Grenze International Journal of Engineering and Technology, January Issue

Personalized Smart Music Recommender Therapy Based on Facial Emotion Recognition

S. Selvi
Department of Computer Science and Engineering, Government College of Engineering, Bargur, Tamil Nadu, India
Email: sel_raj241@yahoo.com

***Abstract*—Music shows profound impact on publics' emotions and can be used as a tool for mood regulation. With the growing availability of wearable devices, facial expression recognition technology has become more accessible. This technology can be leveraged to make a personalized music recommendation system based on the one's sensitive state. This paper examines the neural mechanisms underlying facial expression processing in response to music and compares emotional communication through music with other modalities such as speech. Understanding the intricate relationship between music and facial expression holds promise for enhancing therapeutic interventions, particularly in areas such as emotional expression, communication, and social interaction. In this system, people's expressions captured by a camera, which are then analyzed to decide the user's emotional state. The experimental results based; a music recommendation algorithm suggests a playlist that is the most likely to resonate through the user's current emotional state.**

***Index Terms*— CNN, Emotion, Face Recognition and Music Recommendation.**

I. INTRODUCTION

Music acts as a potent means of social connection, bringing individuals together through shared experiences and a common love for genres or artists. Participating in concerts or festivals with loved ones can strengthen relationships and forge lasting memories. Furthermore, music benefits cognitive development on children. Playing a musical instrument improves hand-eye coordination, fine motor skills, and memory, also encouraging self-expression and creativity. The goal of music therapy centered on facial emotion or expression is to equip individuals with skills that can be applied in everyday life to enhance expressive well-being. By using emotional regulation methodologies that incorporate music and users' facial emotions, users can advance their lifestyles. Essentially, this approach to music therapy fosters resilience, emotional growth, and well-being, investing persons to model fulfilling their lives.

Recommender algorithms are designed to filter and present the most appropriate items to persons by analyzing a vast pool of data, identifying patterns that align with user preferences and interests [1]. While many recommendation systems focus on user behavior and preferences, they often overlook the impact of human expressions. Emotions significantly influence people's daily lives, and for applications such as computer-aided tutoring, neuro marketing, human-robot interaction, socially intelligent software applications, and emotion-aware interactive games, it is crucial to consider the emotions of human interaction partners. Techniques such as facial expression recognition and speech analytics [2, 3] have been employed for emotion identification in these contexts.

Grenze ID: 01.GIJET.11.1.441_1

II. Related Works

The system outlined introduces an innovative approach [4] utilizing the DEAM dataset to classify emotions within audio files, offering a nuanced understanding of musical content. With over 2800 songs annotated for emotions like happiness, sadness, anger, and relaxation, alongside valence and arousal values, it provides a rich foundation for analysis. Kamble and Kulkarni's system [5] employs PCA for feature extraction and Euclidean distance classification to recognize expressions, facilitating efficient processing and accurate identification of emotional states. By tailoring music playback based on the user's emotional state, the system enhances the music listening experience, potentially revolutionizing personalized music recommendation services. This approach could find applications in various domains, including mood-based playlist generation and therapeutic interventions leveraging music's emotional impact. Further enhancements or extensions could be explored to refine the system's accuracy and broaden its applications. Livingstone and Russo [6] have proposed for a Comparison of Speech and No Speech Vocal Emotions. This study compared facial expression responses to music with those elicited by speech and no speech vocal emotions, highlighting similarities and differences in emotional communication across modalities. Studies [7] have proposed deep learning models such as CNN augmented for facial expressions to categorize emotions. Basic ML models [8] such as SVM has been used in the detection of facial emotions based on local region with specific features. An innovative method [9] for recommending songs based on the user's emotional state has been proposed. Utilizing inputs such as heart rate data from smart bands or facial images captured by mobile cameras, the application assesses the user's current emotion. It then suggests songs that correspond to the user's emotional state, categorized into four primary types: angry, happy, neutral, and sad. Future developments could focus on refining the accuracy of emotion detection algorithms and expanding the range of emotional categories for more nuanced song suggestions.

Another innovative fusion of CNN for emotion recognition with ANN for song classification has shown better accuracy [10]. An automated approach [11] has been suggested, which shows the same improvements as the other detection processes, was found as performed well reliably with varying resolutions. Developing a method for emotions recognition under the circumstances of poor resolution images. Research of facial emotions expression using Lightweight CNN such as multi- task cascaded CNN architectures [12] and conditional random fields [13] comparative analysis has proposed with better accuracy.

In modern years, numerous experiments have incorporated emotional and contextual information provided like name of the artist, song title, lyrics of the music based on user activity through mobile device into music recommendation systems [14, 15, 16]. The importance and ability of integrating context and understanding musical emotions in music listening are recognized in both fields [17, 18]. A collaborative recommendation system [19] applying deep learning techniques that considers emotions is exemplified by the MoodPlay application [20] and m-motion application [21] facilitates users' desired emotional state through the played songs list. Also, the Assuncao and Neris [22] algorithm, considers three input constraints: representing the desired and current emotional states, and a set of songs along with emotions represented in a model.

Understanding emotions in music and how music elicits emotional responses [23] are the major keys playing main role in the music psychology. Emotion has gained considerable attention from researchers and has enhance major trends in music discovery and recommendation [24]. Identifying the emotions conveyed by music or the user's current emotional state is crucial for music recommendation systems, as it helps users find music that aligns with their mood. A music recommendation system which utilizing collaborative filtering [25] has introduced that suggests tracks depends on mouse click patterns and keyboard patterns. This approach directly recommends music, avoiding the need to label the emotions, thereby preventing potential errors in emotion estimation from affecting recommendation accuracy.

A. Motivation

Facial emotion recognition methods, utilizing CNN architectures, explore facial acts micro-expressions such as gestures, and expressions to imply emotions with extreme precision. This paper employs Google's deep learning (DL) along with open library "Keras" for facial emotion recognition, operating a CNN for image recognition. It is utilized FERC dataset to train with the proposed network, focusing on accuracy and loss metrics. The dataset contains images representing seven emotions, and to detect these expressions using an emotion recognition model developed with a CNN through deep learning. This paper formulates modifying certain steps in the CNN and the architecture itself, assessed to existing methods, by the Keras library, resulting an improved accuracy.

III. PROPOSED METHODOLOGY

The proposed system for music therapy based on facial expression leverage state-of-the-art deep learning techniques, such as Convolutional Neural Networks (CNNs) is implemented using frameworks like Keras, to advance the integration of technology and therapy. Facial emotion detection algorithms, built upon CNN architectures, analyze facial features, including expressions, gestures, and micro-expressions to infer emotions with high precision.

A. Architecture of the Personalized Music Recommendation System

The workflow of facial emotion detection presented in Fig. 1 illustrates the high-level structure and interactions between components involved in integrating facial emotion detection technology into music therapy interventions. The user interface, designed for seamless interaction, allows users to initiate therapy sessions, provide feedback, and customize their music preferences. By continuously refining its recom- mendations based on user's emotional responses, the system aims to deliver personalized and effective music therapy interventions. Through the integration of CNNs, Keras, and deep learning methodologies, this proposed system represents a cutting-edge approach to harnessing the therapeutic power of music in promoting emotional well-being, paving the way for innovative solutions in the field of affective computing and therapy. A few steps in CNN to be changed and modified CNN architecture which give better accuracy.

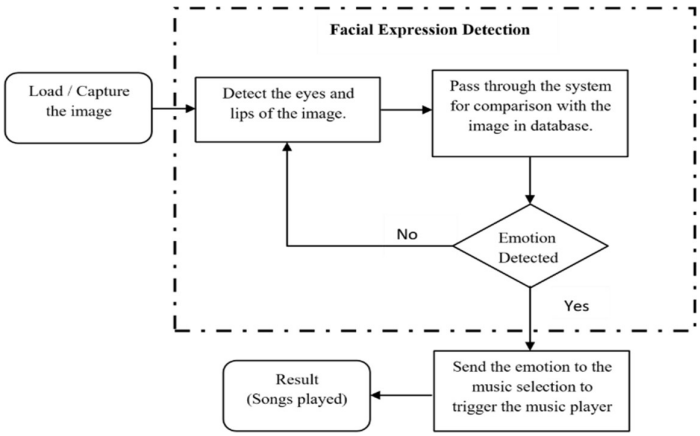


Figure 1. Basic workflow of the proposed system

B. Algorithm of the Personalized Music Recommendation System

Step 1: From FER2013 database, data set of images is assembled. This collection comprises 35,887 grayscale facial images that have been pre-cropped to dimensions of 48-by-48 pixels. Each image is meticulously labeled with one of seven distinct emotion classes: neutral, anger, happiness, disgust, surprise, fear, and sadness.

Step 2: Image Preprocessing as next, various techniques such as resizing, normalization, and filtration are applied to optimize the images for subsequent analysis and model training. These preparatory measures ensure consistency and quality across the data set, laying a solid foundation for accurate and effective processing in later stages of the workflow.

Preprocessed Image=Apply (Resize (Input Image), Normalize (Input Image), Filter (Input Image)) here, the term such as Input Image represents the acquired facial image or video recording; Resize which resizes the input image to a standard size to ensure compatibility with the facial emotion detection algorithm; Normalize to normalize the pixel values of the input image to improve contrast and reduce variation in illumination; Filter (Input Image) applies noise reduction or filtering techniques to remove artifacts or irrelevant information from the input image.

Step 3: Focuses on detecting the presence of a face within each image, a pivotal task for isolating regions pertinent to emotion analysis.

Step 4: handles the conversion of the detected face into grayscale images, streamlining subsequent processing stages.

Step 5: ensures uniformity across all images by formatting them into (1, 48, 48) arrays, facilitating compatibility with the neural network's input layer.

Step 6: Involves passing these arrays through a Convolution2D layer, which generates feature maps by applying convolutional filters.
Step 7: Employs MaxPooling2D, a pooling method that retains only the maximum pixel values within (2, 2) windows across the feature maps, effectively condensing information while preserving crucial features.
Step 8: Encompasses both forward and backward propagation through the neural network, enabling adjustments to model parameters to minimize error.
Step 9: Apply the SoftMax function as next, to yield probability distributions for each emotion class, offering detailed insights into the emotional composition of the detected facial expressions. This comprehensive pipeline facilitates precise emotion recognition and analysis from facial images.
Step 10: Based on the emotion detected the music is played respect to the playlists which has been created by the user itself.

C. Dataset Used

The FER2013 dataset, utilized for training the model, is sourced from the Kaggle. Within this dataset, seven distinct emotional categories are annotated: Neutral (6), Surprise (5), Sadness (4), Happiness (3), Fear (2), Disgust (1), and Anger (0). With a comprehensive training set comprising 28,709 examples and a curated public test set of 3,589 examples, each image encapsulates grayscale facial features measuring 48x48 pixels. Notably, all facial expressions are aligned to ensure consistent centrality and proportional representation within each image. Leveraging this prepared dataset, the proposed network undergoes rigorous validation, meticulously assessing validation accuracy and loss metrics. Through the application of cutting-edge convolutional neural networks (CNNs) within the realm of deep learning, the model exhibits unparalleled proficiency in accurately detecting and classifying facial expressions across the spectrum of seven distinct emotions.

IV. Experimental Results and Discussion

This system applies the DL open library "Keras" from Google platform for the detection of facial emotion, and CNN for image recognition. It is demonstrated that a deep learning method can be useful for personalized music recommender system. To model the system, CNNs use a mathematical operation called convolution to extract features from images. Model training for facial expression analysis in the context of music therapy as in (1) involves several key steps aimed at optimizing the performance of the model. Initially, a suitable deep learning constructive model is selected depending upon on the specific requirements of the task, which can be used for data training module in the music recommender.

$$\text{Min}\left(\sum_{m=1}^{N} N * L\left(Y_i, f(X_i;\theta)\right)\right). \qquad (1)$$

Here, N specifies the count of training samples, L is a loss function which measures the discrepancy among the predicted emotion and the smoothed truth label. X_i represents input facial image. Y_i represents the corresponding ground truth label (i.e., the true emotional state) and (X_i; θ) is the output of the facial emotion detection model with parameters θ.
Preprocessed facial expression data, comprising feature vectors and corresponding emotion labels, is prepared for training, ensuring compatibility with the chosen model architecture. The model parameters are then initialized, and a suitable loss function and optimizer are selected to quantify the prediction error and update the model parameters iteratively during training. In training phase, the model is iteratively fed batches of training data, and the optimizer minimizes the loss function by adjusting the model parameters through backpropagation of gradients.

A. Performance Evaluation of the Personalized Music Recommendation System

The system's performance is monitored on a separate validation set to measure generalization and prevent overfitting, with early stopping mechanisms implemented to halt training when performance ceases to improve. Once training is complete, resultant system is evaluated on an individual test set to assess its generalization ability to unseen data, and hyper parameter tuning may be performed to optimize model performance further. Finally, the trained model parameters and architecture are saved for future use or deployment in real-world applications. This exhaustive regimen of model training guarantees the establishment of precise and durable models specialized for the recognition of facial expressions within the context of music therapy applications.
In model testing for facial expression analysis in music therapy, a separate test dataset is prepared, containing unseen facial expression samples. The trained model parameters and architecture are loaded from disk, and

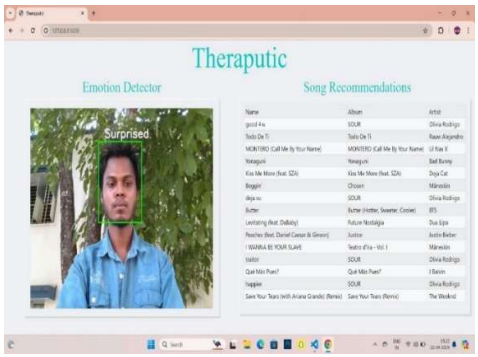
inference is performed on the test dataset to obtain predicted emotion labels. Evaluation metrics such as accuracy, precision, recall, and F1-score are utilized as in table I to assess the model's performance and the corresponding values are listed in table II. This music recommender system shows the better accuracy of 89.72% which visualize its classification performance across different emotion classes. Based on this facial emotion detection system, this proposed music recommender systems suggests the correspondingly classified song as user likes. This playlist is also shown in the Fig. 2.

TABLE I. EVALUATION METRICS

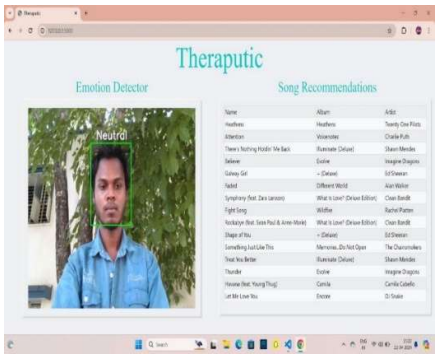
F-measure	$F - Score = 2 * \frac{\text{Pr } ecision * \text{Re } call}{\text{Pr } ecision + \text{Re } call}$
Precision	$\text{Pr } ecision = \frac{TP}{TP + FP} * 100$
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100$
Recall	$\text{Re } call = \frac{TP}{TP + FN} * 100$

TABLE II. EVALUATION MEASURES

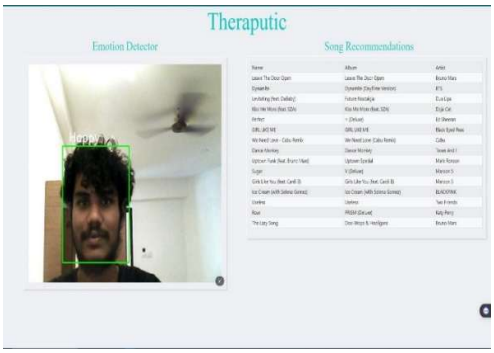
Emotions Identified	Precision	Recall	F1-Score
Angry	0.39	0.32	0.35
Disgust	0.36	0.58	0.45
Fear	0.40	0.39	0.39
Happy	0.45	0.43	0.44
Neutral	0.31	0.39	0.35
Sad	0.36	0.33	0.34
Surprise	0.58	0.57	0.58



(a)



(b)



(c)

(d)

(e)

Figure. 2. Detection of different emotions such as(a) surprise, (b) neutral, (c) happy, (d) angry, (e) sad

Performance analysis is conducted to identify strengths and weaknesses of the model in recognizing specific emotions, and error analysis is performed to understand common types of mistakes made by the model. Optionally, threshold adjustment and cross-validation may be applied to optimize performance and assess robustness, respectively.

V. Conclusion

The integration of facial expression recognition technology into music therapy holds significant promise for enhancing the effectiveness and personalization of therapeutic interventions. Through the analysis of facial expressions, this innovative approach enables a immersed insight of clients' expressive states and responses to music, allowing therapists to tailor interventions to individual needs more effectively. By leveraging CNNs techniques, facial expression recognition systems can accurately detect and interpret subtle nuances in facial expressions, providing valuable insights into clients' emotional experiences during therapy sessions. However, the responsible and ethical use of this technology requires careful consideration of privacy, security, and bias-related concerns, as well as ongoing research and collaboration between therapists, researchers, and technology developers to maximize its potential benefits for clients' well-being. Such systems have the potential to present Individuals benefiting from customized music suggestions that match their emotional state, which could have positive effects on their well-being.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Melville and V. Sindhwani, “Recommender systems”, in Encyc. of mach. learn. Springer, 2011, pp. 829–838. DOI: 10.1007/978-0-387-30164-8_705

[2] N. Sebe et al., “Multimodal emotion recognition”, Handbook of Pattern Recognition and Computer Vision, vol. 4, pp. 387419, 2005.

[3] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: Analysis of affective physiological state”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 10, pp. 1175–1191, 2001. DOI:10.1109/34.954607

[4] A. Arora, A. Kaul and V. Mittal, "Mood based music player," International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 333-337, 2019. DOI: 10.1109/ICSC45622.2019.8938384.

[5] S. G. Kamble and A. H. Kulkarni, "Facial expression based music player," International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016, pp. 561-566. DOI:10.1109/ICACCI.2016.7732105

[6] S, R. Livingstone and F. A. Russo, 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, 2018. DOI:10.1371/journal.pone.0196391

[7] S.k. Sana, G. Sruthi, D. Suresh, G. Rajesh, and G.V. Subba Reddy, “Facial emotion recognition based music system using convolutional neural networks”, Materials Today: Proceedings, Vol. 62, Part 7, 2022, Pages 4699-4706, ISSN 2214-7853. DOI:10.1016/j.matpr.2022.03.131.
[8] D. Ghimire, S. Jeong, J. Lee et al., “Facial expression recognition based on local region specific features and support vector machines”, Multimed Tools Appl, 76, 7803–7821 (2017). https://doi.org/10.1007/s11042-016-3418-y
[9] K. Chankuptarat, R. Sriwatanaworachai and S. Chotipant, "Emotion-based music player", 5th International Conference on Engineering Applied Sciences and Technology (ICEAST), 2019.
[10] K. Patel and R. K. Gupta, "Song playlist generator system based on facial expression and song mood," International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-6. DOI: 10.1109/AIMV53313.2021.9670976.
[11] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches", IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 1-12, March 2015. DOI: 10.1109/TAFFC.2014.2386334
[12] N. Zhou, R.Liang, And W. Shi, “A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection”, IEEE Access, 9, pp.5573-5584, 2021. DOI:10.1109/ACCESS.2020.3046715
[13] J. J. Deng, and C. Leung, “Emotion-based music recommendation using audio features and user playlist”, In: 2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012), IEEE, pp.796–801, 2012.
[14] Abdul, J. Chen, H. Y. Liao, and S. H. Chang, “An emotion-aware personalized music recommendation system using a convolutional neural networks approach”, Applied Sciences, 8(7), 1103, 2018. DOI: 10.3390/app8071103
[15] J. S. Bauer, A. L. Jellenek, and J.A. Kientz, “Reflektor: An exploration of collaborative music playlist creation for social context”, In: Proceedings of the 2018 ACM Conference on Supporting Groupwork, pp.27–38, 2018. DOI:10.1145/3148330.3148331
[16] D. Bogdanov, et al., “Semantic audio content-based music recommendation and visualization based on user preference examples”, Information Processing & Management 49(1), pp.13–33, 2013. DOI: 10.1016/j.ipm.2012.06.004
[17] S. Aalbers, et al., “Efficacy of emotion-regulating improvisational music therapy to reduce depressive symptoms in young adult students: a multiple-case study design”, The Arts in Psychotherapy, 71, 101720, 2020. DOI:10.1016/j.aip.2020.101720
[18] J. J. Deng, C. H. Leung, A. Milani, and L. Chen, “Emotional states associated with music: Classification, prediction of changes, and consideration in recommendation”, ACM Transactions on Interactive Intelligent Systems (TiiS) 5(1), pp.1–36, 2015. DOI:10.1145/2723575
[19] F. Fessahaye, et. al., “T-recsys: A novel music recommendation system using deep learning”, In: 2019 IEEE international conference on consumer electronics (ICCE), IEEE, pp.1–6, 2019. DOI: 10.1109/ICCE.2019.8662028
[20] Andjelkovic, D. Parra, and J. O’Donovan, “Moodplay: Interactive music recommendation based on artists’ mood similarity”, International Journal of Human-Computer Studies, 121, pp.142-159, 2019. DOI:10.1016/j.ijhcs.2018.04.004
[21] W. G. Assuncao, and V. Neris, “m-motion: a mobile application for music recommendation that considers the desired emotion of the user”, In: Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems, pp.1–11, 2019.
[22] W. G. Assuncao, and V. Neris, “An algorithm for music recommendation based on the user’s musical preferences and desired emotions”, In: Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, pp.205–213, 2018.
[23] T. Eerola, and J. K. Vuoskoski, “A review of music and emotion studies: approaches, emotion models, and stimuli”, Music Perception: An Interdisciplinary Journal, 30(3), pp.307–34, 2013.DOI:10.1525/mp.2012.30.3.307
[24] P. S. Iordanis, “Emotion-aware music recommendation systems”, 2021.
[25] S. Y. Jazi, M. Kaedi, and A. Fatemi, “An emotion-aware music recommender system: bridging the user’s interaction and music recommendation”, Multimedia Tools and Applications, 80(9), pp.13559–13574, 2021.