The International Journal of Engineering and Science (IJES)
|| Volume || 13 || Issue || 12 || Pages || PP 90-95 || 2024 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Emotion-Based Music Recommendation System

Sushil Kumar
Computer Science and Engineering, Maharaja Surajmal Institute of Technology, New Delhi, India
Corresponding Author: Sushil Kumar

----------***ABSTRACT***----------
In an increasingly fast-paced world, individuals often struggle to find things matching their current emotional state, leading to inefficiencies in emotional regulation, stress relief, and overall mental well-being. Traditional music recommendation systems rely on user inputs or previous playlist, which may not accurately capture a person's real-time mood. There is a need for a more intuitive and responsive solution that can automatically detect and respond to an individual's emotions to enhance their music listening experience. A Flask web application that utilizes facial expression recognition (FER) to detect emotions in real-time and recommend music that aligns with the detected mood. By integrating the FER 2013 dataset and Spotify API, system provides a personalized and dynamic music recommendation experience. Additionally, implemented a generative AI to deliver engaging content. The aim is to increase user retention and engagement by enriching theoverall music listening experience.
KEYWORDS; - *Facial expression recognition, Generative AI, Convolutional neural network, Music recommendation.*

Date of Submission: 14-12-2024 Date of acceptance: 29-12-2024

I. INTRODUCTION

Emotions are critical to understand the basic human nature and mood. Recent studies suggests that music influences humas mood and sometimes play active role in rejuvenating the human mood. Musical preferences have been demonstrated to be highly related to personality traits and moods [1]. The choice of music on the basis of human mood plays important role in many cases such has happiness, sadness, anger on personal development and the ability to heal disease [2]. Music Therefore, it is basically required to understand human mood first to play music which is highly required in today's stressful world and to understand mood expression plays very significant role. Understand emotion based upon facial expression helps in recommending music. Emotion plays a crucial role in shaping human experiences, and music, as a powerful emotional trigger, can significantly enhance or alter one's mood.

While existing music recommendation systems have made substantial progress in curating playlists based on user preferences, they often overlook the real-time emotional state of the listener. This gap in existing systems limits their ability to fully resonate with user's current feelings and emotional needs. Despite the promising advancements in emotion-basedsystems, several challenges remain, such as the accuracy of emotion detection in varied lighting conditions, the seamless integration of real-time data processing, and the need for intuitive user interfaces.

The proposed system is based upon Flask-based web application with integrated Spotify API for music recommendation, marks a significant milestone in this area by leveraging facial emotion recognition (FER) to create personalized music playlists and provide engaging content about the music.

The proposed system addresses the challenges by integrating state-of the-art facial emotion recognition technology with the SpotifyAPI. By analyzing a user's facial expressions in real-time, oursystem can detect their current emotional state and dynamically recommend music that aligns with their mood. The integration of real-time emotion detection and adaptive content generation holds immense potential to shape thefuture of user-centric digital experiences. The dataset used in this system for training model is FER2013. The objective of this work is as follows

- To create a music recommendation system based on the emotion identified from facial images.
- To propose techniques to enhance accuracy of models using different tools and techniques.

The remaining paper is organized as follows. Section 2 of the focus on the literature survey to discuss related work in area of music recommendation models base on expression recognition. Section 3 presents the tools and techniques used in proposed work. The results of proposed model from music recommendation using facial expression are discussed in section 4 and section 5 presents the conclusion and future scope of the work in the field of music recommendation.

II. Literature Survey

The mental state of a person can be described by his facial expression and recognition of facial expression is done using various facial expression recognition systems and further can be utilized for music recommendation system. Mammadli et al [3] proposed a machine leaning based application using to detect emotions. A CNN models is used to make predictions, comparing different layer combinations for accuracy. Data preprocessing involves cleaning datasets and discarding irrelevant classes. Additionally, the system suggests music playlists based on the determined attributes. Yusuf Yaslan et al.[4] proposed an music recommendation system that uses emotion from signals obtained through wearable computing devices that are integrated with galvanic skin response (GSR) and photoplethysmography (PPG) physiological sensors. In [5] Ayush Guidel et al stated that human being's state of mind and current emotional mood can be easily observed through their facial expressions. A system was developed by taking basic emotions (happy, sad, anger, excitement, surprise, disgust, fear, and neutral) into consideration.

Yading Song et al [6] suggested music recommendation systems based upon emotion recognition in music. It addresses challenges in defining emotions and proposes a model categorizing them based on valence and arousal. It suggests using social tags, lyrics, and music reviews for data collection instead of costly human annotation. Parul Tambe et al. [7] proposed an idea which automated the interactions between the users and music player, which learned all the preferences, emotions and activities of a user and gave song selection as a result. The various facial expressions of users were recorded by the device to determine the emotion of the user to predict the genre of the music. [8] came up with an algorithm that gives a list of songs from the user's playlist in accordance with the user's emotion. The algorithm which was designed was focused on having less computational time and also thus reduces the cost included in using various hardware. The main idea was to segregate the emotions into five categories i.e., Joy, sad, anger, surprise and fear also provided a highly accurate audio information retrieval approach that extracted relevant information from an audio signal in less time.

N. Deshmukh et al. [9] focused on creating a system that fetches the emotion of the user using a camera and then automates the result using the emotion detection algorithm. This algorithm captures the mood of the user after every decided interval of time as the mood of the user may not be the same after some time; it may or may not change. The proposed algorithm on an average calculated estimation takes around 0.95-1.05 sec to generate an emotion-based music system, which was better than previous existing algorithms and reduces the cost of designing.

Chang Liu et al[10] described a system that makes use of Brain-Computer Interfaces, also called as BCI. BCI makes use of devices to send signals to the processing systems. EEG hardware is used in to monitor the person's cognitive state of mind. The drawback of the scheme is that they require the input from the user's brain continuously to perform the classification. An algorithm based on MID is used to continuously monitor and process the signals received from the brain of the user and use these signals to actively monitor and generate emotions that the user is currently experiencing.

III. EMOTION BASED MUSIC RECOMMENDATION

Music Recommendation system based upon emotion works on the principle of facial expression. Facial expression is first classified using convolution neural network and then using expression music is recommended to the user.

The music recommendation accuracy depends upon the accuracy of facial expression. The proposed recommendation system has two variations as shown in fig. 1. The model 1 works on the dataset FER2013 without any augmentation techniques. In model 2 data augmentation was employed to artificially expand the dataset. This was crucial for improvingthe model's generalization ability. Techniques like random rotations, flipping, and zooming, helped introduce variability into the training data, simulating different conditions and perspectives in which facial expressions might appear. This step is particularly important when working with datasets that could otherwise be imbalanced, ensuring that themodel is exposed to diverse representations of each emotion. The proposed models work in stages as described.

Data collection

The dataset utilized in the proposed model is **FER 2013.** The data consists of 48x48 pixel grayscale images of faces. facial images in dataset are categorized into seven emotional states: happiness, sadness, anger, fear, disgust, surprise, and neutral. The dataset in divided into two set namely training and testing. The training set consists of 28,709 and testing test has 28,709 images. To ensure the dataset's reliability, a series of pre-

processing steps were applied to the raw data. Data normalization techniques is used to standardize the pixel values of the images, preventing large values from skewing the model's learning process. This ensures that no single feature dominates the training process and that the model can learn uniformly from all inputs.

Data augmentation

Data augmentation is employed to artificially increase the size of dataset to improve the stability and accuracy of a model by preventing it from overfitting. Data augmentation also helps in reducing the operational cost labeling and cleaning the data. Geometric transformation like rrandom flip, crop, rotate, stretch, and zoom etc has been used for implementing augmentation under certain limits to retain quality of dataset.

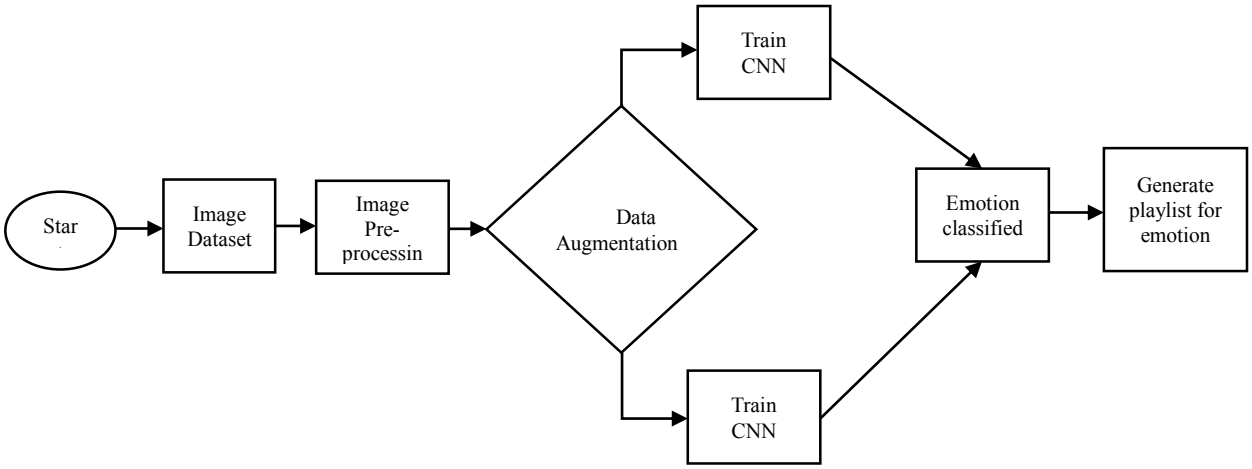

Figure 1: Proposed System

Feature extraction

Feature extraction is the process of identifying key patternsor characteristics within a dataset that the model can use to classify data into different categories, in this case, emotional states. In an emotion-based music recommendation system, feature extraction focuses on identifying the critical facial attributes that correspond to specific emotions such as happiness, sadness, anger, and surprise

Classification

Convolutional Neural Networks (CNN) is central to this process due to their ability to automatically extract and learn from visual patterns. Classification models for Emotion-Based MusicRecommendation System involved creating two different model, each designed to improve the accuracyof emotion classification. Two different CNN models were developed and tested to identify the best-performing.

The CNN model without data augmentation was the initial model used for emotion classification. This basic architecture consisted of sequential convolutional layers followed by max-pooling and dense layers as shown in fig. 2.

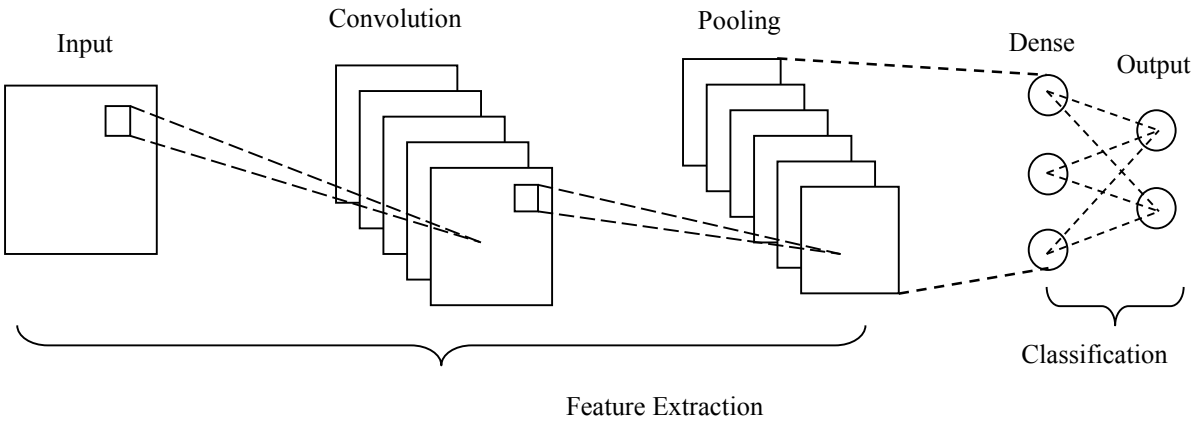

Figure 2: General Structure of Convolution Neural Network

Convolution is an operation performed on image for feature extraction. This operation is performed between the input image and kernel of size N×N for generating features as shown in fig. 3. Further pooling layer is responsible for reducing the spatial size of image. The max polling is commonly used layer. The output of this layer is finally passes to fully connected layer. The CNN uses backpropagation algorithm with rectified Linear Unit (ReLU) as activation function.

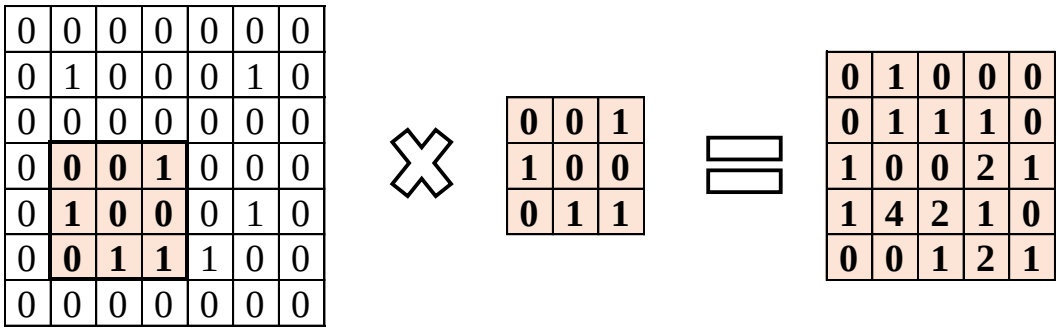

Figure 3: Convolution Operation

Playlist generation

The emotion detection model is most likely based on image or voice analysis and identifies the user's current emotional state (e.g., happy, sad, relaxed). Each emotion is then mapped to specific music genres, moods, or song characteristics. With this emotion-genre mapping, the system sends query to the Spotify API to generate music/playlist corresponding to the current emotional state based upon his facial expression.

IV. RESULT AND DISCUSSION

In the proposed system FER2013 Image dataset is given as input to CNN models. In model 1 image dataset is directly fed to CNN as input. For model 2 Various transformation on images of dataset are applied so that augmented dataset can be prepared and fed as an input the second model. The transformation with parameters are shown in table 1.

Table 1: Data Augmentation factor

Transformation Type	Transformation Factor
Rotation	Random Rotation within range of 30
Shift	Width and height Shift -0.1
Shear	Shear range=0.2
Zoom	0.2 (Random zoom in the range 80-120%)
Flip	Horizontal Flip
Fill Mode	Fill mode=nearest (to main image continuity)

The first CNN model (Model 1) consists of 3 convolutional and all convolutional layers have kernel of size 3×3, Two Maxpooling layers with pool size of 2×2 and two fully connected layers. Two drop layers with factor 0.25 and 0.5 are also added to avoid overfitting in the model. This CNN model is classifying seven emotions as an input to Spotify for music recommendation and working on FER2013 without data augmentation. The Second CNN model (Model 2) is working on augmented FER2013 dataset as input and consists of 5 convolutional layers with kernel of size 3×3, 5 maxpooling layers with pool size of 2×2, 2 fully connected layers and 2 drop layers with factor 0.25. This model is also classifying seven emotions for music recommendation. Both the model are simulated as per the parameters in table 2.

Table 2: Model training Parameter Values

Parameters	Model 1	Model 2
Learning Rate	0.001	0.001
Decay rate (First moment)	0.0001	0.0001
Decay rate (Second moment)	0.999	0.999
Epochs	40	100

Batch Size	32	64
Training Dataset	70	70
Validation Dataset	30	30
Loss	categorical_crossentropy	categorical_crossentropy

FER2013 dataset has been used from training the models. The 70% of image dataset is a is used for training the model and 30% of dataset has been utilized for testing. On the basis of classified emotion music recommendation in the form music or playlist is generated. Along with songs the back history and interesting facts corresponding to songs area also generated

The CNN models are evaluated on the basis of result deduced from the confusion matrix. The confusion matrix is the representation of predicted and actual class in row, column format. The confusion matrix of model 1 and model are shown in fig. 4. The generated results from confusion matrix are based upon true positive (*tp*) and true negative (*tn*) values representing currently classified classes of positive and negative. Similarly misclassified are represented as false positive (*fp*) and false negative (*fn*).

	Angry	Disgust	fear	Happy	Neutral	Sad	Suprise
Angry	116	13	123	241	174	187	104
Disgust	11	3	13	27	22	28	7
fear	104	14	130	269	186	185	136
Happy	186	20	227	502	299	334	207
Neutral	132	12	146	333	223	251	136
Sad	149	14	160	320	233	232	139
Suprise	89	10	116	201	146	164	105

Model 1

	Angry	Disgust	fear	Happy	Neutral	Sad	Suprise
Angry	487	31	65	79	119	145	32
Disgust	21	58	5	8	4	11	4
fear	105	15	308	72	130	249	145
Happy	27	4	26	1553	76	48	40
Neutral	57	6	38	136	229	189	28
Sad	119	8	94	117	240	644	25
Suprise	19	4	43	50	33	19	663

Model 2

Figure 4: Confusion Matrix of model 1 and model 2

The models are evaluated on different evaluations metrics like Accuracy, precision, recall and F1 score. The evaluated results of model 1 and model 1 are shown in table 3.

Accuracy

The accuracy of model is defined as how accurately model predicting emotion for music recommendation and defined as:

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad \ldots \quad (1)$$

The accuracy of model 1 is evaluated to be 18% where as in model 2 the accuracy has shown improvement is 63%

Precision

The exactness of classifies is termed as precision. It is also termed as positive predictive value and is the ratio between positive pattern detected out of the total predicted as positive.

$$Precision = \frac{tp}{tp+fp} \quad \ldots \quad (2)$$

Recall/Sensitivity or the True Positive Rate (TPR)
Completeness of classifier is called recall or sensitivity. It is the ratio of positives identified correctly. And is defined as

$$Recall = \frac{tp}{tp+fn} \quad \ldots. \quad (3)$$

F1 Score/F Score/F Measure
It is defined as harmonic mean between precision and recall.

$$F1\ Score = 2 * \frac{preciosn*recall}{precicion+recall} \quad \ldots. \quad (4)$$

Table 3: Quotative analysis of Model1 and Model1

Emotions	Model1			Model2		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Angry	0.14	0.12	0.13	0.58	0.51	0.54
Disgust	0	0	0	0.46	0.52	0.49
Fear	0.13	0.12	0.13	0.53	0.3	0.38
Happy	0.27	0.28	0.27	0.77	0.88	0.82
Neutral	0.17	0.18	0.18	0.56	0.63	0.6
Sad	0.17	0.18	0.18	0.49	0.52	0.5
Surprise	0.12	0.12	0.12	0.71	0.8	0.75

From table 3 it is revealed that model2 has better performance not in terms of accuracy but in terms of other parameters also. Model1 not only has low accuracy but emotion disgust is not recognised at all in comparison to model2 which has precision of 0.46. The recall value has also improved in case of model2 and for happy emotion it is close to ideal value i.e., 1.0. The F1 score also in case of happy emotion is close to 100%. In model2 significant improvement has been seen considering the sad emotion. Overall, the performance of Model2 is better than model1.

V. CONCLUSION

In this work the focus is to enhance the accuracy of CNN for music recommendation on the basis facial expression. Considering the two models it has been revealed that the performance of the model can be enhanced by improving the dataset so that model can learn more features. The performances of classification models are evaluated considering results deduced from confusion matrix i.e., accuracy, precision, recall value and F1 score. It is revealed from results that CNN with augmented data has comparable precisions, F1 score and recall value for different emotions. It is revealed that increasing the numbers of layers in model has also impact on its performance. In future, the model can be enhanced by considering deep CNN. The use of different preprocessing techniques and feature extraction can also be utilized for better performance or accuracy of model.

REFERENCES

[1]. R. Ramanathan, R. Kumaran, R. Ram Rohan, R. Gupta and V. Prabhu, "An Intelligent Music Player Based on Emotion Recognition," 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 2017, pp. 1-5

[2]. M. Athavle, D. Mudale, U. Shrivastav, and M. Gupta, "Music Recommendation Based on Face Emotion Recognition", J. Infor. Electr. Electron. Eng., vol. 2, no. 2, pp. 1–11, Jun. 2021.

[3]. Ramiz Mammadli, Huma Bilgin, and Ali Can Karaca "Music Recommendation System based on Emotion" 3; Cornell University, December 2022.

[4]. Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, "Emotion-based music recommendation system using wearable physiological sensors", IEEE transactions on consumer electronics, vol. 14, no. 8, May 2018.

[5]. Ayush Guidel, Birat Sapkota, Krishna Sapkota, "Music recommendation by facial analysis", IJRASET, February 17, 2020.

[6]. Yading Song, Simon Dixon, and Marcus Pearce. "A Survey of Music Recommendation Systems and Future Perspectives", The 9th International Symposium on Computer Music Modeling and Retrieval (CMMR),2012.

[7]. Parul Tambe, Yash Bagadia, Taher Khalil and Noor UlAin Shaikh ,"Advanced Music Player with Integrated Face Recognition Mechanism", International Journal of Advanced Research in Computer Science and Software Engineering.2012

[8]. Dureha A 2014 An accurate algorithm for generating a music playlist based on facial expressions International Journal of Computer Applications 100 33-9

[9]. Rabashette MD, Tale MR, Hinge MA, Padale MK, Chavan MR and Deshmukh N Emotion Based Music System.

[10]. Liu C, Xie S, Xie X, Duan X, Wang W and Obermayer K,"Design of a video feedback SSVEP-BCI system for car control based on improved MUSIC method",6th International Conference on Brain-Computer Interface (BCI) 1-4 IEEE, 2018