Machine learning algorithms for analysing music genre recommendations based on emotional factors

[1]Dr. S. Krishna Mohan Rao, [2]T. Nageshwar Rao, [3]Vinodkumar Bijja Hosea, [4]Jatin Rao Bandaru
[1]Professor, [2,3]Assistant Professor, [4]Student, [1,2,3,4]Department of Computer Science Engineering , Siddhartha Institute of Engineering and Technology, Hyderabad, India.

ABSTRACT

Music plays a very important role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners and sometimes even imparts a therapeutic approach. In today's world, with ever increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed (seek & time compression) local playback, streaming playback with multicast streams and including volume modulation, genre classification etc. We propose a new approach for playing music automatically using facial emotion. Most of the existing approaches involve playing music manually, using wearable computing devices, or classifying based on audio features. Instead, we propose to change the manual sorting and playing. We have used a Convolutional Neural Network for emotion detection. For music recommendations, Tkinter is used. Our proposed system tends to reduce the computational time involved in obtaining the results and the overall cost of the designed system, thereby increasing the system's overall accuracy. Facial expressions are captured using an inbuilt camera. Feature extraction is performed on input face images to detect emotions such as happy, angry, sad, surprise, and neutral. Automatically music playlist is generated by identifying the current emotion of the user. It yields better performance in terms of computational time, as compared to the algorithm in the existing literature.

INTRODUCTION

1.1 PROBLEM STATEMENT

Music plays a very important role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners and sometimes even imparts a therapeutic approach. In today's world, with ever increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed (seek & time compression) local playback, streaming playback with multicast streams and including volume modulation, genre classification etc. The motivation of this work comes from the possibility of reducing the human effort in creating music playlists manually, thus generating them automatically based on the user's emotional state. The human face plays an important role in knowing an individual's mood. The required input is extracted from the human face directly using a camera. One of the applications of this input can be for extracting the information to deduce the mood of an individual. This data can then be used to get a list of songs that comply with the "mood" derived from the input provided earlier. This eliminates the time-consuming and tedious task of manually Segregating or grouping songs into different lists and helps in generating an appropriate playlist based on an individual's emotional features.

In old-style music players, a user had to manually browse through the playlist and select songs that would soothe his mood. In today's world, with ever increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed, local playback, streaming playback with multicast streams and including volume modulation, genre classification etc. These features may satisfy the user's basic requirements, but the user has to face the task of manually browsing through the playlist of songs and select songs based on the current mood and behavior. That is the requirement of an individual, a user sporadically suffered through the need and desire of browsing through his playlist, according to his mood and emotions.

Music is often considered to be voice of the soul as it makes people emote their feelings no matter what the situation is. An angry person tries to calm himself by listening to music which might calm his nerves. A sad person listens to motivating song which helps him to come out of the depression phase. Music and emotion coexist.

a. Accurately detect the mood of the person
b. To create a playlist according to the identified emotion by using a real time dataset.
c. Real time dataset allows us to capture the person’s image at the particular instant based on which songs can be suggested which complies with his mood.

EXISTING SYSTEM

For Automatic Facial Expression recognition this research paper [1] uses three phases. These three phases are 1. Face detection2. Feature Extraction and 3. Expression recognition. In the First Phase, YCbCr Colour model are used for face detection, lighting compensation for obtaining face and morphological operations for holding required features of the face i.e. eyes, eyebrows and mouth. This System also uses Active Appearance Model Method (AAM) for facial feature extraction. In this method the features on the face like eye, eyebrows and mouth are located and a data file is created which gives information about the model points detected. Different facial expressions are given as input to the AAM Model which changes according to expression.

Three different ways are used in this paper [2] for emotion classification and context-based music recommendation. They are 1. Emotion State Transition Model (ESTM) 2. Context-based music recommendation (COMUS) 3. Nonnegative matrix factorization (NMF). ESTM is predominantly used to model various human emotions and their transition to music. It acts like a bridge between an individual’s mood and low-level music features. With the help of ESTM the most legitimate music can be recommended to the client for travelling to the ideal state. COMUS ontology is utilized for demonstrating user's musical inclinations and setting, and for supporting thinking about the client's ideal feeling and inclinations. COMUS are a music dedicated ontology developed by including particular classes for music suggestion which incorporates mood, situation and other features. In order to reduce the dimensions data related to music are gathered after which NMF are applied to map them to ESTM.

PROPOSED SYSTEM

We use facial expressions to propose a recommender system for emotion recognition that can detect user emotions and suggest a list of appropriate songs.

The proposed system detects the emotions of a person, if the person has a negative emotion, then a certain playlist will be shown that includes the most related types of music that will enhance his mood. And if the emotion is positive, a specific playlist will be presented which contains different types of music that will inflate the positive emotions. The dataset we used for emotion detection is from Kaggle Facial Expression Recognition. Dataset for the music player has been created from Bollywood Hindi songs. Implementation of facial emotion detection is performed using Convolutional Neural Network which gives approximately 95.14% of accuracy.

MODULES

Upload Image

In the first step of our emotion-based music recommendation system, the user uploads an image that represents their current emotional state. The image could be a photograph of themselves or any other visual representation that reflects their emotions. This image serves as input to our system, allowing us to analyze and interpret the user's emotional state.

Preprocessing

Once the image is uploaded, we proceed to preprocess it to extract relevant information for emotion recognition. Preprocessing may involve resizing the image, normalizing the color channels, and removing any unnecessary background or noise. We apply image processing techniques to enhance the quality and clarity of the image, ensuring that it provides a reliable representation of the user's emotions.

Training:

In this step, we train a machine learning model using a dataset of labeled facial expressions or visual cues associated with different emotions. The training dataset consists of images with known emotional labels, allowing the model to learn the patterns and correlations between facial features and specific emotional states. Through the training process, the model becomes capable of recognizing and classifying emotions based on facial expressions in new, unseen images.

Emotion Recognition

Using the trained model, we apply emotion recognition algorithms to the preprocessed image. The model analyzes the facial features, such as eye movements, eyebrow position, mouth shape, and overall expression, to infer the user's emotional state. The output of this step is a prediction or classification of the user's emotion, which could include categories such as happiness, sadness, anger, surprise, or neutral.

Play music:

Based on the recognized emotion, we proceed to recommend and play music that aligns with the user's emotional state. We have a database of music tracks tagged with corresponding emotional labels. Using the predicted emotion from Step 4, we select and play music that matches the user's emotional state. For example, if the user's emotion is recognized as sadness, we may recommend calming or soothing music. Conversely, if the user is predicted to be happy, we may suggest upbeat and energetic songs. The recommended music provides a tailored and immersive experience that resonates with the user's emotional needs and preferences.

ALGORITHM

Working of CNN

Generally, a Convolutional Neural Network has three layers, which are as follows:

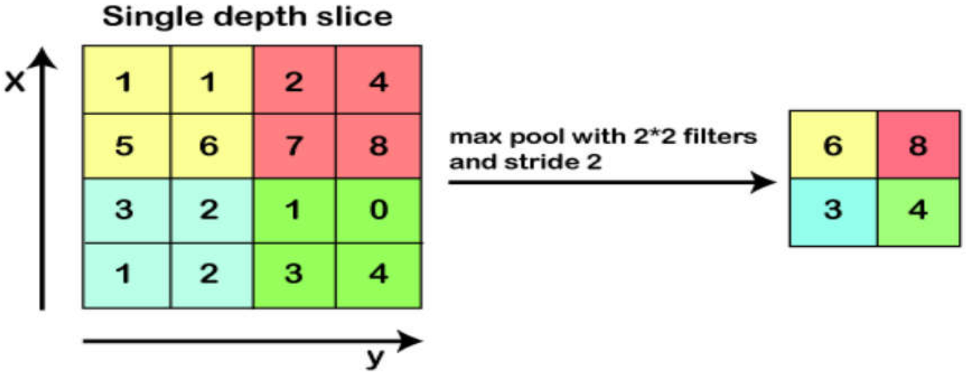
Input: If the image consists of 32 widths, 32 height encompassing three R, G, B channels, then it will hold the raw pixel values of an image.

Convolution: It computes the output of those neurons, which are associated with input's local regions, such that each neuron will calculate a dot product in between weights and a small region to which they are actually linked to in the input volume.

ReLU Layer: It is specially used to apply an activation function elementwise, like as max (0, x) thresholding at zero. It results in which relates to an unchanged size of the volume.

Pooling: This layer is used to perform a down sampling operation along the spatial dimensions (width, height) that results in [16x16x12] volume.

Flattening: It converts 2D array into 1D array.

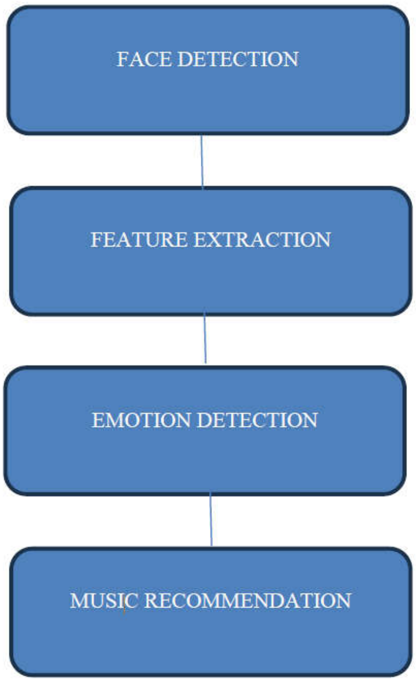


1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

6	8
3	4

METHODOLOGY

We built the Convolutional Neural Network model using the Kaggle dataset. The database is split into two parts training and testing dataset. The training dataset consists of 24176 and the testing dataset contains 6043 images. There are 48x48 pixel grayscale images of faces in the dataset. Each image in dataset is labeled as one of five emotions: happy, sad, angry, surprise, and neutral. The faces are automatically registered so that they are more or less centered in each image and take up about the same amount of space. The images in dataset contain both posed and unposed headshots, which are in grayscale and 48x48 pixels. The dataset was created by gathering the results of a Google image search of every emotion and synonyms of the emotions. Dataset being trained on an imbalanced dataset may perform well on dominant emotions such as happy, sad, angry, neutral, and surprised but they perform poorly on the under -represented ones like disgust and fear. Usually, the weighted-SoftMax loss approach is used to handle this problem by weighting the loss term for each emotion class supported by its relative proportion within the training set. However, this weighted-loss approach is predicated on the SoftMax loss function, which is reported to easily force features of various classes to stay apart without listening to intra-class compactness. One effective strategy to deal with the matter of SoftMax loss is to use an auxiliary loss to coach the neural network. To treating missing and Outlier values we have used a loss function named categorical crossentropy. For each iteration, a selected loss function is employed to gauge the error value. So, to treating missing and Outlier values, we have used a loss function named categorical crossentropy.



Face Detection

Face detection is one of the applications which is considered under computer vision technology. This is the process in which algorithms are developed and trained to properly locate faces or objects in object detection or related system images. This detection can be real-time from a video frame or images. Face detection uses such classifiers, which are algorithms that detect what's either a face (1) or not a face (0) in an image. Classifiers are trained to detect faces using numbers of images to get more accuracy. OpenCV uses two sorts of classifiers, LBP (Local Binary Pattern) and Haar Cascades. A Haar classifier is used for face 11 detection where theclassifier is trained with pre -defined varying face data which enables it to detect different faces accurately. The main aim of face detection is to spot the face within the frame by reducing external noises and other factors. It is a machine learning-based approach where the cascade function is trained with a group of input files. It is supported the Haar Wavelet technique to research pixels inside the image into squares by function. This uses machine learning techniques to urge a high degree of accuracy from what's called "training data".

B. Feature Extraction

While performing feature extraction, we treat the pre-trained network that is a sequential model as an arbitrary feature extractor. Allowing the input image to pass on it forward, stopping at the pre-specified layer, and taking the outputs of that layer as our features. Starting layers of a convolutional network extract high-level features from the taken image, so use only a few filters. As we make further deeper layers, we increase the number of the filters to twice or thrice the dimension of the filter of the previous layer. Filters of the deeper layers gain more features but are computationally very intensive. Doing this we utilized the robust, discriminative features learned by the Convolution neural network. The outputs of the model are going to be feature maps, which are an intermediate representation for all layers after the very first layer. Load the input image for which we want to view the Feature map to know which features were prominent to classify the image. Feature maps are obtained by applying Filters or Feature detectors to the input image or the feature map output of the prior layers. Feature map visualization will provide insight into the interior representations for specific input for each of the Convolutional layers within the model.

C. Emotion Detection

Convolution neural network Architecture. Convolution neural network architecture applies filters or feature detectors to the input image to get the feature maps or activation maps using the Relu activation function. Feature detectors or filters help in identifying various features present in the image such as edges, vertical lines, horizontal lines, bends, etc. After that pooling is applied over the feature maps for invariance to translation. Pooling is predicted on the concept that once we change the input by a touch amount, the pooled outputs don't change. We can use any of the pooling from min, average, or max. But max-pooling provides better performance than min oraverage pooling. Flatten all the input and giving these flattened inputs to a deep neural network which are outputs to the class of the object. The class of the image will be binary, or it will be a multi-class classification for identifying digits or separating various apparel items. Neural networks are as a black box, and learned features in a Neural Network are not interpretable. So basically, we give an input image then the CNN model returns the results. Emotion detection is performed by loading the model which is trained by weights using CNN. When we take the real-time image by a user then that image was sent to the pre-trained CNN model, then predict the emotion and adds the label to the image.

D. Music Recommendation Module

Songs Database We created a database for Bollywood Hindi songs. It consists of 100 to 150 songs per emotion. As we all know music is undoubtedly involved in enhancing our mood. So, suppose a user is sad then the system will recommend such a music playlist which motivates him or her and by this automatic mood will be delighted. Music Playlist Recommendation By using the emotion module real-time emotion of the user is detected. This will give the labels like Happy, Sad, Angry, Surprise, and Neutral.

RESULTS

The project "Emotion-Based Music Recommendation" follows a systematic approach to provide personalized music recommendations based on users' emotions. The process begins with capturing or accessing an image, such as a photo or video frame. This image serves as the input for the subsequent steps. In the preprocessing stage, the image is prepared for analysis by applying various techniques like resizing, normalization, and noise reduction. These steps ensure that the image is in a suitable format for emotion recognition.

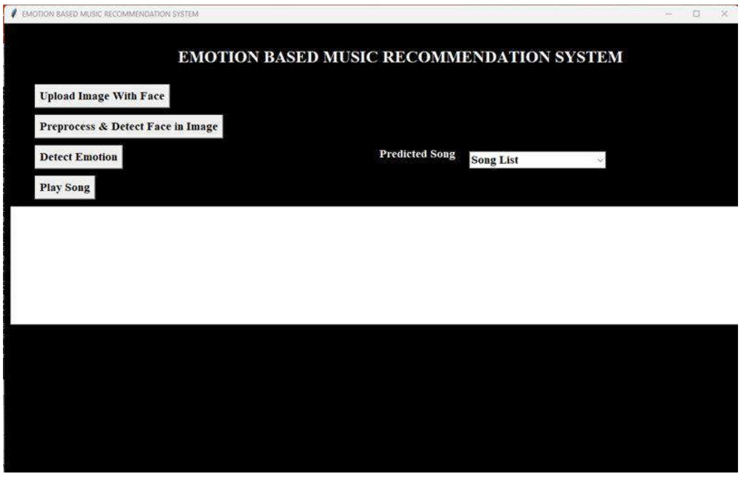
Once the image is preprocessed, the next step in the project is training. This involves using machine learning algorithms to analyze a dataset of images labeled with corresponding emotions. The model is trained to recognize patterns and features indicative of different emotional states. Through an iterative process, the model learns to associate visual cues with specific emotions, enabling it to make accurate predictions.

After the training phase, the project moves on to emotion recognition. The trained model is applied to the preprocessed image to detect and classify the predominant emotion present. This could involve identifying emotions such as happiness, sadness, anger, or surprise based on facial expressions, body language, or other visual cues. The recognition algorithm assigns a confidence score or probability to each detected emotion.

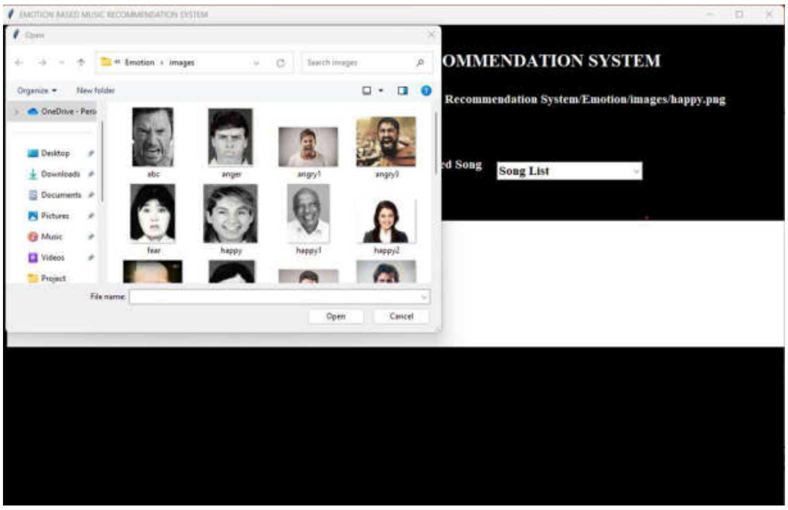
Once the emotion is recognized, the final step is to recommend and play suitable music based on the identified emotion. A music recommendation system suggests songs or playlists that align with the detected emotion. For example, if the emotion recognized is happiness, the system might suggest upbeat and cheerful songs. The recommended music is then played for the user to enjoy and enhance their emotional experience.

Overall, the "Emotion-Based Music Recommendation" project leverages image analysis, machine learning, and music recommendation techniques to provide personalized music suggestions based on users' emotions, creating a more engaging and tailored music listening experience.

Double click on 'run.bat' file to get below screen

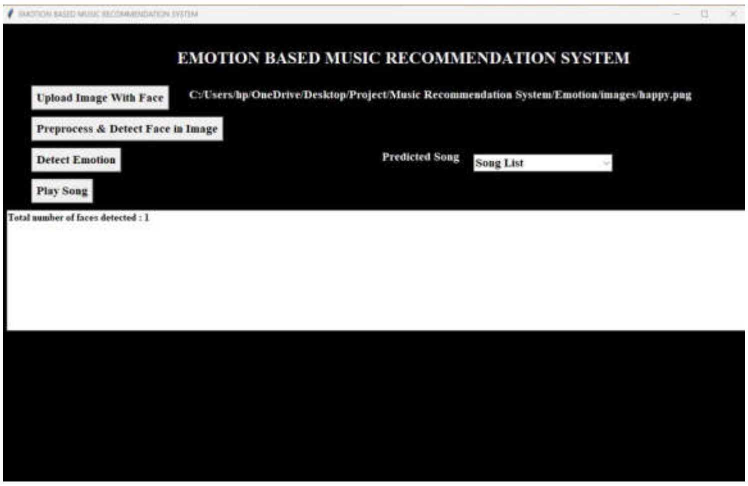

EMOTION BASED MUSIC RECOMMENDATION SYSTEM
EMOTION BASED MUSIC RECOMMENDATION SYSTEM
Upload Image With Face
Preprocess & Detect Face in Image
Detect Emotion
Play Song
Predicted Song
Song List

In above screen click on 'Upload Image With Face' button to upload image



Recommendation System/Emotion/images/happy.png

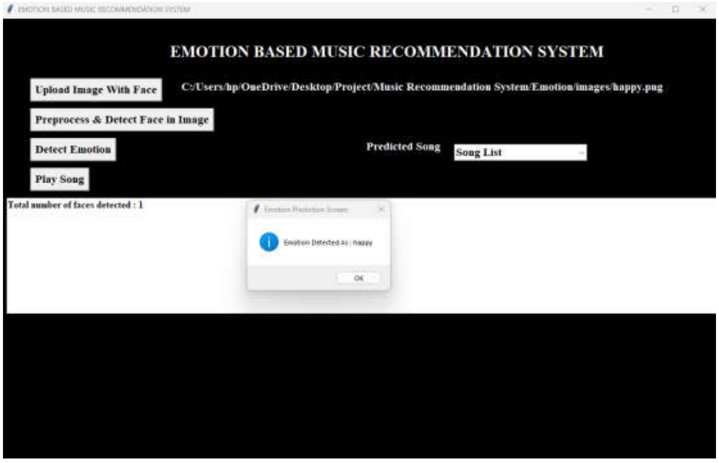
Double click on 'run.bat' file to get below screen In above screen click on 'Upload Image With Face' button to upload image


EMOTION BASED MUSIC RECOMMENDATION SYSTEM
EMOTION BASED MUSIC RECOMMENDATION SYSTEM
Upload Image With Face
Preprocess & Detect Face in Image
Detect Emotion
Play Song
Predicted Song
Song List

C:/Users/hp/OneDrive/Desktop/Project/Music Recommendation System/Emotion/images/happy.png

Total number of faces detected : 1

In above screen we can see in uploaded image one face is detected. Now click on Detect Emotion button to detect emotion


EMOTION BASED MUSIC RECOMMENDATION SYSTEM
EMOTION BASED MUSIC RECOMMENDATION SYSTEM
Upload Image With Face
Preprocess & Detect Face in Image
Detect Emotion
Play Song
Predicted Song
Emotion Prediction Screen
OK

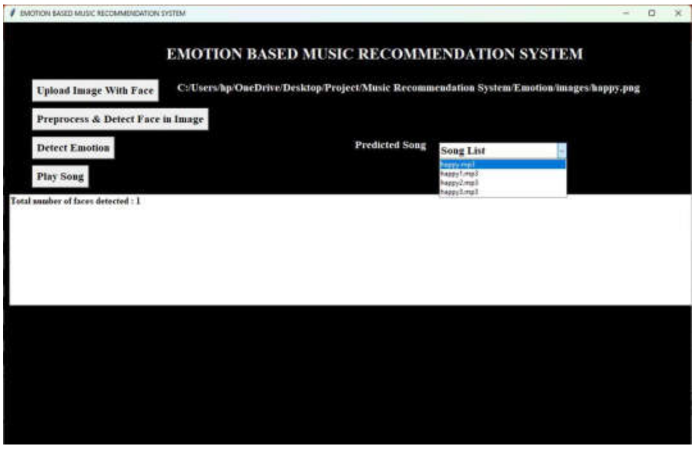
C:/Users/hp/OneDrive/Desktop/Project/Music Recommendation System/Emotion/images/happy.png

Song List

Total number of faces detected : 1

Emotion Detected As : happy

In above screen we can see emotion happy is detected and now click on drop down arrow link to get all disgust songs list


EMOTION BASED MUSIC RECOMMENDATION SYSTEM
EMOTION BASED MUSIC RECOMMENDATION SYSTEM
Upload Image With Face
Preprocess & Detect Face in Image
Detect Emotion
Play Song
Predicted Song

C:/Users/hp/OneDrive/Desktop/Project/Music Recommendation System/Emotion/images/happy.png

Song List

- happy.mp3
- happy1.mp3
- happy2.mp3
- happy3.mp3

Total number of faces detected : 1

In drop down box we can see 'happy.mp3' songs is showing, select that song and click on 'Play Song' button to play song.

CONCLUSION

Since music has the power to emote user's feelings a generic model is implemented to recommend music based on the user emotions. Human emotions play an important in expressing the thought of an individual. The main goal of the system is to detect changes in the emotional state of the user and play music according to the user's preferences by exploring various music tracks. The system uses CNN algorithm for emotion classification which can be determined by change in shape, size and movement of eyebrows, eyes and mouth. They fall into one of the six basic types of emotions which are sadness, happiness, anger, fear, disgust and surprise based on which a playlist is generated. The main reason for using CNN algorithm over SVM is its ability to recognize the most important features in an image without any help from humans. Also, SVM's prediction accuracy is found to be less when compared to CNN's accuracy. The proposed system has delivered results with significant accuracy. Since human emotions are not consistent and they are actually a result of internal and external circumstances happening around an individual it is difficult to get 100% accuracy. But with better algorithm and intense research a perfect emotion-based music recommendation system can be developed. The proposed system is tested against a web camera. The total cost involved in implementing this project is almost negligible. Average estimated time for various modules of proposed system.

FUTURE SCOPE

The future scope for an emotion-based music recommendation project is promising, as it aligns with the growing demand for personalized and context-aware music experiences. Here are some potential directions for expanding and enhancing a project:

REFERENCES

[1] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632. https://doi.org/10.31979/etd.w5fs-s8wd

[2] Hemanth P,Adarsh ,Aswani C.B, Ajith P, Veena A Kumar, “EMO PLAYER: Emotion Based Music Player”, International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 4, April 2018, pp. 4822-87.

[3] Music Recommendation System: “Sound Tree”, Dcengo Unchained: Sıla KAYA, BSc.; Duygu KABAKCI, BSc.; Işınsu KATIRCIOĞLU, BSc. and Koray KOCAKAYA BSc. Assistant : Dilek Önal Supervisors: Prof. Dr. İsmail Hakkı Toroslu, Prof. Dr. Veysi İşler Sponsor Company: ARGEDOR

[4] Tim Spittle, lucyd, GitHub, , April 16, 2020. Accessed on: [Online], Available at: https://github.com/timspit/lucyd

[5] A. 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, vol. 8, no. 7, p. 1103, Jul. 2018.