International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Music Recommend System Using Facial Emotion Recognition

Nripendra Dwivedi , Shivam Kumar Goit
Dept. of computer science
Galgotias University
Greater Noida, India

shivamkumar0563378@gmail.com
nripendra.dwivedi@galgotiasuniversity.edu.in

ABSTRACT:

The system plays music based on detected facial expressions through an automatic process. The method deploys facial emotion recognition instead of standard manual song selection or wearable devices or audio-based classification systems to recommend music.

Convolutional neural networks, or CNNs, have been applied to as the central component in our system to recognize emotions that people express through their faces. The system operates by playing automatic music through a custom interface which uses combination of Pygame and Tkinter programming.

The efficiency design of our system enables faster processing speed and reduced system expenses. Through this method the system achieves both efficiency and precision.

Our model received evaluation through testing against the emotion labelling (2013) dataset because it is a common standard norm for study on facial expression analysis. The built-in camera records facial expressions that get processed to detect emotions from a list of happiness, anger, sadness, surprise and neutrality.

The system retrieves music content either from individual playlists or from pre-selected mood-based playlists depending on the emotion detection result. By doing so users no longer require manual effort to organize songs according to their mood-based categories.

Through modern computer vision technology integration with basic tools we created an emotional music application which delivers personalizable and fluid listening sessions. Our system provides a smoother music listening experience because it understands users' emotions as they occur.

Key words: Convolutional neural network, pygame, tkinter, music, player, camera, facial detection, sentiment analysis, euphony suggestion, utterances, and pattern recognition

1.INTRODUCTION

Scientific research validates the double effect of musical stimuli on human mental operations and brain neurological reactions. A specific research project investigated musical listening because people depend on music to control emotional states and body arousal. The research participants highlighted that music serves two main purposes to improve emotional state as well as enhance self-knowledge. Research indicates that musical preferences strongly relate to several human personality traits and the emotional states people experience [1].

The fundamental musical components including meter, timbre, rhythm and pitch receive processing within brain regions that regulate emotions and moods [2]. We depend on human interaction for all everyday activities because it involves combining body movements, vocalization, facial expressions and emotional displays [3]. Modern applications increasingly rely on emotion detection capability because it serves purposes that span from smart card technologies to security systems and beyond into video tagging and crime investigations and adaptive human-machine interfaces.

Envision a reality where your music application shuffles your playlist, but also takes note of your mood and plays the most suitable song for the moment. That is precisely what our project aims to achieve. The innovation of The Music Recommendation System using Facial Emotion Detection technology provides a whole new perspective to music experiences personalization. Unlike previous systems that relied on playlists or historical data, this one utilizes deep learning and computer vision to detect human emotions through facial recognition and suggest music that resonates with someone in real time. Facial features and expressions can be analysed to determine an individual's emotional state – whether they are in a state of happiness, sadness, calm, or even excitement. Machines have been trained on large datasets to successfully interpret human expressions, but doing so autonomously is the true test. This is the point in time when machine learning and deep leaning come into play. Our system employs cutting-edge algorithms to facial feature analysis and emotion recognition with even the smallest changes to the face. When the emotion is determined, the system looks for songs in a music database where songs are classified based on emotions. Be it energetic music or a soothing tune, the system provides the 'right' music to help you cope—and feel connected—through sound. But it does not stop there. The system can learn how an individual emotionally responds to various songs over time, making its suggestions more intelligent and tailored. This offers a powerful listening experience that responds to each person's mood and evolves alongside them. There are numerous advantage New recommendations of songs which will vibrate with your mood

• Listening can be a way to manage your feelings - uplifting tracks when you are feeling low, or calming ones when you are stressed

• It gives another layer of feeling to how we feel the music

In a nutshell, this project involves amalgamation of artificial intelligence, intelligence of emotions and music to change the way we hear, feel and associate. It's about turning music recommendations into playlists for your life–with real emotion at the heart of it. So really, we're not creating playlists at all; we're creating soundtracks for life. The framework of the paper is as listed below:The relevant scrutiny on facial expression analysis and music suggestion systems is covered in Section 2. We'll go over our system's architecture, data collecting, and preprocessing procedures in Section 3. The collaborative filtering method for creating music suggestions and the deep learning model for picking out facial expressions are described in Section 4. In Section 5, The assessment findings are exhibited and the system's performance is evaluated. Section 6 outlines the work and identifies possible directions for further investigation.

2.LITERATURE REVIEW

Music recommendation systems have advanced pretty much over the years in a bid to give users more personalized experiences. One of the first methods used, collaborative filtering, recommends music according to a user's previous selections habits along with the habits of like-minded users. While it works well sometimes, collaborative filtering cannot quite get a hold of the mood of a user and hence suggest music according to that state. Facial expression analysis, used in areas like psychology, neuroscience, and computer vision for understanding one's feelings, has now been incorporated into systems that suggest music. Research shows that mixing face expression reading with music suggestions can make recommendations more personal. For instance, work by Yang et al. (2014) made a system which used facial expression check to find out a user's feelings and give music suited to their mood. Hi there! The setup used a deep belief network to link face looks with feelings, plus collaborative filtering to make suggestions. In the next study in 2016, Yang and others grew this idea by mixing face expression checks with body signals like heart rate and skin sweat. This mixed way used deep learning to spot emotions and combined content-based and gathered filtering for developing music suggestions. They deployed a k closest neighbour technique and convolutional neural networks (CNNs) to classify moods based on facial expression identification.

to recommend music that matched the user's emotional state.

The emotional state of the person at present becomes visible through facial indications. People primarily use bodily signs During social interactions, people use movements of the hands, facial expressions, and voice tones to express their emotions.

Interaction happens through body signals together with facial expressions and vocal tone in expressing emotions to others. Preema et al [6] stated

The process of developing numerous playlists takes considerable time while also being complex to handle. The paper demonstrates that the music player utilizes

A music player system picks songs dependent on user mood at the present moment. Audio file classification by the application utilizes audio features for the generation of playlists based on mood detection. The system employs Viola-Jonas algorithm which detects faces along with extracting expressive facial features.

detection and facial expression extraction. The extracted features were categorized using Support Vector Machine (SVM) technology.

The system detects five basic universal emotional states consisting of anger and joy and surprise with sadness and disgust.

A recommendation engine developed by Yusuf Yaslan et al uses signals to determine the emotions of users during system operation.

The authors in paper [3] showed how they obtained signals by using wearable computing devices combined with GSR and PPG physiological sensors. Emotions function as intrinsic features found within all human individuals. People need their emotions for success throughout every phase of their existence. The research examines valence prediction and arousal through physiological signals with several channels as the core emotion recognition issue in this work. The human mental and emotional conditions condition can easily be seen through the emotions on the face according to authors Ayush Guidel et al in their research [7]. The developers focused on establishing this system from fundamental emotional categories that included happy, sad, anger, excitement and surprise, disgust, fear alongside neutrality as base points. The face detection system implemented in this project had an easy execution process.

The system uses a convolutional neural network for its implementation. The statement "language of emotions" describes music as an emotional expression which people use everywhere on Earth.

The written document Ramya Ramanathan et al. [1] described the track composer with intelligence system that uses acknowledgment of emotions algorithms.

Human nature contains emotions as a fundamental element.

Our approach uses a deep learning-based model for identifying facial emotions and a collaborative filtering algorithm for music recommendation. We evaluate the system on facial expression images and music preference datasets and show that it can effectively predict emotional states and provide accurate and personalized music recommendations

3.PROBLEM DEFINITION

Create a music player system with multi-platform compatibility that recommends songs based on current user mood readings using Machine Learning techniques from a web camera.

A system should use supervised learning Algorithms to identify the client's feelings in real time via the webcam data which it will then employ for music recommendation.

4.PROJECTED MODEL

This projected system enables user interaction with the music player through its features. The system functions to acquire proper facial recordings through its camera. The Convolutional Neural Network conducts prediction of emotions based on images sourced from camera capture. The system obtains songs from the database through the analysis of captured image emotional data. Our proposed system's core job is to generate autonomous playlists that change the user's desired mood from pleased to sad, natural, or astonished. Before choosing appropriate playlists, the detection algorithm determines the feelings present in the themes. which will create positive emotional effects for the user. The system which recommends music based on facial emotion detection includes four distinct sections.

The system must correctly detect user faces under real-time operation in this module.

At this point the system requires an input of user face data. The convolutional neural network acts as a program that assesses the featured aspects of user-provided images.

This section analysed user image features for emotion detection then The software will generate labels. that reflect the user's feelings. The tune is recommended by the preference system to users through mood sketching of their feelings.

5.METHODOLOGY

5.1DATABASE DEPICTION

We designed the Convolutional Neural Network architecture through the environment provided by Kaggle. The FER2013 data warehouse contains training and testing sections that make up its partition. The overall quantity of pixels in the experimental data reaches 24176 while testing data has 6043 images. All images in the database feature faces with grayscale 48x48 pixels dimensions. All images in FER-2013 carry labels indicating happiness, sadness, anger and surprise and lack of emotion (neutral). Each face in the dataset receives an automatic alignment to central positioning and standardizes overall face dimensions. Both posed and unposed shots are stored in FER-2013 and are shown in monochrome with a resolution of 48 x 48 pixels.

The developers constructed FER-2013 by conducting Google image searches for all emotions and their respective synonyms. FER systems using imbalanced training datasets show high accuracy on major Feelings like joy, sadness, and anger and Startled while they present low success rates for the underrepresented emotions consisting of Dread and revulsion. The loss scaled by softmax technique exists for fixing this challenge through the use of weight distributions which reflect the emotion class proportions found in training data. The SoftMax loss function operates as the basis for weighted-loss while demonstrating strong capabilities in separating different class features while disregarding intra-class compactness. Using an auxiliary loss as training for the neural network represents an effective solution to resolve SoftMax loss. Categorical crossentropy loss function has been selected to address missing values and Outlier values. A set loss function serves as an error measurement tool during every iteration. The training method received missing and Outlier values through categorical crossentropy as its loss function. The research presented a fresh music recommendation system which used user taste data alongside face expression monitoring. The project utilized images obtained from participants who listened to music while facing the camera. The participants received either positive or negative musical selections and their reactions were recorded through this process. Users provided emotional ratings to each song during the evaluation process. The deep learning model consisted of convolutional and fully connected layers to study facial expressions of participants. The images gathered for this purpose helped develop the model to identify emotions spanning happy and sad and angry and neutral. Music recommendations meant for the user were generated using collected emotional data. The recommendation engine deployed collaborative filtering algorithms through the input features that stemmed from participant music rating data.

Users can access personalized recommendations from the system which matches their current emotional state against their past preferences and connected user preferences. System evaluations occurred through standard metrics accuracy, retrieve and F1 score (the average F1 score was 0.85) for emotional state recognition accuracy while also assessing music recommendation relevance and enjoyment in user testing. The findings were positive because users found their mood fit the recommended music and reported positive feedback throughout the evaluation. Our approach maintains an automatic balance between music emotions since it integrates facial expression features into a collaborative filtering system.

User engagement will substantially rise through a music recommendation platform which delivers individualized and emotionally tuned musical content.

5.2EMOTION DETECTION MODULE

5.2.1FACE DETECTION

The field of computer vision contains face detection as an important technology application. Engineers create detection algorithms through a training process that enables the identification of faces and objects within photographs using object detection or a similar technique. Detecting faces in moving video frames and static pictures can be achieved in real-time. A classifier system operates within face detection by identifying objects through algorithms that confirm if a face exists (1) or lacks (0) within an image. Classifiers gain higher accuracy after receiving training from processing numerous photographic images to locate faces. OpenCV operates with detection programs that utilize Haar Cascades combined with LBP (Local Binary Pattern) classifiers. A Haar classifier operates for face recognition to achieve consistent identification of various faces. The training process utilizes pre-established variables that define face characteristics. Face detection systems operate to identify facial appearances in digital frames by reducing irrelevant noise from background elements. Machine learning-based training of the cascade function relies on a set of input files for its development. A method for understanding image pixels using square divisions can be accomplished through the Haar Wavelet approach which is specified in [9]. The machine learning techniques demand precise "training data" which must achieve high accuracy levels.

5.2.2FEATURE EXTRACTION

The pre-trained sequential network serves as an arbitrary feature extractor to extract features by allowing image input to continue until reaching the specified layer then collecting the layer output. The preliminary convolutional network layers extract valuable information from images so only a small number of filters need to be used. When layers deepen in the model The quantity of filters grows up to double or triple the extent of the preceding filter. The computational complexity of more severe layer filters grows but they collect more properties from the input data. Strong discriminative features of the Convolution neural network enabled us to accomplish this task [10]. Output from the model comes from subsequent feature maps after its initial layer.

We need to examine the feature map because it helps determine the important image categorization elements. The process begins with the visual that was provided or the preceding structure map results that filters or feature detectors apply to generate feature maps. Each convolutional layer succeeds its specific inputs by showing feature maps that explain its internal representations.

5.2.3EMOTION DETECTION

The convolution neural network design implements Relu activation function while processing input images through feature detectors which produce feature maps [11]. Images contain multiple elements which can be detected by feature detectors or filters that identify bends along with edges as well as vertical and horizontal lines. The translation-invariant operation pools the feature maps. During pooling the output keeps its values consistent with any slight change made to the input data. Users have three options when selecting pooling methods: either selecting minimum or maximum or computing averages. The performance of max-pooling surpasses average or min-pooling despite all other options. A deep neural network requires flattened input data as its base for producing class-based outputs. A CNN must determine between binary or multi-class coding based on its intended use for identifying numbers or separating different clothing items. The interpretability of neural network learned characteristics remains unfeasible which makes these networks function similarly to a black box. We grant images to the CNN model which generates its output outputs [10]. After receiving CNN weight-based training the model becomes able to detect emotions. The previously trained CNN model processes real-time images which users provide. The CNN model proceeds to classify the image while generating a prediction of emotion.

5.3MUSIC RECOMMENDATION MODULE

5.3.1SONGS DATABASE

We created a database which contained 100 to 150 Hindi Bollywood songs. The system contains between 100 to 150 songs per emotional state classification. Musical compositions serve to enhance how we feel according to general understanding. When a user shows signs of feeling down then the system activates an automatic playlist selection process that provides mood-lifting songs to improve their state.

5.3.2MUSIC PLAYLIST SUGGESTION

During real-time operation the emotion module recognizes user emotional states which produces reaction plates of Happy, Sad, Angry, Sur prise and Neutral. We attached these tag names to database music folders which we created using os.listdir() from the Python library. Obtaining a list of all files located in target folders is made possible through the execution of the os.listdir() function.

```
if label== 'Happy':

os.chdir("C:/Users/deepali/Downloads/Happy")

Your current mood appears cheerful so I am playing the song for you.

# Fetching Songs

songtracks = os.listdir()

# Inserting Songs into Playlist

for track in songtracks:

self.playlist.insert (END, track)
```

The user interface shows emotion-based captions which generates a playlist suggestion through its graphical interface. The application uses Pygame because it supports multimedia format playback including audio, video and other content types. The music player features play song, pause song, resume song and stop song functions obtained from this library to operate the songs. The songs list names are stored in variables playlist, song status and root. These variables additionally function in both displaying the main GUI and tracking currently playing songs. The development of the GUI interface uses Tkinter for its implementation.

6.DESIGN MODEL

The recommendation system connects machine learning algorithms to facial expression detection for delivering custom music solutions with emotional appeal. It all starts with data collection: user share pictures of their facial expressions as well as their music preferences. Training a deep learning model requires the processing of this collected information so it can identify faces and emotional states. The central system component uses convolutional and fully connected layers for facial expressions analysis to match detected emotions between joy and sadness and excitement. After mood detection the system applies its collaborative filtering algorithm to suggest music that aligns with the current mood state of users thus improving their experience. The system accepts user commands through a user-friendly interface while receiving user feedback to develop its capabilities continuously. The system performance operates through standard evaluation metrics which include precision and recall to ensure proper system functioning. A user study evaluated the connection between recommended music and user emotional state and the user's satisfaction levels. The system can be updated with new recommendation methods or facial analysis models through its modular structure which allows easy system flexibility. More personalized interesting musical recommendations aim to boost user contentment alongside user participation rates.

7.RESULT AND ANALYSIS

Multiple scientific studies evaluated support vector machines (SVM) together with extreme learning machines (ELM) and convolutional neural networks [12]. Table 2 shows the comparison of comparable algorithms. Each research study contains both accuracy values as well as information about its matching approach. Using a convolutional neural network leads to increased accuracy of emotion recognition.

Table 2. Validation and Testing accuracy for the three algorithms on the Fer2013 Dataset.

Algorithm	SVM	ELM	CNN
Validation Accuracy	0.66	0.62	0.95
Testing Accuracy	0.66	0.63	0.71

The accuracy ratings reached respectively were 0.66 for both Validation and Testing while CNN stood as high as 0.71. Table 3 shows hyperparameters for the skilled CNN network. The weight update takes place during every batch through the learning rate control mechanism. The complete training dataset executes numerous epochs which subject the network during training time. Before weight alteration occurs the network shows a number of patterns equivalent to the batch size. Nonlinear prediction bounds become possible because of the integrated activation functions in the model. During deep learning model training Adam presents a suitable substitute to stochastic gradient descent as an optimization algorithm. The measurement of deep learning model mistakes depends on the categorical-crossentropy loss function in most single-label multi-class classification scenarios.

Table 3. Hyperparameter for trained CNN network.

Hyperparameters	Values
No of units	128
No. of Sessions	5
Optimizer	Adam
Rate of learning	0.001
Epoch	48
No. of Layers	28
Activation function	Relu, SoftMax
Loss function	Categorical-crossentropy

8.CONCLUSION

This study delivered a mood based music selector that aligns music suggestions to users based on their facial expressions and emotional state, using machine learning coordination based on their emotions. The data records were created by taking photographic evidence of the facial expressions and ratings of music preferences from participants. A deep learning model was developed with this data for recognizing emotions through facial expressions and then collaborated with a collaborative filtering algorithm in recommending suitable music according to the user's mood. The mean F1 score for accurately identifying the exact emotional state was 0.85, and feedback from the user survey showed that participants found the recommendations relevant and enjoyable. This study shows how joining face expression check with smart machines can make music suggestion systems better. By knowing a user’s feel in the moment, the system can give a more interesting and personalized listening experience. To sum up, our method gives a new and creative path to customizing music suggestions. In the future, the setup can be improved by adding user responses and situational elements to make its precision and alertness much better.

9.FUTURE SCOPE

The operational system is currently useful but additional improvements remain possible. The system can be modified in multiple ways to offer enhanced user results and smoother functionality. The alternative approach reflects methods using emotions which our system does not incorporate like disgust and dread. The feeling supported the automatic music playback feature. Future development of the system will contain mechanisms allowing Music therapists to help people with emotional strain, anxiety, sadness and agony. Future development possibilities exist for the system since its camera resolution is low and its performance deteriorates severely under poor lighting conditions.

11.REFERENCE

[1] Ramya Ramanathan, Radha Kumaran, Ram Rohan R, Rajat Gupta, and Vishalakshi Prabhu, an intelligent music player based on emo tion recognition, 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions. (2017)

[2]Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar, Smart music player integrating facial emotion recognition and music mood recommendation, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, (IEEE),2017.

[3]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.

[4] Ahlam Alrihail, Alaa Alsaedi, Kholood Albalawi, Liyakathunisa Syed, Music recommender system for users based on emotion detection through facial features, Department of Computer Science Taibah University, (DeSE), 2019.

[5]Research Prediction Competition, Challenges in representation learning: facial expression recognition challenges, Learn facial expres sion from an image, (KAGGLE).

[6]Preema J.S, Rajashree, Sahana M, Savitri H, Review on facial expression-based music player, International Journal of Engineering Re search & Technology (IJERT), ISSN-2278-0181, Volume 6, Issue 15, 2018.

[7]AYUSH Guidel, Birat Sapkota, Krishna Sapkota, Music recommendation by facial analysis, February 17, 2020.

[8]. Anagha S. Dhavalikar and Dr. R. K. Kulkarni, “Face Detection and Facial Expression Recognition System” 2014 International Conference on Electronics and Communication System (ICECS - 2014).

[9]. Yong-Hwan Lee, Woori Han and Youngseop Kim, “Emotional Recognition from Facial Expression Analysis using Bezier Curve Fitting” 16th International Conference on Network-Based Information Systems.(2013)

[10]. T.-H. Wang and J.-J.J. Lien, “Facial Expression Recognition System Based on Rigid and Non-Rigid Motion Separation and 3D PoseEstimation,” J. Pattern Recognition, vol. 42, no. 5, pp. 962-977, 2009.

[11] Mehendale N Facial emotion recognition using convolutional neural networks (FERC). SN Appl Sci.(2020)