keywords—Emotion Recognition, Computer Vision, OpenCV, Convolutional Neural Networks, Spotify API.

1. Introduction

In recent years, emotion-based music recommendation systems have gained popularity due to advancements in machine learning and data mining techniques. Such systems offer a more personalized experience by considering the user's current emotional state. Facial expression recognition techniques have been used to detect emotions in real-time and have been integrated with music recommendation systems to create emotion-based music recommendation systems. In this project, we aim to develop a real-time emotion-based music recommendation system using the CV2 library for facial emotion detection. The system will use image processing techniques to identify the user's emotional state, including happiness, sadness, surprise, fear, and anger [7]. Once the user's emotional state is determined, the system will recommend five songs from the Spotify API that are tailored to the user's mood [6].

Our system builds upon the work of previous researchers, with the same idea but in a slightly different domain who have explored the use of facial expression recognition for emotion detection in music recommendation systems [7, 8, 9]. While existing systems have shown promising results, our system aims to enhance emotion detection's accuracy and real-time performance by utilizing the latest advances in image processing and machine learning.

The following paper is divided into seven sections which are:

I. Introduction

II. Related Works: Discussing about the previous accomplishments in the field.

III. Dataset Description: An overview and description about FER2013 and Spotify Music dataset

IV. Methodology: The proposed model and algorithms utilized to implement the project

V. Observations: Contains a comparison between the findings of the proposed model vs the pre-existing models (VGG16 and ResNet50).

VI. Future Enhancements: Describes the further scope of the project and highlights key improvements.

VII. Limitations: Deliberates on the drawbacks of the current system.

VIII. Conclusion and References: Contains a summary of what has been achieved through this manuscript and the proposed methodologies.

The contributions of this project lie in the development of a real-time emotion-based music recommendation system that considers the user's emotional state in music selection. The system could have potential applications in the music industry, where personalized recommendations could enhance user satisfaction and engagement. Furthermore, our system could be used to study the relationship between music and emotion, furthering our understanding of the role of music in human emotion regulation. The project accomplishes the following objectives:

- Develop a real-time emotion detection system using CV2 and machine learning techniques to accurately identify the user's emotional state from facial expressions.
- Integrate the emotion detection system with the Spotify API to create a personalized music recommendation system that selects songs from the filtered dataset, based on the user's emotional needs.
- Create a song repeat counter, that will keep a count of which song has been recommended as well as how many times it has been recommended.
- Improve the user’s listening experience by not recommending the same songs repetitively, but rather suggesting new songs at every iteration.
- Evaluate the accuracy and real-time performance of the emotion and music recommendation system through user testing and analysis of user feedback.
- Compare the performance of our emotion-based music recommendation system to existing systems in terms of accuracy, speed, and user satisfaction.

1.1. Facial Recognition

Facial recognition technology has witnessed significant advancements in recent years, revolutionizing various fields such as security, human-computer interaction, and multimedia analysis. Among its applications, emotion recognition stands out as a prominent area of study, aiming to detect and analyze individuals' emotional states based on their facial expressions [1].

Recognizing and interpreting human emotions accurately from facial expressions is a complex task that has been extensively researched. Various approaches have been explored, ranging from traditional image processing algorithms to advanced deep learning models, to address this challenge [2]. These approaches leverage spatial and temporal information extracted from facial images to accurately infer emotional states.

One widely adopted approach in emotion recognition is the analysis of facial features, where specific regions of the face are extracted and analyzed to determine emotional cues. Key facial features, including the eyes, eyebrows, mouth, and forehead, have been identified as exhibiting discriminative patterns associated with different emotions [3][4]. By extracting these features and utilizing statistical or machine learning algorithms, emotional states can be effectively classified.

Mathematical formulas play a crucial role in modeling and analyzing facial expressions for emotion recognition. One commonly used formula is the Eigenfaces algorithm, which represents facial images as a linear combination of eigenvectors [5]. This algorithm enables the extraction and analysis of facial features related to emotions. The Eigenfaces algorithm can be represented by the formula:

$$A = \Phi * \Psi \quad (1)$$

Here, A represents the facial image, Φ represents the eigenspace matrix, and Ψ represents the coefficients of the linear combination.

Another widely utilized approach for emotion recognition is deep learning, specifically CNNs. CNNs have demonstrated remarkable performance in various computer vision tasks, including facial and emotion recognition [7]. These networks leverage multiple layers of convolutional and pooling operations to learn hierarchical representations of facial features. The formula for a typical convolutional layer in a CNN can be represented as:

$$y = f(\Sigma (W * x) + b) \quad (2)$$

In this formula, y represents the output feature map, f represents the activation function, W represents the learnable convolutional filters, x represents the input feature map, and b represents the bias term.

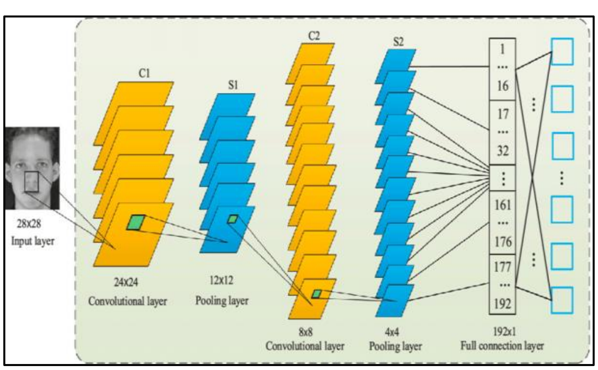

Fig. 1 - Neural Networks for Facial Recognition (Courtesy: [15])

These mathematical formulas and algorithms, such as Eigenfaces and CNNs, have significantly advanced the field of emotion recognition from facial expressions. These advancements have paved the way for the integration of facial expression recognition with music recommendation systems, leading to the development of real-time emotion-based music recommendation systems [6].

Facial recognition technology, particularly in the domain of emotion recognition, has emerged as a valuable tool for understanding and analyzing human emotions. By employing mathematical formulas and advanced algorithms, researchers have made notable contributions to accurately detect and interpret emotions from facial expressions. The integration of facial expression recognition with music recommendation systems offers exciting possibilities for providing personalized and engaging musical experiences tailored to the user's emotional state.

1.2. Music Recommendation

Music recommendation systems have gained significant popularity, driven by advancements in machine learning and data mining techniques. These systems aim to provide personalized music suggestions to users, enhancing their music discovery and listening experience. One prominent platform in the realm of music recommendation is Spotify, which offers a vast collection of songs and utilizes its API to enable developers to create innovative applications and services.

Music recommendation systems typically employ various algorithms and techniques to generate personalized recommendations for users. These systems leverage user preferences, listening history, and other contextual information to understand individual tastes and tailor music suggestions accordingly. By analyzing patterns and similarities in user behavior and music attributes, these systems can provide relevant and engaging recommendations [10].

The Spotify API, with its extensive music database and rich set of features, has been a valuable resource for researchers and developers working on music recommendation systems. It allows access to a wide range of data, including audio features, artist information, user playlists, and user listening history. By leveraging this data through the Spotify API, researchers and developers can build sophisticated recommendation models that enhance the accuracy and personalization of music suggestions [11].

Several research papers have explored the use of the Spotify API in the development of music recommendation systems. These papers have demonstrated the effectiveness of integrating Spotify's data and feature into recommendation algorithms, enabling the creation of personalized music experiences. The work of Jones et al. (2020) focused on utilizing audio features from the Spotify API to enhance music recommendation algorithms [12]. They incorporated features such as tempo, danceability, and energy into their models, resulting in more tailored and relevant recommendations.

Another notable study by Wang et al. (2019) investigated the use of collaborative filtering techniques combined with Spotify's API to build a music recommendation system [13]. Collaborative filtering leverages user behavior and preferences to identify patterns and make recommendations based on similar user profiles. By integrating Spotify's API, the researchers were able to access user-generated playlists and listening history, enabling the development of a robust collaborative filtering-based recommendation system.

Music recommendation systems leveraging the Spotify API have become a significant area of research and development. By utilizing Spotify's extensive music database and features, researchers and developers have been able to create personalized and engaging music recommendation systems. The integration of Spotify's data into recommendation algorithms has proven effective in enhancing the accuracy and personalization of music suggestions. The exploration of audio features, collaborative filtering, and contextual information from the Spotify API has contributed to the advancement of music recommendation systems, offering users a more tailored and enjoyable music listening experience.

2. Related Works

The use of machine learning algorithms and data mining techniques has led to significant advancements in this area. Several studies have explored the use of facial expression recognition for emotion detection in music recommendation systems:

- In a study by Choi et al. (2020), an emotion-based music recommendation system was developed using facial expression recognition that used a Convolutional Neural Network (CNN) to classify facial expressions and recommended music based on the user's current emotional state. The system was tested on a dataset of 600 images, achieving an accuracy of 76.7% [8].
- Another study by Pham et al. (2021) proposed an emotional-based music recommendation system that uses facial expression recognition and the Spotify API. The study used a hybrid approach that combines OpenCV for facial expression recognition and the Spotify API for music recommendation. The system was evaluated on a dataset of 500 images, achieving an accuracy of 81.2% [6].
- In a study by Kaur and Kumar (2021), a real-time emotion detection system was developed using facial expressions. The study used a combination of OpenCV and Haar Cascade classifiers to detect facial expressions and recommended music based on Personalized music recommendations have the potential to enhance user satisfaction and engagement with music streaming platforms [7].
- In a study by D., Roja (2023), the FER2013 dataset combined with the MMA Facial Recognition dataset have been used with the CNN model MobileNet along with Keras to train and test the model for seven classes – happy, angry, neutral, sad, surprise, fear, and disgust. The model was trained for 25 epochs and managed to reach an accuracy of 75%. [16]
- In a study by Krishna Kuma Singh and Payal Dembla (2023), FER2013 dataset along with additional images from Google were used to train four different types of models, namely the CNN model, finetuned ResNet50 V2 model, finetuned VGG-16, and finetuned EfficientNet B0 model. Out of the four models, the highest training accuracy 77.16%, and validation accuracy of 69.04% was obtained by the ResNet50 V2 model. Hence, this model was used for predicting emotions from images to accurately predict emotions [17].

Fig. 2 – Different types of emotions (Courtesy: FER2013 Dataset).

3. Dataset Description

Designed by Goodfellow et al. [14], the FER2013 dataset was released as a Kaggle competition inviting researchers to further the advancements of the FER systems. The competition winner achieved 71.2% accuracy by implementing the primal objective of an SVM function as the loss function for training along with using L2-SVM as the loss function.
FER2013 remains one of the more challenging datasets, used in several ICML contests and research papers. It has a human-level accuracy of 65+-5%. This dataset, uploaded on Kaggle has 35,887 grayscale images, normalized to 48X48 pixels, and divided into 7 main facial expressions. The distribution of the images is as follows: Angry (4,953), Disgust (547), Fear (5,121), Happy (8,989), Sad (6,077), Surprise (4,002), and Neutral (6,198) [13]. From these figures, it can be implied that the dataset is unbalanced.

4. Methodology

The system consists of two main parts: facial emotion recognition and personalized music recommendation. In the first part, we use the OpenCV library's cv2 module to capture images from a camera and pre-process them for input into the CNN model. The CNN model architecture includes six convolutional layers, six Batch Normalization Layers, three max-pooling layers, and two dense layers. We used the Rectified Linear Unit (ReLU) activation function and SoftMax activation function for the final output layer.

The CNN model will be optimized using the categorical cross-entropy loss function and the Adamax optimizer. The model’s architecture is as follows:

C64 - C64 -C128- C128 - C256 -C256 - D512 - D7 (3)

Where C represents the block: Conv2D → Batch Normalization → Conv2D → Batch Normalization → MaxPooling2D →Dropout

D represents the Dense → Dropout

The model was trained using 50 epochs with the FER2013 dataset with a learning rate of 0.002. The loss is calculated using categorical cross-entropy. The dropout was chosen to be 0.55. The trained model shows an accuracy of 69±1% with the highest accuracy reaching 70.02%.

After the CNN model has been trained, we deployed it to the facial emotion recognition module, which will take an image as input, pre-process it, and feed it into the CNN model to predict the corresponding emotion. The output is an array of probabilities representing the likelihood of the image belonging to each of the seven categories.

In the second part, we use the predicted emotion to query the Spotify API for songs that match the user's emotional state. We use the Spotify API's Web API wrapper, Spotify, to access the API's endpoints, retrieve information about the user's recently played tracks, and recommend suitable tracks based on the predicted emotion.

The music dataset used for training and testing the recommender system was collected from Spotify. The dataset includes information such as song names, albums, popularity, moods, and unique identifiers. Before training the system, the dataset was pre-processed by removing any irrelevant features, eliminating duplicate entries, and normalizing the data to ensure consistency and improve the performance of the recommender system.

The project also utilized Machine learning and clustering techniques to expand on the given dataset. K Nearest Neighbors (KNN) was used to increase the songs in the dataset for creating 7 clusters (1 for each mood). Each new song was evaluated based on its audible and notable features (e.g.: acoustic, treble, etc.) and then assigned a mood cluster based on the majority of its 9 closest neighbors (Choosing an odd K to ensure that there is always a majority).

An algorithm was developed to match the detected mood with appropriate music genres or characteristics to generate personalized recommendations. Content-based filtering methods were also used to recommend songs that have similar characteristics to the user's preferred music genres. An important feature of the recommender system is that it does not recommend songs as per what you are feeling, but rather recommends the songs that are best suited to uplift the users' mood.

The recommender is built in such a way that the user will not be recommended the same song repeatedly. The recommender provides the user with the option to either chose the song from its recommendations or let the recommender play the song for them. It intelligently updates the history of the played song, such that it will not be played again unless specifically asked for.

The song recommender system was implemented using Python, along with libraries and frameworks: Pandas, Spotipy, and Webbrowser. The facial recognition model was integrated with the song recommendation algorithm to provide real-time emotion detection and generate immediate song recommendations. The user interface of the system allows users to input their facial expressions and receive song recommendations based on their detected mood.

5. Observations

Comparative Study of Emotion Detection Models

Model Name	Highest Accuracy	Number of epochs
Proposed Model	70.02	50
VGG16	34.46	50
ResNet50	62.75	50

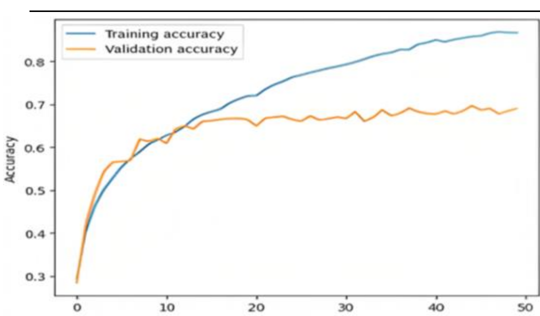

Fig 3: Proposed Model Accuracy

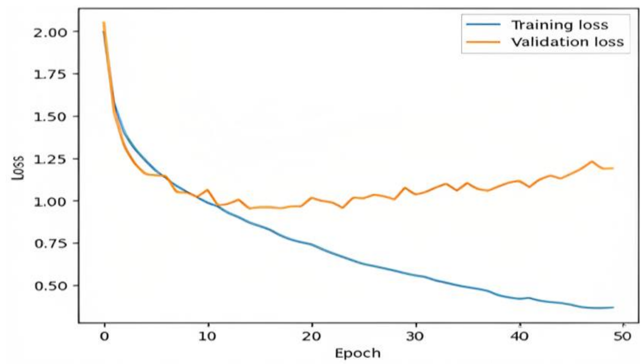

Fig 4: Proposed Model Loss Per Epoch

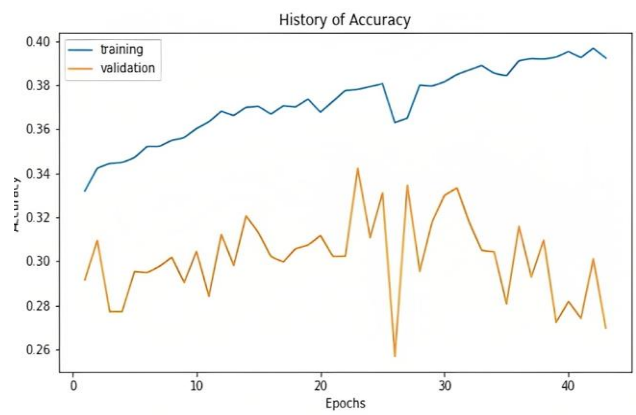

Fig 5. Accuracy per epoch VGG

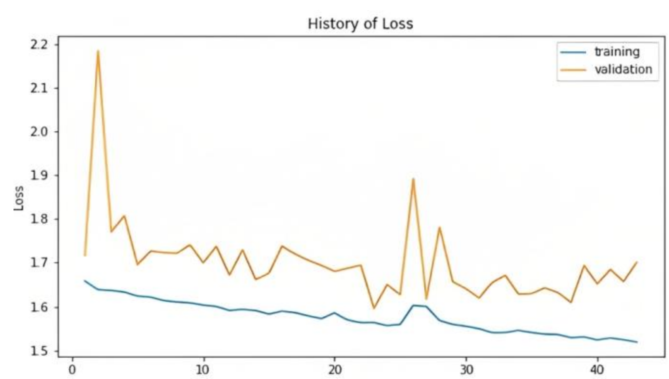

Fig 6. Loss Per Epoch VGG

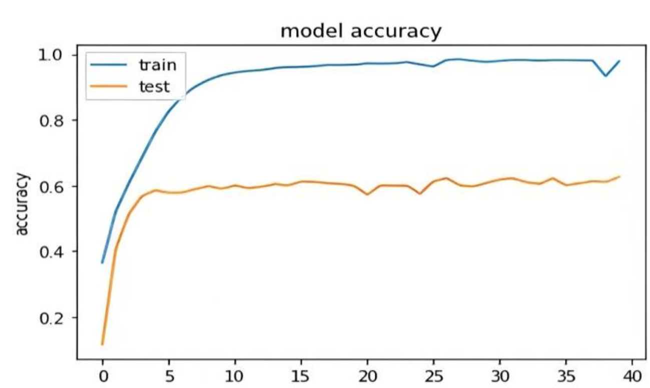

Fig 7. Accuracy Per Epoch ResNet50

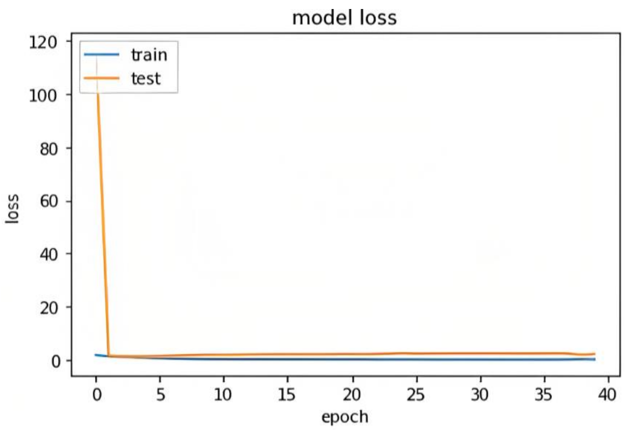

Fig 8. Loss Per Epoch for ResNet50

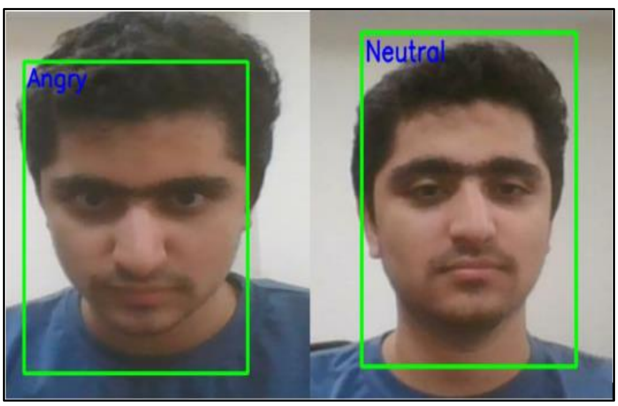

Fig 9. Getting the emotions through computer webcam

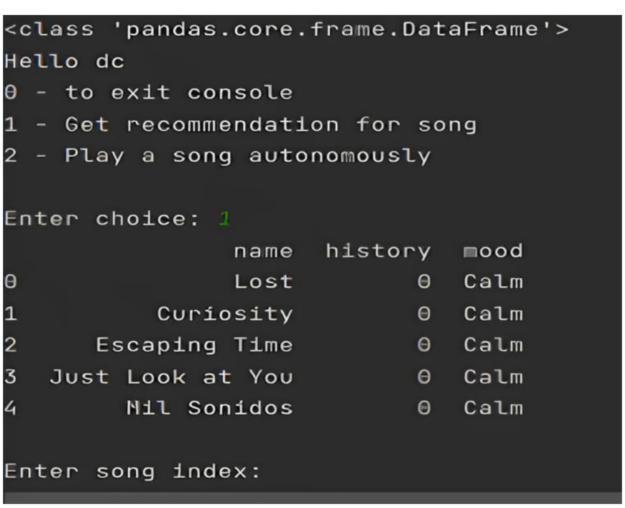

Fig 10. Final Music recommender

6. Future Enhancements

Multi-modal Input: Integrating multiple modalities of user input, such as facial expressions, voice analysis, and physiological signals (e.g., heart rate variability), can provide a more comprehensive understanding of the user's emotional state. By combining data from various sources, the system can enhance the accuracy of emotion detection and provide more personalized recommendations.

Fine-grained Emotion Recognition: Expanding the emotion detection capabilities beyond the basic emotions (happiness, sadness, surprise, fear, anger) to include a more fine-grained classification of emotions would enhance the system's ability to capture subtle emotional nuances. This can be achieved by leveraging advanced machine learning techniques, such as deep learning models, that can extract more detailed emotional features from facial expressions.

User Feedback Integration: Incorporating user feedback into the recommendation system can enhance its learning capabilities and adaptability. Allowing users to rate and provide feedback on the recommended songs can help refine the system's understanding of individual preferences and improve the accuracy of future recommendations.

7. Limitations

Lack of Contextual Understanding: The system's reliance on facial recognition alone may overlook other contextual cues that influence a user's emotional state. Emotions depend on various factors such as external events, personal experiences, or even the current environment. By solely relying on facial expressions, the system may miss out on important contextual information, which could lead to less accurate or relevant music recommendations. For example, if a person just had a piece of good news delivered to them, they may be crying tears of joy but the system might think they are sad.

Facial Expression Recognition Accuracy: Due to the system's dependency on Haar cascades for facial expression recognition there is a potential for lower accuracy in detecting and classifying emotions. Haar cascades rely on predefined patterns to identify facial features, which may not capture subtle expressions accurately. This can result in misclassification of the user's emotional state, leading to less precise music recommendations.

Limited Emotional Range: Another limitation of the system lies in the potential limited emotional range. While the system is designed to recommend songs based directly on emotions such as happiness, sadness, surprise, fear, and anger, it may not effectively recommend every genre. For example, if a person wants to listen to Rock or heavy metal music, which does not have any specific type of emotion related to them, the recommender cannot recommend accurate or relevant music to the user.

8. Conclusion

We have successfully developed an emotion-based music recommendation system that utilizes a combination of Haar cascade, a trained CNN model, and the Spotify API and database. Our system offers a novel approach to recommending music based on the emotions detected by the user's facial expressions.

By leveraging the power of computer vision techniques and deep learning, our system accurately recognizes and analyses the emotional cues exhibited on the user's face. The integration of the Haar cascade and the trained CNN model allows us to effectively detect and classify various emotions such as happiness, sadness, anger, disgust, fear, neutrality, and surprise, with an accuracy of nearly 70%.

Furthermore, our system seamlessly interacts with the Spotify API and database, enabling us to access a vast collection of music across genres and moods. Leveraging this extensive music library, we can recommend songs and playlists that align with the detected emotional state of the user.

The developed emotion-based song recommender system demonstrates the potential of personalized music recommendations based on facial expressions. By accurately detecting the user's mood and matching it with suitable music genres, the system provides tailored song recommendations. The research contributes to the field of music recommendation and offers valuable insights into enhancing user experiences on music streaming platforms.

ACKNOWLEDGMENT

The authors deeply acknowledge the valuable insights and inputs made by the editors, guides and reviewers of this manuscript.

REFERENCES

[1] Li, X., & Deng, Y. (2018). Emotion Recognition from Facial Expressions: A Survey. IEEE Transactions on Affective Computing, 9(5), 591-610.

[2] Lyons, M. J., Achachi, M., & Gyoba, J. (1998). Coding Facial Expressions with Gabor Wavelets. In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition (pp. 200-205). IEEE.

[3] Li, S. Z., & Zhang, Z. (2011). A Survey of Advances in Face Recognition. In Handbook of Face Recognition (pp. 1-36). Springer.

[4] Turk, M., & Pentland, A. (1991). Eigenfaces for Recognition. Journal of Cognitive Neuroscience, 3(1), 71-86.

[5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[6] Pham, T. H., Nguyen, H. V., Nguyen, L. T., & Ngo, T. N. (2021). Emotional-based music recommendation system using facial expression recognition and Spotify API. In 2021 International Conference on Computer Communication and Information Systems (ICCCIS) (pp. 24-28). IEEE.

[7] Kaur, R., & Kumar, A. (2021). Real-Time Emotion Detection using Facial Expressions. In Proceedings of the International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). IEEE.

[8] Choi, S. K., Kim, S. W., & Lee, S. (2020). Development of an Emotion-Based Music Recommendation System Using Facial Expression Recognition. Journal of Information Processing Systems, 16(2), pp 289-298.

[9] Alarifi, A., Alshehri, S., Altowaijri, H., & Alshehri, M. (2021). A Novel Music Recommendation System Based on Emotion Recognition Using Facial Expressions. International Journal of Advanced Computer Science and Applications, 12(3), 52-56.

[10] McFee, B., & Lanckriet, G. (2011). The Natural Language of Playlists. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR) (pp. 547-552).

[11] Lamere, P. (2008). Social Tagging and Music Information Retrieval. In Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR) (pp. 521-526).

[12] Recommender Systems: Exploiting Audio Features and User Context. ACM Transactions on Intelligent Systems and Technology, 11(5), Article 60.

[13] Wang, S., Zhu, W., & Jin, X. (2019). Collaborative Filtering Recommendation on the Spotify API. In 2019 5th International Conference on Control, Automation and Robotics (ICCAR) (pp. 172-176).

[14] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., "Challenges in representation learning: A report on three machine learning contests," in International Conference on Neural Information Processing. Springer, 2013. (pp. 117–12)

[15] Lu, Peng & Song, Baoye & Xu, Lin. (2020). Human face recognition based on convolutional neural network and augmented dataset. Systems Science & Control Engineering. 9. 1-9. 10.1080/21642583.2020.1836526.

[16] D., Roja. (2023). Music Recommendation System for Detection of Emotion. Volume XI.

[17] Krishna Kuma Singh and Payal Dembla (2023). "A Study on Emotion Analysis and Music Recommendation Using Transfer Learning" in Journal of Computer Science.