International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

"Music Recommendation System Basedon Detection of Real-Time Facial Expressions"

Digvijay Shingare[1], Prof. Shweta Shah[2]

[1]Third Year Students, [2]Assistant Professor, Computer Engineering, Pune Institute of Computer Technology, Pune, India

***Abstract*— Music is used by people to elevate their mood. When anticipating a person's emotions and mood, the face is crucial. Positive mental health boosts creativity enhances ability to make decisions and promotes interpersonal connections. Additionally, being in a good frame of mind relieves stress, which can otherwise negatively impact our health and wellbeing. To create a user- friendly environment, it is intended to automate user-music player interactions.**

We provide a fresh method for automatically playing music while analysing facial expression. Most of the current methods include manually playing music, wearing wearable computers, or categorising based on auditory attributes. We suggest changing the manual sorting and playing instead. An internal camera is used to record facial expressions. On the input face photos, feature extraction is done to identify emotions including happiness, rage, sadness, surprise, and neutrality. A user's current sentiment is usedto suggest music accordingly. Compared to the existing algorithmin literature, it performs better in terms of computational timecomplexity.

In this case, the fundamental concept is that we are creating a prototype for a dynamic music recommendation system based on human emotions. The emotion on a real person's face is recognized using an integration of feature extraction and machine learning techniques. Once the mood is determined from the input image, the appropriate music will be played to keep the users' attention.

***Keywords*— Facial Expression and Emotion Recognition, CNN Model, Music Analysis, Image Processing, Machine Learning, Neural Networks, Haar cascade, Media pipe.**

I. Introduction

The human face is the most significant organ for an individual. With the help of it, people will more often communicate with one another while expressing and recognizing their own feelings. Music is one of the most important things that keeps people happy and at peace with themselves, and it is the art of living that makes people completely happy. Music is a fantastic technique for reducing tension and stress.

Any individual can benefit from music regardless of their mental state. With the use of the end user's input, our suggested system seeks to forecast the human emotion of eachunique user. In this case, the user will be prompted to choose between listening to music that will improve their mood or match their present mood before any songs are played on YouTube.

Our approach just considers the individual's Different moods, such as happy, sad, angry, neutral, or surprised.

This system offers an intelligent agent that will classify all different emotional states as distinct playlists that will eventually play music and suggest appropriate playlists to the user based on their current mood or of their choosing to improve their mood. This system can tell someone's mood by the look on his face. Complex neural networks are used to identify the facial expression (CNN). The device's camera is used to capture the collection of photographs, which are then provided to a trained CNN that can identify facial expressions and return them to the application. A song playlist is suggested in accordance with facial emotion. This is an enhancement to a music player's already-existing features. To solve this problem, the typical facial expression changes quickly and inconsistently, which may result in the erroneous playlist suggestion. When a program first launches, it gathers N images and captures. The most frequent face expression.

II. Problem Definition And Scope

It is widely accepted that music has a significant emotional impact on people. Studies have shown that rhythm itself is a powerful sedative. Most people, however, struggle with choosing songs, particularly those that reflect their current emotions. People will become even less inclined to search for the songs they want to listen to when they see the lengthy lists of unsorted music. Most users simply select songs at random from the song folder and play them using the music player. Most of the time, the songs that are played don't reflect the user's present mood and this method of searching for and choosing songs gets boring quickly. The person preferred to randomly select songs or simply press "play all" to play every song he owned.

III. Existing System

Even if programs like WYNK offer music, they are not based on mood; rather, they are based on previous music history, making the current system a manual one where one must choose songs independently.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

By making efforts to develop algorithms that can process naturally occurring human affective behavior, Hafeez Kabiniet al. [1] addressed the issue that existing methods typically handle only deliberately displayed and exaggerated expressions of prototypical emotions despite the fact that deliberate behavior differs in visual appearance, audio profile, and timing from spontaneously occurring behavior. They discussed crucial concerns such as the gathering and accessibility of training and test data, as well as the various ways that have been developed to address the issue of a machine comprehending human affective behavior.

The term "mind" has always drawn scientists to an all-encompassing study of it. Utilizing facial expressions is the most straightforward approach to convey emotions. Humans frequently convey our feelings in interpersonal relationships by using nonverbal indicators including hand gestures, facial expressions, and voice tones. Large playlist creation, management, and music selection take a lot of time and effort,according to Nikhil Zaware et al [2].

The work by Henal Shah et al. [4] explains how sentimental or emotion analysis could be used to create an intelligent music player. Emotions in people are intended to promote understanding and the sharing of sentiments and goals. Both verbal and facial expressions serve to convey emotions. Writing down one's feelings is another way to communicate them. This paper focuses primarily on the methodologies available for detecting human emotions for developing emotion-based music players, the methodologies used by existing music players to detect emotions, the methodology a music player follows to detect human emotions, and how it is preferable to use the proposed system for emotion detection. Additionally, it provides a quick overview of the operation of our systems, playlist creation, and emotion classification.

IV. Disadvantages

Ignoring the mental state of the individual, the same collection of songs will be recommended, even if the person isin a different mood. We list and describe a few of the major issues that, in our opinion, the study area of music recommender systems is now dealing with, including how to solve the cold start problem, continue an automatic playlist, and appropriately assess music recommender systems.

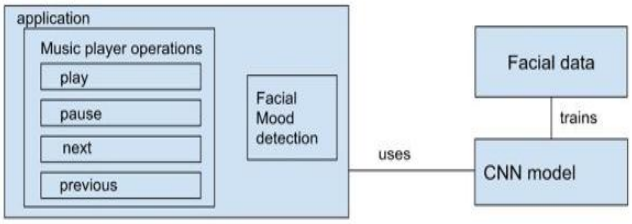
V. Proposed System

Music recommendations depending on the user's expression. Use deep learning principles to infer a person's mood from their facial expression.

This would be an upgrade to the current music system since mood is now considered when identifying songs.

The suggested method aims to give the user an engaging experience when generating a playlist. The working relies on various systems doing their duties in a specific order to get the intended results. The following is a statement of the working:

1. The suggested system functions by first offering a user interface that is clear and basic enough
2. The system classifies the input and creates an output that is an emotion (mood) based on the expression retrieved from The real-time graphical input after first checking for the presence of a face using the face detection procedure.
3. The user camera is used in the following step with the appropriate permissions, and a real-time facial expression input (picture) is given to the system.
4. The songs from the playlists are then played using the categorized expression as an input to choose an acceptable playlist from the ones that were initially prepared.



SYSTEM ARCHITECTURE

Fig. 1 System Architecture

VI. Methodology

A. CNN (Convolutional neural network) :

Convolutional neural network (CNN) model is employed in this instance for both mood detection and image processing. A CNN is a particular type of network design for deep learning algorithms that is utilized for tasks like image recognition and pixel data processing. Although there are different kinds of neural networks in deep learning, CNNs are the preferred network architecture for identifying and recognizing objects. For recognizing emotion intensity, SVM, RF, and KNN, three well-known ML techniques, were applied. A convolutional neural network separates facial recognition into the following key elements:

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

- Identify each face in the photo
- Despite other influences like light, angle, posture, etc.,concentrating on each face.
- Finding distinct characteristics
- Matching a face with a name by comparing all the acquired data with data that has already been stored inthe database.

One of the most crucial algorithms for image classification and picture extraction, with convolutional neural network (CNN) techniques, each neuron just has to feel the local portion of the image rather than the entire image. The benefit of the CNN algorithm is that it learns intuitively from training data rather than explicitly extracting features. Network topology and input data may go well together. It offers special benefits for picture processing.

Convolutional neural networks, which are used frequently in face recognition, object identification, and scene labelling, among other tasks, take an image as input and classify and process it according to many categories, including voice, image, and object, among others. Depending on the image's resolution, the system perceives an image as an array of pixels.

Convolution Layer: This layer's primary objective is to extract features from images. A small portion is picked from a large image and is drawn through every point in the picture. The data are complicated at every moment in time when it is passed, all at once. Each tiny portion of this larger image thatit incorporates is referred to as a kernel or filter. The output image's map is created by this activation function.

Pooling Layer: By demodulating the image using these layer sub sampling techniques, each activation function's dimension is also decreased. However, it retains most of the relevant data. This layer's primary goal is to increase robustness, generalisation, and convergence quickly.

B. MediaPipe:

Google created a cross-platform library called Media pipe that offers incredible pre-built machine learning (ML) solutions for computer vision problems. OpenCV library in python is a computer vision library that is widely used for image analysis, image processing, detection, recognition, etc.

It was initially created for the analysis of YouTube video and audio in real-time. It was gradually included into many additional products, such as the Netcam Perception system and the Object identification by Google Lens.

C. NumPy:

Working with arrays is made possible by the Python package NumPy. Comprehensive mathematical functions, random number generators, linear algebra functions, Fourier transforms, and other features are all available in NumPy. Numerical Python is referred to as NumPy.

In contrast to lists, NumPy arrays are stored in a single continuous location in memory, making it very easy for programs to access and manipulate them. In computer science,this characteristic is known as locality of reference. This is the primary factor that makes NumPy faster than lists.

A Python package called OpenCV makes it possible to carry out image processing and computer vision tasks. It offers a variety of capabilities, such as tracking, face recognition, and object detection. The OpenCV Python Tutorial will teach us more about the library.

D. The Haar cascade:

Haar cascade is an algorithm that can detect objects in images, irrespective of their scale in image and location.

This algorithm is not so complex and can run in real-time. We can train a haar-cascade detector to detect various objects like cars,bikes, buildings, fruits, faces etc.

Haar cascade uses the cascading window, and it tries to compute features in every window and classify whether it could be an object. Sample haar features traverse window-sized across the picture to compute and match features.

Haar cascade works as a classifier. It classifies positive data points → that are part of our detected object and negative data points → that don't contain our object. Human face detection is the first example of object detection using a haar cascade, where we will detect human faces from a picture using a pre- trained haar cascade.

VII. ALGORITHM

Emotion Detection from Face and suggesting music using the Mediapipe library and CNN Algorithm:

Face emotions recognition is also one of the widely applied use cases of face detection with the help of the Media pipe library. Convolutional neural network (CNN) model is employed in this instance for both mood detection and image processing. It detects the features of the face i.e. eyes, nose, and mouth though we will also try to detect some landmarks here as well this is not the optimal way we will do that only because the MediaPipes face detection algorithm offers us the same.

1. The very first step will be to import all the necessary libraries. i.e. NumPy, MediaPipe, Matplotlib, CNN.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

2. The very first step will be to initialize the MediaPipe's face detection model.
3. After initializing the model we will call the face detection function by using the relevant parameters and their values.
4. At the last it is also necessary to see the results for that we will use the drawing_utils function to draw the resultson the image/frames.
5. As discussed here we have read the image after providing its path in the function.

Now, we will use the matplotlib function to plot the image as working with Jupyter notebook so cv2's show function will not work in this environment hence with Matplotlib's show function we will plot the image before that we will also usethe figure size with the help of figure function of matplotlib.

Step 1: Using the user's webcam, the system receives inputfrom the user.

Step 2: Our algorithm analyses the image and categorises it as either representing a happy, sad, neutral, or furious emotion.

Step 3: Using the end user's data, the trained datasets areretrieved and detected.

Step 4: The user's present mood or an improved version of itis reflected in the playlist or songs that are played.

Step 5: Following the successful sentiment identification, music is played to improve the user's mood.

Music recommendations depending on the user's expression:

- A music player that makes playlist recommendations depending on the user's expression.
- Uses deep learning concepts to detect one's mood based on facial expression
- This would be an improvement to the present music system as mood is also added as a factor of detecting song.

System will suggest a playlist for tracks based on the mood. Seven different facial expressions were included in the batch of data used to train the intricate neural network. The Haar Cascade and CNN are used to extract faces from photographs from the data collection, and these images are then given to the neural network. Since the neural network was trained on face-extracting pictures.

Give the Haar Cascade network the facial images from the photos the camera took as input for prediction. The figure below shows the system architecture of the application.

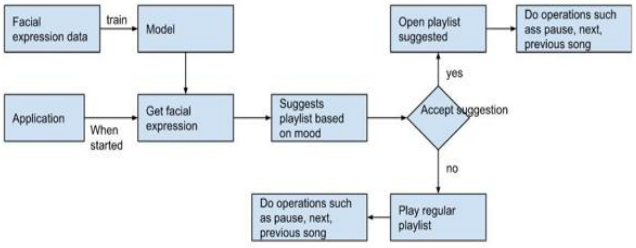


Fig. 2 Flow Chart for System

VIII. Future Scope

We can gradually introduce learning into the application so that it gains knowledge from fresh data produced by the application. Based on user response, the programme will determine whether it made a correct prediction. The approach described above improves model quality and boosts model accuracy. To further improve the model's accuracy, we may also include new features like heart rate, which is partly correlated with emotional states in people. To anticipate emotions more accurately than the prior method, we can also take the environment into account. For instance, if we are in a gym, the program needs to recognize gym-appropriate objects and play upbeat music to get us in the right frame of mind.

IX. Conclusions

This program can be added to modern advanced music players as an extra function that recommends songs based on real time facial expression with more accuracy. The frequencywith which the system suggests the song he needs would increase with the addition of a facial expression detecting system in the music player, increasing user happiness. This facial recognition technology can also be applied in a variety of other contexts, such as suggesting movies and activities.

REFERENCES

[1] "Emotion Based Music Player" by Hafeez Kabini, Sharik Khan, Omar Khan, and Shabana Tadvi, International Journal of Engineering Research and General Science, Volume 3, Issue 1, 2015.

[2] "Emotion Based Music Player" by Nikhil Zaware, Tejas Rajgure, Amey Bhadang, and D.D. Sakpal, published in International Journal of Innovative Research & Development, Volume 3, Issue 3, 2014.

[3] Setiawardhana, Peni Rahayu, and Nana Ramadijanti "ELECTIONS TO THE MUSIC PLAYLIST USING FACIAL EXPRESSIONS RECOGNITION USING BACKPROPAGATION NEURAL NETWORK" 2012, Issue 3 of Volume 6 of Jurnal Ilmiah Kursor.

[4] International Journal of Innovative and Emerging Research in Engineering, Volume 2, Issue 4, Henal Shah, Tejas Magar, Purav Shah,and Kailas Devadkar, "AN INTELLIGENT MUSIC PLAYER USING SENTIMENTAL

[5] Emotion Review, vol. 7, no. 2, pp. 189–197, April 2015. Swathi Swaminathan and E. Glenn Schellenberg, "Current emotion research inmusic psychology."

[6] Examined Existence, [online] "How music alters your mood." It's possible to change your mood using music at http://examinedexistence.com/.

[7] Byeong- jun Han, Seungmin Rho, RogerB. Dannenberg, Eenjun Hwang " SMERS MUSIC EMOTION RECOGNITION USING SUPPORT VECTOR Retrogression " 10th International Society for Music Information Retrieval Conference(ISMIR 2009).

[8] Benoit Mathieu, Slim Essid, Thomas Fillon, Jacques Prado, Ga ¨ el Richard " YAAFE, AN EASY TO USE AND Effective AUDIO point birth SOFTWARE " Institut Telecom, Telecom ParisTech, CNRS/ LTCI.

[9] Mary Duenwald (2005). The science behind facial expressions. on October9, 2012, was taken from http://discovermagazine.com/2005/jan/physiology-of-facial-expressions

[10] Jacob C. Hager (2003). Introduction to the DataFace website: Physiognomy, Nonverbal Communication, Emotion Expressions, and Facial Expressions. From http://face-andemotion.com/dataface/general/homepage.jsp, retrieved on October 10, 2012.

[11] Facial Expression Detection and Recognition System, W.K. Teo, Liyanage C. De Silva, and Prahlad Vadakkepat, Journal of The Institution of Engineers, Singapore, Vol. 44, 2004.

[12] "Sentimental analysis by machine learning tool for generalized solution on twitter"by s shah ,Ashish Soni IJRCET ,VOLUME 7 ISSUE 3 April 2018.

[13] "A Study on Neural Networks used for Traffic Signal Classification",by K Gangrade, S Shah ,IRE Journals Volume 6 Issue 7 | ISSN: 2456-8880.