IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11559

Song Recommendation Using Emotion Detection

Kirti Jain[1] Shruti Swarup Srivastava[2] Tushar Vij[3]

[1,2,3] Department of Computer Science & Engineering, Inderprastha Engineering College,Ghaziabad, Uttar Pradesh, India

Abstract: In recent years, recommendation systems have been used to make people's lives easier with product recommendations used by Amazon/Flipkart, movie recommendation used by Netflix/Amazon Prime. While recommendation systems alone help the user make decisions, integrating emotion analysis with this system would automate a variety of things. The analysis of human sentiment, which is also referred to as mining of opinions or Emotion AI in circumstances, is the study of different states of the human brain. The use of emotion detection technologies has garnered a lot of interested patrons over the years so the project is based on detecting user's real time emotion and then using the detected emotion to generate a playlist of songs, recommending songs based on a person's emotions.

Keywords: Emotion Detection, Deep Learning, Web Development, Recommendation System, Machine Learning, Image Processing

I. INTRODUCTION

Music has always been an essential component in expressing oneself, it is a way in which people connect with each other despite their differences. With the evolution of technology, many music streaming platforms have gained worldwide recognition. With the rapid improvement in Machine Learning and Artificial Intelligence, emotion detection systems and recommendation systems have been on the rise with big platforms like Amzon and Spotify implementing recommendation systems in one way or another (collaborative approach).

1.1 Problem and Motivation

Sometimes, people don't want to do the extra work of manually searching for a song, either they are sad or just angry but want to listen to music which would suit their mood accordingly. This music recommendation system is aimed for people experiencing any kind of emotions but most importantly, people going through depression or a sad phase in their life can benefit greatly because of our application, they can easily let the webcam click their picture and get a playlist of the required songs all playable in the webapp.
The objective is to make the user's experience seamless so we built this application which combines frontend and backend technologies to give fast results which will help users access different songs faster, everyone deserves to listen to quality music based on how one is feeling at any given moment and we built this app in hopes of achieving this goal.

II. LITERATURE SURVEY

Emotion Detection and Characterization using facial features [1] aims to detect faces from any given image, extract facial features (eyes and lips) and classify them into 6 emotions (happy, fear, anger, disgust, neutral, sadness) The training data is passed through a series of filters and processes and is eventually

characterized through a Support Vector Machine (SVM), refined using Grid Search. The testing data then tests the data and their labels and gives the accuracy of classification of the testing data in a classification report.
The best result achieved so far is by passing the training images through Histogram of Oriented Gradients (HOG) followed by characterization by SVM, which gives an average precision of 85%.

A novel music recommendation system using deep learning [2] focuses on an approach to improving music recommendation systems. The proposed solution could be applied to many different platforms and domains, including YouTube (videos), Netflix (movies), Amazon (shopping), etc. Current systems lack adequate efficiency once more variables are introduced. Our algorithm, Tunes Recommendation System (T-RECSYS), uses
a hybrid of content-based and collaborative filtering as input to a deep learning classification model. The authors apply their approach to data obtained from the Spotify Recsys Challenge, attaining precision scores as high as 88% at a balanced discrimination threshold.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11559

In Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition [3], the authors use a combination of micro-expression recognition technology of convolutional neural network and automatic music recommendation algorithm is developed to identify a model that recognizes facial micro- expression and recommends music according to corresponding mood. This model was successfully built with a recognition rate of 62.1%. After identifying the corresponding expression, a content-based music recommendation algorithm is used to extract the feature vector of the song and a cosinesimilarity algorithm is used to make the musicrecommendation.

Emotion Recognition using Facial Expressions [4], the results of recognition of seven emotional states (neutral, joy, sadness, surprise, anger, fear, disgust) based on facial expressions. The classification of features performed using k expressions play an important role in recognition of emotions and are used in the process of person. Person's face, as the most exposed part of the body, allows the use of (usually cameras) to analyze the image of the face for recognizing emotions. In this experiment, the authors usedMicrosoft Kinect for 3D face modeling
mainly because of its low price and simplicity.

In Android based Emotion Detection Using Convolutions Neural Networks [5], the proposed study uses Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) to conduct a comparison of which deep learning technique works best foremotion recognition. Both neural networks are trained using FER2013 dataset of Kagglewith seven emotion classes. The trained models are evaluated where CNN attains the accuracy of 65% and RNN lack behind with the accuracy of 41%. The trained models are then applied using music player based on one's facial expressions.

In Audio Based music classification with pretrained convolutional network [6], the authors built a convolutional network that is then trained to perform artist recognition, genre recognition and key detection. The network is tailored to summarize the input features over musically significant timescales. We find that our approach improves accuracy for the genre-recognition and artist recognition tasks. The Million Song Dataset isa collection of all information that is available through API 2 for popular songs. This meansthat a lot of the data was automatically derived from musical audio signals.

Music Emotion Classification and Context- based Music recommendation [7], the authorsfocus on three different ways in which the idea of context awareness is incorporated in the recommendation systems. SVM is used as an emotional state transition classifier. Firstly, the authors proposed a novel emotion state transition model (ESTM) to model human emotional states and their transitions by music.With ESTM, we can recommend the most appropriate music to the user for transiting to the desired emotional state. Secondly, the authors presented context-based music recommendation (COMUS) ontology for modeling user's musical preferences and context, and for supporting reasoning about the user's desired emotion and preferences.
Thirdly, for mapping low-level features to ESTM, the authors collected various high- dimensional music feature data and applied nonnegative matrix factorization (NMF) for their dimension reduction. We also used support vector machine (SVM) as emotionalstate transition classifier.

III. METHODOLOGY

The music recommendation system using emotion detection is a web application that combines frontend and backend features along with a model which predicts a user's emotion effectively. There are three major modules and some minor modules involved inthe basic architecture of the web application with major modules like face recognition, minor modules like setting up the UI corresponding to major guidelines. The discussion of system architecture and design will be based on the three major modules –face recognition, emotion detection and musicrecommendation.

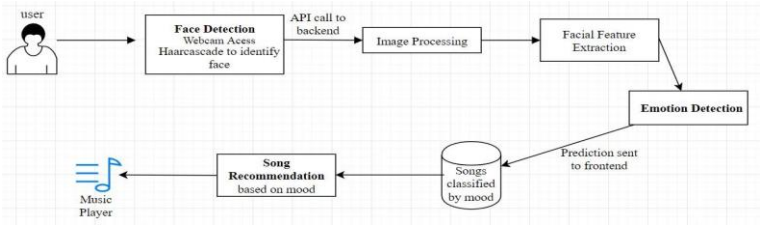


Fig 3.1- Architecture Diagram of music recommendation using emotion detection

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11559

Module 1 – Facial Recognition

This is an important module since we are capturing the image of the user which is then sent for the further process of image processing and facial feature extraction in the backend. We request the user to grant access to their webcam which is then used to capture a real time image of the user, this image is then sent to the backend through a Flask API in the converted base64 format. Since this is a web application, we have a frontend and backend containing different elements bind together using Flask. In this project, we are using real-time face detection. Our system will make use of Haar Cascade files to determine the presence of faces in the captured image. Real time face detection is an easier problem than detecting face in a static image.

Module 2 – Emotion Detection

The emotion detection module focuses on feature extraction from the captured image and running the image through a trained model. The classification is done for
{“angry”,”disgust”,”fear”,”happy”,”neutral”,” surprise”}. Emotion detection module takes place in the backend of the application. When the captured image is sent to the backend, it is in base64 format which is then decoded in the backend server which is later utilized to predict the real time emotion of the user. In the backend of the project, the decoded image is first converted to the image that the model is used to - feature extraction is done, cropping of the image so that only the faces in the image are captured using Haarcascade files, then the image is converted into grayscale after which it is passed through the emotion detection model which accurately predicts the emotion in the captured image.
The facial points mainly are eyes, eyebrows and lips (basic constraints). The movement of all these feature points is recorded for further detection.

Testing of the API was done using Postman, where a POST request was made with a base64 encoded image which is then rescaled and feature extraction is performed, the output is what the API will send – the top two predicted emotions from the captured image

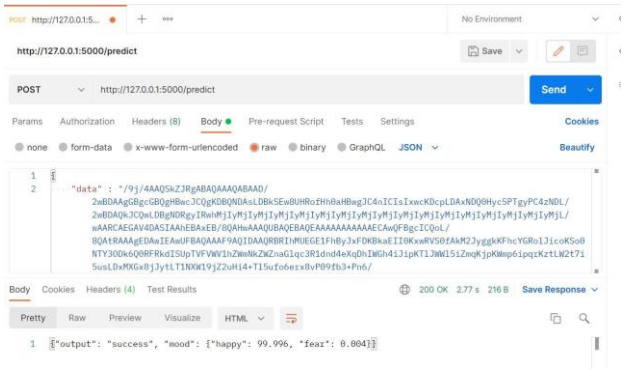


Fig 3.2 - API sending output as the top 2 predicted emotions by the CNN model

Module 3 – Recommendation System

The image captured by the webcam will be used for analyzing the mood of the user using an emotion detection module. After the mood is predicted, the already classified songs (based on mood) will be recommended to the user. Suppose a user’s current mood is ‘happy’ the recommendation system will take that into consideration, search for “happy” songs (all the music data will be stored in a Flask database which will be accessed when recommendation process begins), and looks for songs which fall into the same category as that of the predicted mood. The song is then displayed on the user interface in the form of a playlist where the user can listen to the music. The user also has the ability to manually search for his/her preferred song.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

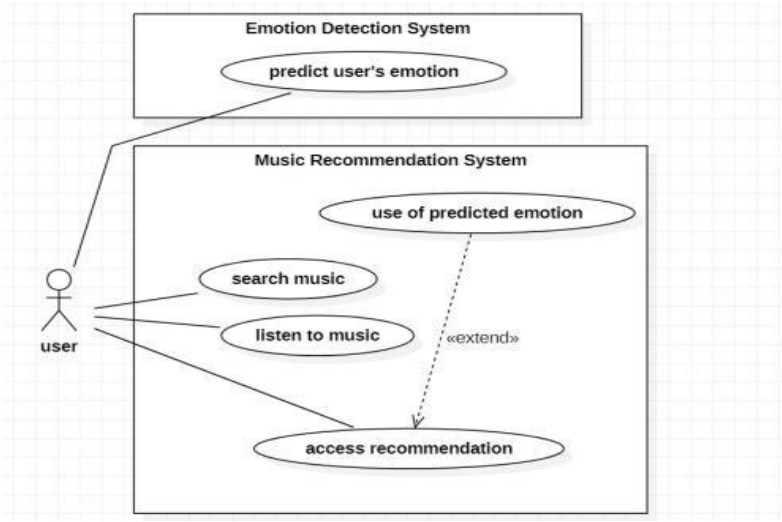
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11559

IV. WORKING OF THE APPLICATION

Fig 4.1 - Use Case diagram of the song recommendation system using emotion detection

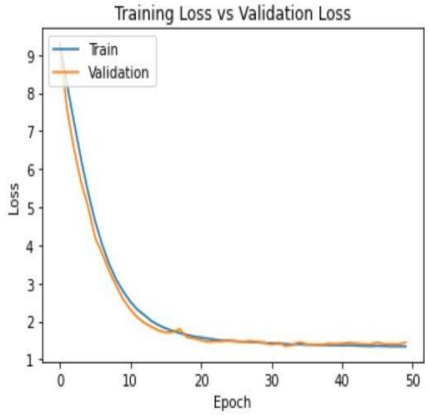


The main objective of the application is to make use of automation in selecting the right song for the end user. The facial recognition module will be responsible for detecting the face of a person within a given frame, we have made use of the extensive python libraries to help in this feature, after the face is detected within the frame, the emotion detection module comes into play. Using this module, facial features are extracted from the captured image which is used by the model to predict the most probable mood/emotion of the user in real time. The songs will be displayed in the form of a playlist dependingon the mood of the user.

Model Features

The Kaggle FER dataset was used in this application, the dataset consists of both training and testing/validation data containingseveral images of each kind of emotion namely happy, sad, surprise, angry, neutral and disgust (6 classes). We trained the model in two different ways – using MobileNetV2 and CNN Sequential model, the accuracy of the latter gave it an edge, we used CNN Sequential model in this application to train and test our emotion detection model.

Fig 4.2 - Training Loss vs Validation LossFig 4.3 - Training Accuracy vs ValidationAccuracy



The model is working well and can detect real time emotion/mood of the user through facial feature extraction and CNN Sequential model.Whenever a user needs to change the playlist, he can open the web camera again in order to capture the live image and then a new playlistwill be displayed to the user based on the mood. CNN architecture makes use of various features in order to get the optimal result while detecting the user's mood. Feature extraction helps in detecting facial

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

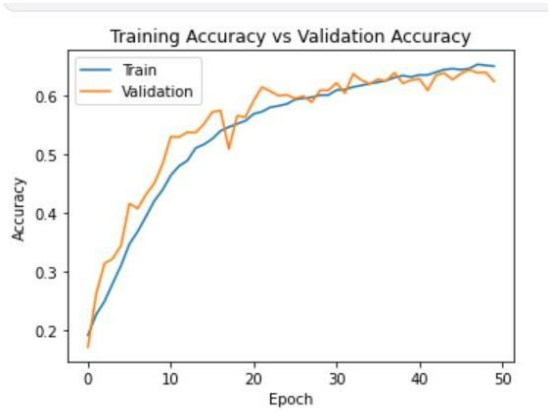
Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11559

features preset on a person's face such as the vertical lines across eyes, lips and nose. Feature extraction is done by all the different layers ina convolutional neural network (BatchNormalization, Conv2D and MaxPooling among others).

V. RESULTS

We were able to successfully capture a user'sface, decode the image being sent to the backend and then we were able to get the predicted emotion of the user, a playlist of favorable songs is then displayed to the user, the user can then choose a song of his liking and has the option to browse through a variedlist of songs.



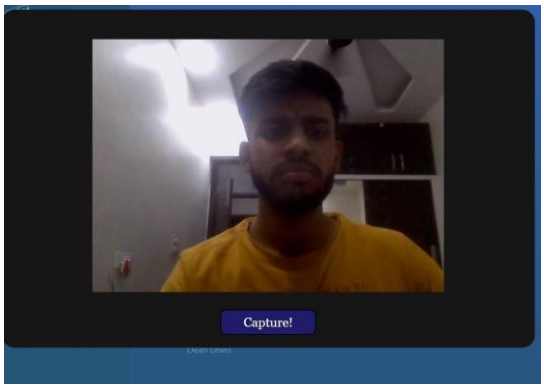


Fig 5.1 - Capturing the image of the user in realtime through webcam access

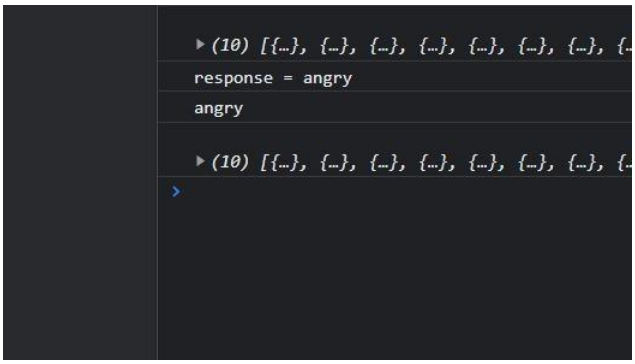


Fig 5.2 - The predicted emotion being shown as'Angry'

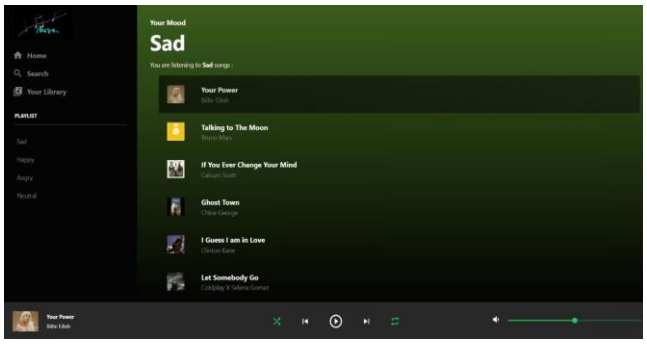


Fig 5.4 - The predicted emotion being shown as'Sad'

Fig 5.5 - The predicted emotion is shown on top with the playlist of songs

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

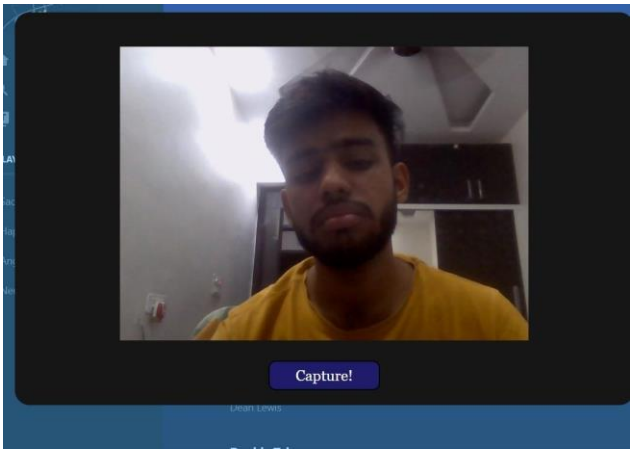
Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11559

Model	Accuracy
MobileNetV2	64.32
CNN Sequential Model	85.71

VI. FUTURE SCOPE ANDCONCLUSION

By identifying a person's mood, we can find what song or playlist he/she would love to listen. Using DL/ML techniques we can achieve the solution of the problem of automating both the emotion recognition as well as song recommendation. Automating generation of songs/ playlists depending on the users' real time facial emotion solving theproblem of manually searching for desired songs is the objective achieved.



As of right now, the song directory is filled with downloaded songs of each category that can be predicted by the model, there are 10 songs for each category and the songs are shown in a playlist manner where the user gets the freedom to select any song, he/she wants. In future, we want to make use of the Spotify API which will directly be able to get access to a large database of songs of the various kinds of moods that can be predicted by our model. Another option is for us to use a relational database like MySQL to store all the different kinds of songs which can be later used to generate different playlists containing random songs from the category.

VII. REFERENCES

[1] Jain, C., Sawant, K., Rehman, M., & Kumar, R. (2018, November). Emotion detection and characterization using facial features. In 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE) (pp. 1- 6). IEEE

[2] Fessahaye, F., Perez, L., Zhan, T., Zhang,R., Fossier, C., Markarian, R., ... & Oh, P. (2019, January). T-recsys: A novel music
recommendation system using deep learning. In 2019 IEEE international conference on consumer electronics (ICCE) (pp. 1-6). IEEE.

[3] Tarnowski, P., Kołodziej, M., Majkowski, A., & Rak, R. J. (2017). Emotion recognition using facial expressions. Procedia Computer Science, 108, 1175-1184.

[4] Han, Byeong-jun, et al. "Music emotionclassification and context-based music recommendation." Multimedia Tools and Applications 47.3 (2010): 433-460.

[5] Dieleman, S., Brakel, P., & Schrauwen, B. (2011). Audio-based music classification with a pretrained convolutional network. In 12th International Society for Music Information Retrieval Conference (ISMIR 2011) (pp. 669-674). University of Miami.

[6] Rosa, Renata L., Demsteneso Z. Rodriguez, and Graça Bressan. "Music recommendation system based on user's sentiments extracted from social networks." IEEE Transactions on Consumer Electronics61.3 (2015): 359-367.

[7] Moscato, Vincenzo, Antonio Picariello, and Giancarlo Sperli. "An emotional recommender system for music." IEEE Intelligent Systems 36.5 (2020): 57-68.