2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)
March 9-10, 2023, Amity University Dubai, UAE

Emotion Based Music Recommendation System

Micah Mariam Joseph
Department of Engineering
Amity University
Dubai, United Arab Emirates
micahJ@amitydubai.ae

Diya Treessa Varghese
Department of Engineering
Amity University
Dubai, United Arab Emirates
diyaV@amitydubai.ae

Lipsa Sadath
Department of Engineering
Amity University
Dubai, United Arab Emirates
lsadath@amityuniversity.ae

Ved Prakash Mishra
Department of Engineering
Amity University
Dubai, United Arab Emirates
vmishra@amityuniversity.ae

2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) | 979-8-3503-3826-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCIKE58312.2023.10131874

***Abstract*— A person often finds it difficult to choose which music to listen from different collections of music. Depending on the user's mood, a variety of suggestion frameworks have been made available for topics including music, festivals and celebrations. Our music recommendation system's main goal is to offer users recommendations that match user's preferences. Understanding the user's present facial expression can enable us to predict the user's emotional state. Humans frequently use their facial expressions to convey their intentions. More than 60% of users have at some point believed that the count of songs in their music playlist is so much that the user's are unable to choose a music to play. By creating a suggestion system, it might help a user decide which music to listen, allowing the user to feel less stressed. This work is a study on how to track and match the user's mood using face detection mechanism, saving the user time from having to search or look up music. Deep learning performs emotion detection which is a well-known model in facial recognition arena. Convolution neural network algorithm has been used for the facial recognition. We use an open-source app framework known as Streamlit to make a web application from the model. The user will then be shown songs that match his or her mood. We capture the user's expression using a webcam. An appropriate music is then played on, according to their mood or emotion.**

Keywords— emotion recognition, convolution neural network, streamlit.

I. INTRODUCTION

Nowadays, music services make vast amounts of music easily accessible. People are constantly attempting to enhance music arrangement and search management, in order to alleviate the difficulty of selection and make discovering new music works easier. Recommendation systems are becoming increasingly common, allowing users to choose acceptable music for any circumstance. Recommendations for music can be used in a range of situations, including music therapy, sports, studying, relaxing, and supporting mental and physical activity. [1] However, in terms of personalization and emotion driven recommendations, there is still a gap. Humans have been massively influenced by music. Music is a key factor to for various effects on humans such as controlling their mood, relaxation, mental and physical work and help in stress relief. Music therapy can be used in a variety of clinical contexts and practices to help people feel better. In this project, we're creating a web application that recommends music based on emotions. It influences how people live and interact with one another. At times, this could seem that we have been controlled by our emotion. The emotion we are encountering at any given moment have an effect on the final decision that we choose, actions that we undertake, and the impression that we form. Neutral, angry, disgust, fear, glad, sad, and surprise are the seven primary global emotions. The look on a person's face might reveal these basic emotions. This study presents a method for detecting these basic universal emotions from frontal facial expressions. After, implementing the facial recognition machine learning model, we then further continue to make it into a web application by using Streamlit. The Emotion detection is performed using Deep learning. Deep Learning is a well-known model in the pattern recognition arena. The keras library is being used, as well as the Convolution Neural Network (CNN) algorithm. A CNN is indeed an artificial neural network with some machine learning component. Among other things, CNN can also be used to detect objects, perform facial recognition and process images. [2]

II. LITERATURE REVIEW

Humans frequently convey their emotions through a variety of ways like hand gestures, voice, tonality and so on, but they mostly do through facial expressions.An expert would be able to determine the emotions being experienced by the other person by observing or examining them. Nevertheless, as there is technological advancement in today's world, machine are attempting to become more smarter. Machines are aiming to operate in an increasingly human-like way. On training the computer on the human emotions, the machine would be capable to perform analysis and react like a human. By enabling precise expression patterns with improved competence and error-free emotion calculation, data mining can assist machines in discovering and acting more like humans. A music player which is dependent on emotions takes less time to find the appropriate music that the user can resonate with. People typically have a lot of music on their playlist, this would make it difficult for the user to choose an appropriate song. Random music does not make the user feel better, so with the aid of this technology, users can have songs played automatically based on their mood. [3] The webcam records the user's image, and the pictures are stored. The system recorded user's varied expressions to assess their emotions and select the apt music.

The ability to read a person's mood from their expression is important. To capture the facial expressions, a webcam is used. This input can be used, among other things, to extract data that can be used to infer a person's attitude. Songs are generated using the "emotion" which has been inferred from the previous input. This reduces the tedious job of manually classifying songs into various lists. The Facial Expression Based Music Recommender's main objective is scanning and analyzing the data, and then it would suggest music in line with the user's mood. [4]

By utilizing image processing, we have developed an emotion-based music system that would allow the user to

create and manage the playlist with less effort even while delivering the best song for the user's current expression, giving the listeners an outstanding experience. [5] The way a person is feeling can be inferred from their facial movements. The image of the person is captured using a webcam, and information is then extracted from the image.

Emotion detection research is influenced by a range of fields, including machine learning, natural language processing, neurology and others. They investigated for several universal expressions of emotions in the face expressions, voice feature and textual data in previous study. Some of the categorization for emotions include Happy, Sad, Disgust, Fury, Fear and Surprise.
Later on, the image, audio, and textual data are combined to better the work. The combination of these data yields the most accurate result. [6]

A. Convolution Neural Networks

In recent years, the development of Convolution neural networks has had a considerable impact on the computer vision sector, as well as a substantial step and ability to recognize objects[6]. Neural networks include CNNs. Neural Networks, a subtype of machine learning, are the backbone of deep learning approaches. They are comprised of node tiers with input layer on each tier. Each node has a distinct weight and threshold and is linked to the other nodes. Building a convolution neural network is indeed a practical approach for categorizing photographs using deep learning. A Keras library module for python is available which makes building CNN incredibily simple. To view images, computer require pixels. Photos typically have connected pixels. For instance, a certain collection of pixels can represent a pattern or an edge in an image. Convolutions employ this to help with image recognition. The nodes become active and sends the data to the network's next layer if its output exceeds a predetermined threshold. No data is forwarded to the next tier of the hierarchy if this is not the case.

Neural Networks comes in several types, these neural networks are used for various use cases and data types [6]. For example, recurrent neural networks are majorly used for Speech Recognition and Natural Language Processing whereas the Convolution Neural Networks (CNNs or ConvNets) are usually used for classification purposes or any computer vision tasks. Before CNNs came into action, a manual, extremely time-consuming process of feature extraction methods were made use of for the purpose of identification of objects in images. However, now for the purpose of image classification and for any object recognition task we can make use of Convolution Neural Network as the impact a much more scalable approach. CNN does it's job by deriving principles or concepts from Linear Algebra, Matrix Multiplication, for retrieving the patterns from the image. CNN can be computationally demanding, which makes use of the Graphical Process Unit (GPUs) for training the models [7].

Artificial Intelligence includes Machine Learning as a subtype, here we would be providing a particular data to the system or machine, and this system would be able to derive certain patterns using the data presented. The system will then be able to forecast solutions to a variety of similar problems. The Neural Network of the human brain is where the inspiration of the Neural Network (NN) comes from. A field in Artificial Intelligence is Computer Vision and it focuses in any problems related to image. CNN paired up with Computer Vision has the potential of executing various complex operation such as they would be able to classify images, providing solutions to scientific problems of astronomy and can build Self-Automated Vehicles[8]. CNN is a mixture of Convolution Layer as well as Neural Network. Any Neural Network which is used for Image Processing contain various layers such as the Input Layer, Convolution Layer, Pooling Layer, Dense Layer.

Convolution is a type of filter which is used on images, this filter helps to extract feature from it. The Convolution makes a filter of certain size (Default size is 3X3) [8]. Element wise multiplication is performed from the image's top left corner after the filter is created. Multiplying elements with the same index is known as element-wise multiplication. A pixel value is created by adding these computed values together and is then stored in the new matrix. This recently formed matrix will be put to use in additional processing [9]. Following the application of convolutions, there is a notion refered as pooling. Pooling is a method for reducing the size of an image. The first convolutional layer, when building a neural network, requires the shape of the picture that is provided to it as input. After the image has been passed, the image will be transmitted through all convolutional layers and pooling layers before being sent to the dense layer [10].

B. Mediapipe

MediaPipe is an open-source framework by Google, which is been utilized for applying in a machine learning pipeline. Being multimodal, it can be applied for various media and audio. The processes of a program that used the MediaPipe must be developed in a pipeline configuration. A pipeline is made up of components known as calculators, each of which is connected by streams through which data packets pass. Developers can create their own application by replacing or defining custom calculators anywhere in it [11].

C. Streamlit

Streamlit is an open-source app framework built on python. It allows us to swiftly build data science and machine learning web apps. Scikit-learn, SymPy(latex), Matplotlib, Keras, Pandas, Pytorch, Numpy, and other Python libraries are compatible with it. [12]

D. OpenCV

OpenCV is an amazing tool to perform image processing and computer vision task. Object Detection and Object Tracking is one of the major functionalities of this open-source library. OpenCV performs a vital role in today's world as it can be used in various real-time application. It can be used to detect faces, objects and handwriting in an image or video [13].

III. METHODOLOGY

This work takes into account the major challenges which the machine learning system faces and the core of the system is the data training part. The data instructing portion of the system is instructed using the real data of people's facial emotions. For instance, for the system to determine an angry facial emotion the system should first be trained with the angry reaction. Similarly, for the model to determine a happy facial emotion they will have to first be trained using the happy emotion. To antecedents the model with these emotion types, we make use of the re-training process. Re-training data has been assembled from the physical world. The challenge in the system was the retraining portion. Various other parts of the system are also considered as challenging. Machine Learning is an extremely potential tool that provides for more efficient and rapid data processing of large databases. This provides the ability of detecting emotion more accurate. The System is able to provide feedbacks in actual-time. The model need not wait to get the final result in later time, and the photo taken need not be stored.

A. Data Collection

The mediaPipe assigns different landmarks to different points in the face. The data contains different landmark points of our face, and one particular row would be comprising of all the key points as face key points, left hand landmark, right hand landmark and various other samples would have all these properties. We compare the differences in those landmarks during each emotion to train the model on different emotions passed by the user, like happy, sad, etc. Hence, the model is able to classify each of the emotion passed by the user.

We use the video capture class to capture the video feed coming from the webcam. After capturing the video, the system will read the frame and show it to the user.

We make use of holistic solution inside the mediaPipe. The holistic solution would take in the frame, and it would return all the facial key points such as left hand, right hand. Then the frame is converted from cv2 colour to RGB because cv2 reads integer formats. Basically, we use process function inside holis and would pass the frame and get the result out of it.

We then use drawing function to draw on our face and mark the face landmark, right-hand landmark and the left-hand landmark on the frame, of the result variable. We then store all these drawings in a list.

The collected data of various emotions are then stored in a numpy file format with specific names associated to it (such as happy.npy, sad.npy, angry.npy, neutral.npy, rock.npy, surprise.npy.)

B. Data Training

Creating a Convolutional Neural Network (CNN) is an exceptional method to categorize images using deep learning. We make use of a library called Keras in python to construct a CNN model. PCs recognizes photos as pixels. The pixels present in photos are normally associated to each other. For instance, an edge in the image or any related pattern might be represented by a specific set of pixels. Convolution utilizes the pixels to recognize these photos. The process occurring in the convolution layer is that it multiplies the matrix of pixels with a filter matrix (also known as 'Kernel ') and then adds up these multiplication values. After process one part of the pixel it then moves over to the next pixel portion to process it in similar method. This process continues until the whole image pixel have been covered. We make use of a model which has the type as Sequential. Sequential is the easiest way to generate a model in Keras. Layer by layer model is created. We make use of various layer in our model. The first 2 layer in the convolution layer works with the input images, these are represented in 2-Dimentional matrix form. Another function in this algorithm is Activation. We make use of the ReLU which is the Rectified Linear Activation as the activation property for the initial 2 layers. The ReLU activation property is proved to work well with neural networks. A Flatten layer is present in between both the Conv2D layer and also in the Deep layer. The flatten/reduce layer serves as a interrelation between both the conv2D layer and the Dense layer. The result most probably will be predicted based on the highest probability. The next step is compilation of the model. The system is compiled using the important three parameter: metrics, optimizer and loss. Out of the three the study rate is being managed using the optimizer. For the loss function a 'Categorical cross entropy' was used. This is considered as a widely used option for classification. Lower the score, better the performance of the system. To make items even much better to understand, when processing the system, the 'accuracy' measure will be utilized.

Initially all the files that we have created are searched (which are happy.npy, sad.npy, angry.npy, neutral.npy, rock.npy, surprise.npy). the npy files are filtered using the split function. The files are then stored in an array(X array) with the labels associated with it in another(Y array).

For example, while talking about happy.npy, all the data under happy.npy will be the input to the model and will be stored in the X array. For a particular input data, we require the model to predict something. What we require the model to predict is the emotion happy. This prediction data will be present in the y variable.

Once the initialization is completed, we are going to concatenate the X and Y array. Basically, it concatenates the input data to the X array and the prediction result in the y array. The name of the file will have an integer associated to it.

The model then passes the file through the CNN algorithm to predict the emotion of the person.

C. Frontend

Earlier we have already created a model which can detect different emotions like sad, angry, happy etc.

We then deploy this model into a web app. The trained model is going to give us a model.h5 file. It's important to note that the structured data is been stored in the h5 file format, not a

model in and of itself. Since the weights and the model setup can easily be stored in a single file, Keras stores the model in this manner. Keras is a powerful and user-friendly free open-source Python tool for building and evaluating deep learning models.The model.h5 are been used to create a web app. Model.h5 will help in creating a web app using different python libraries mainly streamlit and streamlit webrtc.

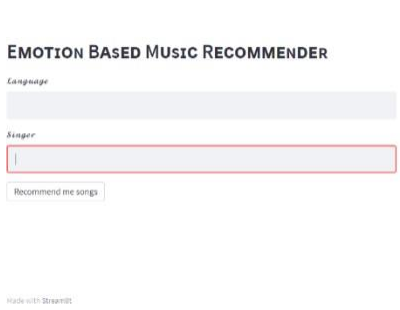

Fig 1: Streamlit Webapp

Streamlit is considered as open-source app toolkit which is based on Python. It allows us to create web applications for machine learning and data science quickly. Scikit-learn , Keras , Pytorch , SymPy (latex), Numpy , pandas , Matplotlib and other Python libraries are compatible with it [14].

Firstly, in the coding part we are going to give two variables for getting the language and the singers name so that based on the user's choice the songs can be recommended. Then we use webrtc to capture the video of user's face so that based on the different emotions made, songs can be recommended. WebRTC (Web Real-Time Communication) is a technology that enables websites and Web applications to capture and possibly broadcast audio and/or video content as well as send arbitrary data between browsers without the need for a third party.. Streamlit webrtc is a python library is used here for real-time video and audio streams over the network, with streamlit. [15]

We have used different python libraries in our project each of them performs different functions. We used load_model to load the model. We used mediapipe library to detect the landmarks of face and hand. We used Cv2 for drawing different functionalities. CV2 is considered as an open-source library which can be utilized for doing tasks such as face detection, object tracking, landmark detection, and more. We also used webbrowser module to recommend different songs from YouTube. The users are provided with a high-level interface for viewing web-based material by the webbrowser module.

The webrtc captures the video and from that it will enable us to predict the emotions. [16] The predicted emotion is going to be saved locally into a file and when required we load the file. We use numpy library to save the prediction or that particular emotion.

Now when the emotion is detected, a web browser tab is going to be opened and it should be a YouTube tab that consists of all the recommended songs based on the emotions, language and the singer. To recommend different songs from the YouTube we import another module named webbrowser. The webbrowser module passes the URL. We inject all the keywords which has been retrieved from the frontend to the URL query. [17] The user is then redirected to the YouTube page as per as the URL query. For example, if we input the language as English, singer as Taylor Swift and the emotion is detected as happy, then the web app will recommend happy songs by Taylor Swift.

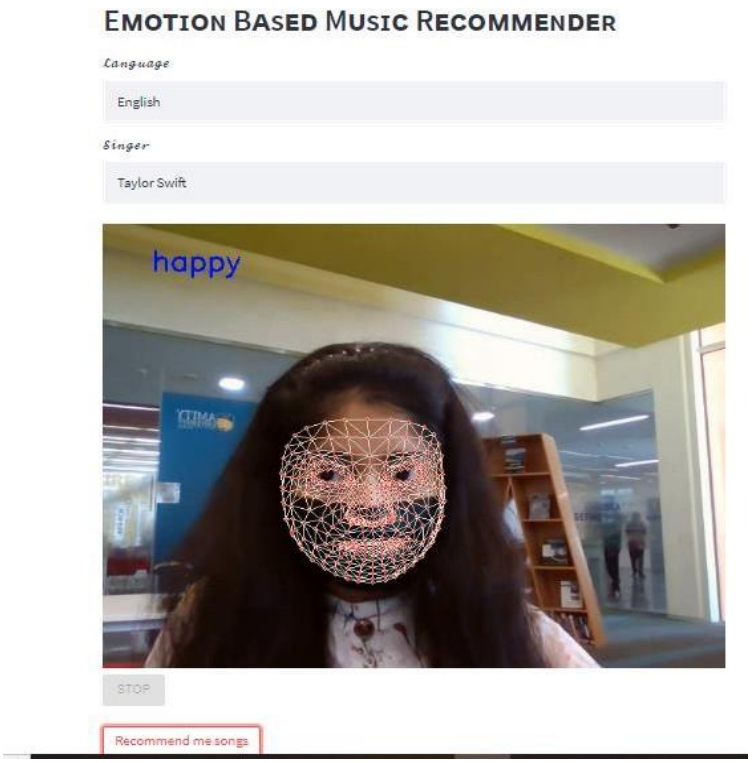

Fig 2: Key words are received from the User

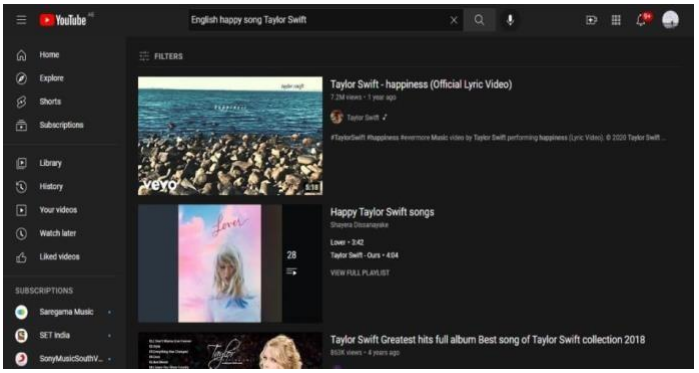

Fig 3: The user is redirected to the youTube page

IV . RESULT AND DISCUSSION

The figure below showcases how the model performs. After the language and singer is entered, the webcam starts capturing the emotion of the user. The captured emotion is then analyzed by the model, the inputs such as “Language”, “Singer” and “Emotion” would thenbe injected into the URL Query. The user is then redirected to a YouTube page as required.

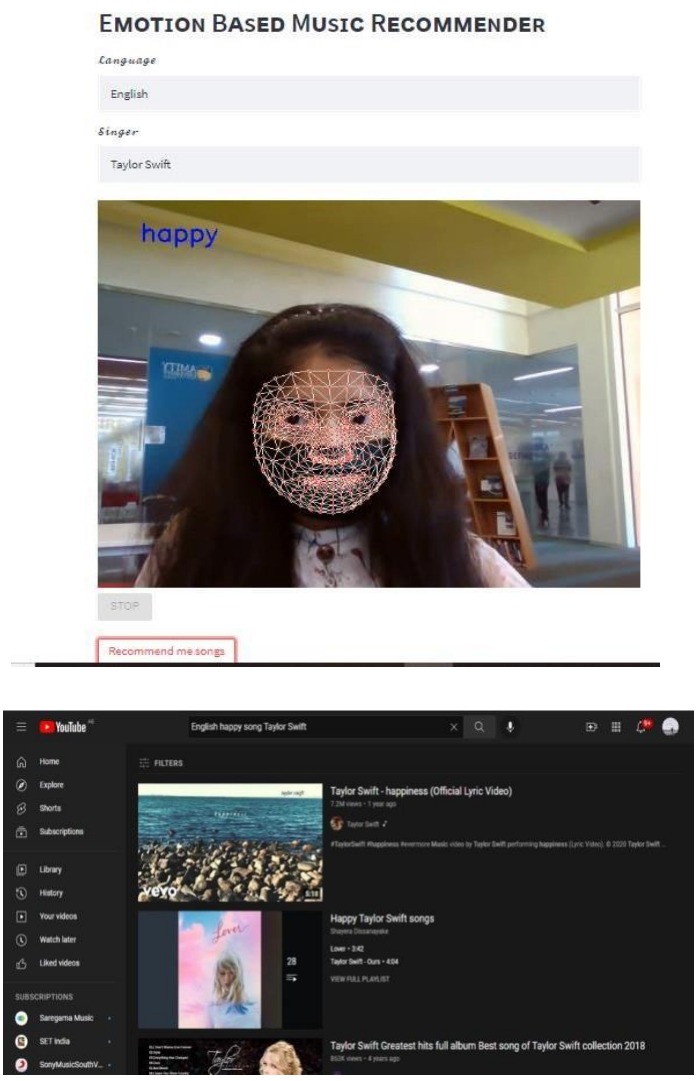

Fig 4: Final Result

V. CONCLUSION AND FUTURE WORK

One of the essential areas of study is the identification of emotions from the facial expressions, which has previously attracted a lot of interest. It is clear that the difficulty of emotion recognition using image processing algorithms has been growing daily. By utilizing various features and image processing techniques, researchers are constantly looking for solutions to this problem.

In this paper we have implemented a system where two predicates such as language and singer is been used to understand the preference of the user. Once the predicates are entered, the webcam would start to capture the image of the user, the captured emotion is then analyzed by the model, the inputs such as "Language", "Singer" and "Emotion" would then be injected into the URL Query. The user is then redirected to a YouTube page as required.

This study presents a method for detecting the basic universal emotions from frontal facial expressions. After, implementing the facial recognition machine learning model, we then further continue to make it into a web application by using Streamlit. The Emotion detection is performed using Deep learning. A well-known model in the field of pattern detection is Deep Learning. The keras library is being used, as well as the Convolution Neural Network (CNN) algorithm. A CNN indeed is an artificial neural network that includes machine learning components. Among other things, CNN can also be used to detect objects, perform facial recognition and process images.

Some modifications that could be made to this system are:

- Add advices that would help the user according to his/her emotions (for example, if the system the emotion of the user as sad, then the system would provide some motivational quotes or other advices that would cheer up the user.)
- The system can also provide small activities that would help improve the mood of the person.
- Improve the face detection accuracy.

VI. REFERENCES

[1] Rumiantcev, M. and Khriyenko, O., 2020. Emotion based music recommendation system. In *Proceedings of Conference of Open Innovations Association FRUCT*. Fruct Oy.

[2] Ali, M.F., Khatun, M. and Turzo, N.A., 2020. Facial Emotion Detection Using NeuralNetwork. *the international journal of scientific and engineering research*.

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

[4]James, H.I., Arnold, J.J.A., Ruban, J.M.M., Tamilarasan, M. and Saranya, R., 2019. Emotion based music recommendation system. *Emotion*, *6*(03).

[5] Gupte A, Naganarayanan A and Krishnan M Emotion Based Music Player-XBeats International Journal of Advanced Engineering Research and Science 3 236854

[6] Ruchika, A. V. Singh, and M. Sharma, "Building an effective recommender system using machine learning based framework," in 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), Dec 2017, pp. 215– 219.

[7] L. Shou-Qiang, Q. Ming, and X. Qing-Zhen, "Research and design of hybrid collaborative filtering algorithm scalability reform based on genetic algorithm optimization," in 2016 6th InternationalConference on Digital Home (ICDH), Dec 2016, pp. 175–179.

[8] Will Hill, Larry Stead, Mark Rosenstein, George Furnas, and South Street. Recommending andEvaluating Choices in a Vitual Community of Use. Mosaic A Journal For The Interdisciplinary Study Of Literature, pages 5–12, 1995.

[9] M.A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes,and Malcolm Slaney. Content-based Music Information Retrieval: Current Directions and Future Challenges.Proceedings of the IEEE, 96(4):668–696, 2008.

[10]Qing Li, Byeong Man Kim, Dong Hai Guan, and Duk Oh. A Music Recommender Based on AudioFeatures. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information

retrieval, pages 532–533, Sheffield, United Kingdom, 2004. ACM.
[11]Bhat, A. S., Amith, V. S., Prasad, N. S., & Mohan, M. (2014). An Efficient Classification AlgorithmFor Music Mood Detection In Western and Hindi MusicUsing Audio Feature Extraction. 2014 Fifth International Conference on Signal and Image Processing, 359- 364.
[12]Talele, M., Gurnani, Y., Rochani, H., Patil, M. and Soneja, K., SMART MUSIC PLAYER USINGMOOD DETECTION.
[13]Ninad Mehendale, ,Facial emotion recognition using convolutional neural networks (FERC),‘ 18 February 2020

[14]Fan, X., Zhang, F., Wang, H., & Lu, X. (2012). The System of Face Detection Based on OpenCV. In 24th Chinese Control and Decision Conference (CCDC),Taiyuan, China. IEEE.

[15]Gilda, S., Zafar, H., Soni, C., & Waghurdekar, K. (2017). Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India. IEEE

[16]V. Bhandiwad, B. Tekwani, Face recognition and detection using neural networks,in International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, India,

[17]MediaPipe Team, Face Mesh. Mediapipe, 2020 [online].