2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS) | 979-8-3503-4314-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/CSITSS60515.2023.10334237

An Emotion Aware Music Recommendation System using Flask and Convolutional Neural Network

Shashank Gajula
Computer Science & Engineering (AI&ML)
Vardhaman College of Engineering
Hyderabad, India
gajulashashank99@gmail.com

Tigulla Hruthika Goud
Department of Information Technology
Mahatma Gandhi Institute Of Technology
Hyderabad, India
hruthikatigulla@gmail.com

Bhawani Sankar Panigrahi
Department of Information Technology
Vardhaman College of Engineering
Hyderabad, India
bspanigrahi@vardhaman.org

M.A.Jabbar
Computer Science & Engineering (AI&ML)
Vardhaman College of Engineering
Hyderabad, India
jabbar.meerja@gmail.com

***Abstract*—Music has been instrumental in influencing the emotions of a person. Every person develops a choice of music depending on varied factors like- philosophy, situations and personal emotions. Many organizations have been providing different ways to recommend a good playlist of music to their users based on their previous choices and emotion of user but couldn't bridge the gap of personalization and emotion driven recommendation. This paper emphasizes on building this gap by providing a recommendation of music based on user's choices and their current emotions. The proposed model is an innovative approach by harnessing the power of Convolutional Neural Networks (CNN) and music therapy approaches to ensure to provide people with an apt recommendation that enhances the user's experience of music. This model has achieved an accuracy of 71%. It can easily be integrated with music platforms to enhance the user experience and provide better services.**

Keywords—Artificial Intelligence, Deep Learning, CNN model, Music Therapy Approaches, NLP

I. INTRODUCTION

The ease with which people can now access huge music archives has completely changed how people listen to and engage with music in the modern digital age. As a result of this paradigm change, intelligent platforms called music recommendation systems have emerged. These systems provide tailored music recommendations that are in line with user tastes in an effort to increase user engagement. In order to create playlists that connect with consumers on a deep level, these systems must be able to navigate the complex world of musical genres, performers, and compositions. The seamless access to huge amounts of music being provided by many applications in the market need to tailored in accordance with user preferences and emotions to keep the user engaged and provide an pleasurable music experience. By such provisions of the music applications the user is being recommended a very narrowed down list of music from an enormous collection of music from different genres . This provides the user a hustle-free experience. These recommendation systems will engage more and more users to the music applications bringing a good capital to the application.

The traditional music recommendation systems had a considerate amount of success in providing a recommendation based upon the study of user's previous choices of music. But when it came to the emotion-based recommendation more advanced techniques were required. Basically, there are two main methodologies, one method is traditional way of content-based filtering approach which emphasizes more on the previous choices of the user and recommends the music with relevant content.

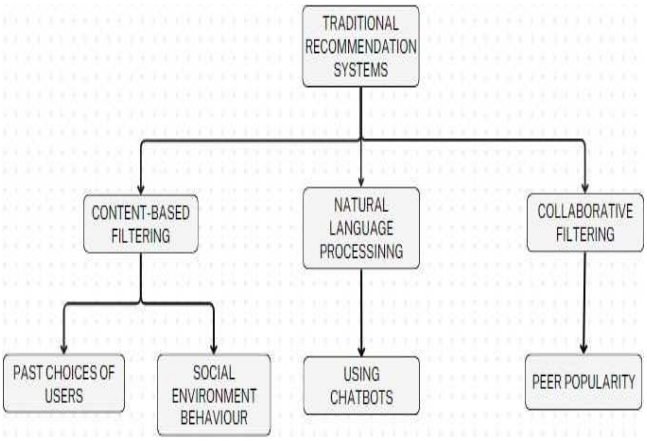

Fig. 1. Traditional Recommendation Systems

Failing to recognize the existing user interests is considered to be the main drawback of the existing method. [1] The other content-filtering methods gave recommendations based upon the social network logs of the user. Using the social network logs, there is an analysis made of the user's behavior in social environment. The other approach is collaborative filtering approach where the user is recommended the most popular music among the peer group with similar preferences. The drawback of such model comes where there is clash between the popularity bias and personal choices i.e., people have choices which are quite different from the mass group of people and such music choices are less popular resulting as less recommended music. The other drawback comes when few music albums (for example the old classics) though are ignored or less recognized by the peer of present generation but still remain the favourites of many music lovers.

One more way recommending music is through human-computer interaction through chatbots and help of API's.[2]The user interacts with the chatbots and answers few simple questions asked by the computer and these answers are analyzed by Natural Language Processing(NLP).Based upon the tone of the answers given by the user the users current emotion and preference is identified and the suitable music is recommended. The IBM analyzer is typically used in this recommendation system to determine the user's tone based on their responses and to anticipate their emotion. The FM API has been supplied with the predicted emotion and will then suggest a playlist of songs appropriate for that particular feeling. But the drawback of this model is manual work of the user to make the computer understand his emotion and one more major problem is not all people will be able express themselves in the form of written text.

II. Background and Literature Survey

A. Challenges Noticed in Existing Systems

The preliminary problem faced during surveying the existing traditional systems is that they failed to encapture the emotion or mood of the user at the particular moment of time. As the mood of user is erratic in nature the models failed to dynamically provide the results. And few systems provided only the results based upon the peer society's choices which would largely vary from the choices of an individual. Due to such problems, it developed a gap between the user's personalized choices and the playlists recommended by various applications. As music is on the major art form that all people get indulged in and enjoy it on regular basis, there is great demand and need in the market for intelligent recommendation systems.

B. Existing System Methodologies

The most basic approach of designing a recommendations is through maintaining a database of information and then based upon the meta data like artist, playlist, album, year of release etc. also we can recommend music to the users[5]. Such music recommendation systems are very basic and are more like an Search engines which helps in extracting information based upon keywords or text similarity factor. This system of recommendations is mostly based on the content-based filtering method.[4][5] The collaborative filtering approach can also be used but it would be unable to provide solutions for those new songs which are not yet rated by many audience or it will not be helpful for those users who have not given any rating or feedback to the songs yet.

Few recommendation systems are developed in collaboration with a interactive chatbot[3].Here the user interacts with the chatbot by answering few questions asked by the chatbot and based upon the answers given by the user , the tone of the user will be analyzed by Natural Language Processing models and other APIs [1][3] which will result the emotion will be returned by the model. And based upon the received emotion, a music classifier will be used to classify the music and a playlist is created which contains the music containing only the music of particular genre [1][3] . These systems had a drawback of manually answering the questions again and again which causing an inconvenient user experience. The other recommendation systems have complete content-based filtering approach which analyses a user personality and recommends the music [2]. The analysis of user personality traits begins with sound psychological observations that are discovered through an analysis of behaviour by users in social contexts. The social environment behavior is the user's activities in social media platforms like twitter, LinkedIn, Instagram and other public social media platforms. The latest updated applications for recommendation is developed using the Mobilenet models and keras which were developed as an mobile application[4].The application interface and development of model is done by Android and machine learning integration. This application has been divided into two parts-Face Detection which was initially considered using the python library OpenCV but later an FaceDetector class of java was considered so that it can be integrated with android app, the following phase is mood detection, that categorises an individual's facial expression as sad, neutral, disgusted ,happy, surprised, or fearful.

In this application Mobilenet which also an CNN architecture have been developed to detect the emotion from a dataset from Kaggle which has got about 40,450 training images and 11,924 testing images and this model gained a good results as an app.

Few other technologies which have been used are:
An external web application can be integrated with Sound Tree's music recommendation algorithm and made available as a web service. It takes advantage of person-to-person association based on previous user behaviours, including having downloaded or listened to music.

The music recommendation service lucyd was developed by four graduate students in the Master of information and Data science (MIDS) programme at UC Berkely. Users can ask Lucyd for music recommendations using whatever terms they choose.

III. METHODOLOGY

The recommendation system for mood-based music focuses on incorporating real-time mood recognition. It is a working model for an innovative device with two main modules: Recognizing facial expressions and moods, and recommending music.

A. Emotion Detection Module

Facial Expression Capture -The cv2 library(OpenCV) has been used for video capture and manipulation. The utils.py python file is encapsulates the video streaming functionality and also starts a sperate thread which continuously reads frames from the specified webcams source. This python file enables webcam frame capturing without interfering with the operation of the primary program. The train.py file contains the python script which has imported the necessary Keras and ImageDataGenerator classes are imported. It contains the Keras dataset FER2013 which is contains the 48*48 pixel grey scale images of faces. Depending on the emotion expressed in the facial expression, each face is to be assigned to one of seven categories (0 = angry, 1 = disgust, 2 = fear, 3 = happy, 4 = sad, 5 = surprised, 6 = neutral). The public test

set has 3,589 examples, whereas the training set comprises 28,709 cases. A sequential Keras model (emotion_model) is defined. The model starts with two pairs of Convolutional layers (Conv2D), each followed by a MaxPooling layer (MaxPooling2D). In order to extract features from the photos, these layers are used.

Dropout layers (Dropout) are inserted to prevent overfitting. Prior to the completely connected layers, the feature maps are flattened using the Flatten layer. There are then two fully connected (Dense) layers, the first of which has 1024 units and is activated by ReLU, and the second of which is activated by SoftMax and has 7 units (for the seven emotions).The network can learn associations between various features thanks to fully linked layers, which link all neurons from the previous layer to the following layer. In the context of emotion recognition, the fully connected layers receive the learned features and perform classification to predict the emotion category.

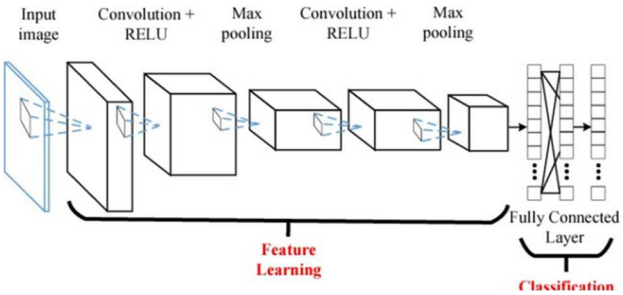

Fig. 2. CNN Model Architecture[4]

B. *Music Playlist Recommendation and Integration Module*

This module involves two main python scripts which are used to retrieve the emotion and create a Dataframe where playlists are added using pandas of python and then an API call to retrieve various playlists for different emotions. The spotify.py is python file which has the python script capable of obtaining track information from Spotify playlists and producing CSV files with the track information for various moods. The Spotify library and the Spotify API appear to be used to obtain and analyze track data from different playlists based on distinct moods. Sleep delays are also incorporated into the code to prevent the Spotify API from being overloaded. The created CSV files each represent the playlist for each emotion and the song title, album title, and artist name are all included in each CSV file.

The Integration is done by the camera.py file contains the python script which is capable of recognizing the facial expressions using webcam feed and makes music suggestions depending upon observed emotions. The emotion label is used to choose a CSV file with music recommendations for the given feeling. A Pandas DataFrame (df1) is built containing music suggestions. Overall, this script collects webcam footage, recognizes faces in each frame, predicts emotions with a pre-trained CNN model, and recommends music depending on the observed emotions. To facilitate modularity and maintainability, the script is split into classes for various functionality such as video streaming, emotion prediction, and music suggestions.

C. *Interface Module*

The app.py is containing the code for a Flask application which is acting like an interface dynamically analyzing and displaying the user's real time emotion using the webcam video stream and also a dynamic display of playlist is also being shown corresponding to the emotion detected by the model. The code structure is modular, with separate routes for rendering HTML templates and serving video streams. The Flask library has been imported. The main route renders and HTML file named index.html with music recommendation table. The '\video_feed' route generates a video stream by calling the gen function, which provides video frames and updates the Dataframe with music recommendations. The '\t' route returns the JSON representation music recommendations Dataframe. The JSON data can be requested asynchronously using AJAX to update the recommendations dynamically. The app is run when the script is executed as the main program.

Fig.3. displays the System Architecture Diagram and Fig.4. displays the system's data flow diagram.

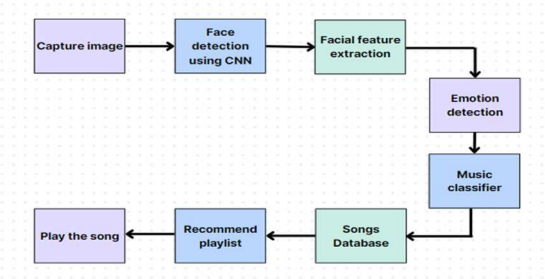

Fig. 3. Architecture of the Proposed Model

The architecture of the system depicts the relationships between various components that include the userinterface,the CNN module,Webcam Video Stream etc.When the user logged-in and the Flask application has been started then the Webcam Video Tream starts taking the live video recording of the user's face at a particular FPS rate defines in the module . Then simultaneously the face is detected and set to the CNN model to detect the emotion of the user based upon his facial expressions.The emotion is identified and it is passed to the Music classifier which extracts the playlists corresponding to the detected emotion and saves it as a CSV file.This CSV file stored in an HTML template and it is routed back to the Flask application and displayed as a table.

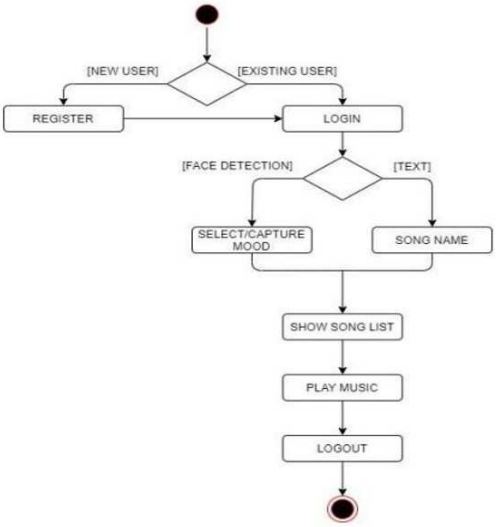

Fig. 4. Data Flow Diagram Of The System

D. *Hardware Requirements*

The standard set of specifications provided by any operating system or piece of software are the computer's physical resources, or hardware . The necessary hardware for this paper is as follows:

- Minimum 4 Gigabyte RAM for processing
- Webcam
- Input/Output devices to interact with the computer
- Network Interface to connect to Internet or local network

E. *Software Requirements*

Software requirements for a model are the computer software resources which define the specifications of the operating system, the required libraries and other computational efficiencies to be fulfilled for the smooth running of the application.

- Libraries- Keras, NumPy, Pandas, Flask, OpenCV
- Spotify Developer Account with all required credentials to be maintained
- Gunicorn-Web Server(optional)
- Visual Studio

IV. RESULTS AND DISCUSSION

The "Facial Expression Recognition 2013" dataset, which was introduced as a part of the Facial Expression Recognition Challenge at the 2013 IEEE International Conference on Automatic Face and Gesture Recognition (FG), was used in the model-building process. The data consists of grayscale portraits of 48 by 48 pixel faces. The training set has 28,709 cases, whereas the public test set contains 3,589 instances. The model will be trained with these images. The biggest advantage of image dataset is images inherently provide contextual information. Anger, disgust, fear, happiness, sorrow, surprise, and neutral the seven facial expressions that are portrayed by the photos in the dataset. Each image has one of these emotions labeled on it.Few test results have been obtained by real-time experimenting with different people.

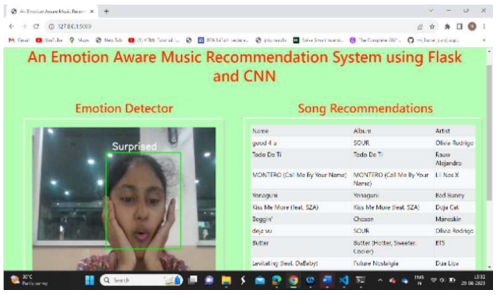

Fig. 5. Recommendation of a music playlist for “Surprised” expression

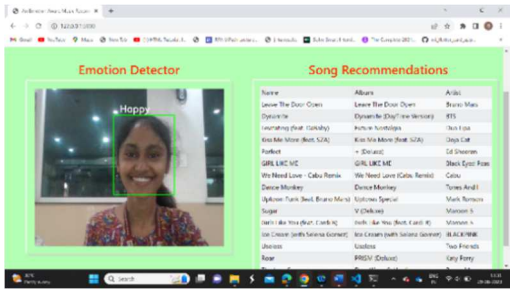

Fig. 6. Recommendation of a music playlist for “Happy” emotion

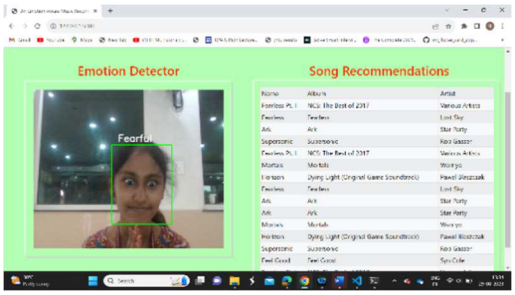

Fig. 7. Recommendation of a music playlist for “Fearful” emotion

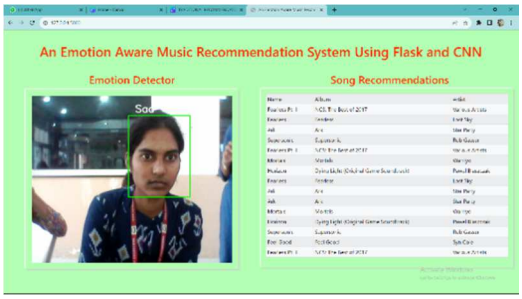

Fig. 8. Recommendation of a music playlist for “Sad” emotion

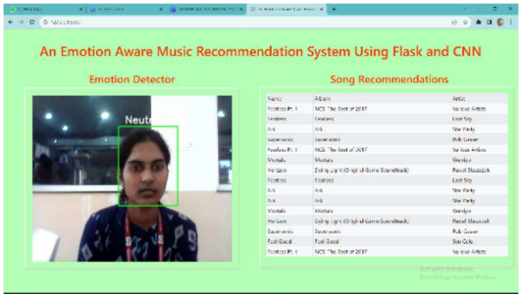

Fig. 9. Recommendation of a music playlist for “Neutral” emotion

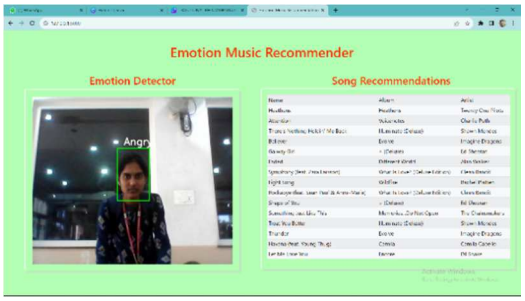

Fig. 10. Recommendation of a music playlist for “Angry” emotion

In Fig.11.Represents a Confusion Matrix. A Confusion Matrix performs as a useful tool for assessing and enhancing an emotion-based music recommendation system.When a classification model's true values are known, a confusion matrix is used to evaluate the model's performance and determine how well it can divide a set of data into distinct classifications.The accuracy of the classification model using the confusion matrix is given based upon four important components: True Positives(TP),True Negatives(TN),False Positives(FP),False Negatives(FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

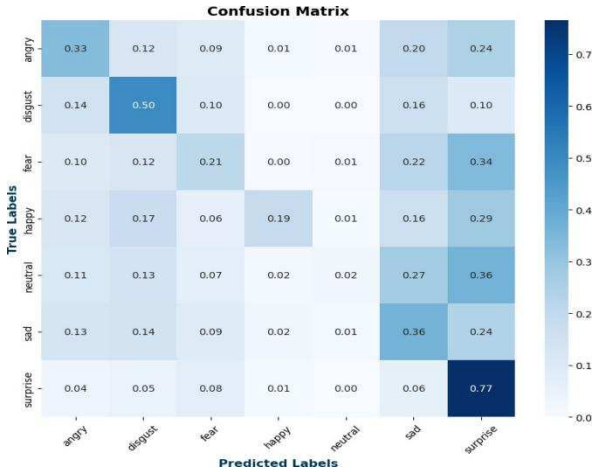

Fig. 11. Data Flow Diagram of The System

In a confusion matrix, the rows represent the actual labels of the data, while the columns represent the predicted labels generated by the model. The number or frequency of data points that fall into a certain set of true and expected labels is represented in each column of the matrix. The matrix's diagonal entries correspond to instances where the true label and expected label match, proving that the predictions were accurate. Off-diagonal elements are instances where the model's predictions were wrong.

In Fig.12, The graph shows the accuracy of our model, with the accuracy shown on the y-axis and the number of epochs specified on the x-axis. As per the results and the plotted graph we observe that our model has achieved an accuracy of 70% approx.

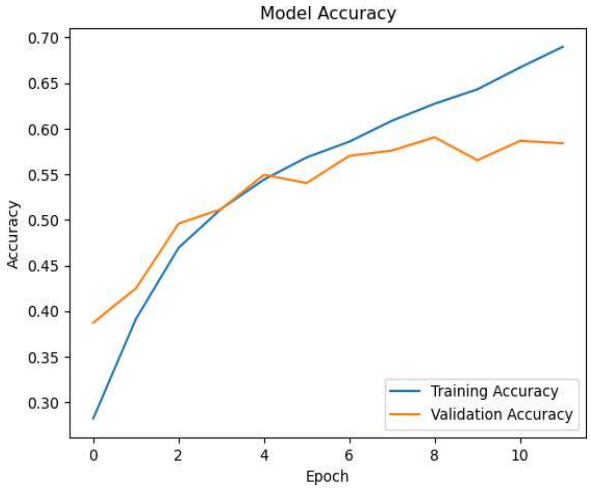

Fig. 12. Training and Validation Accuracy

It is important to strike a balance between high training accuracy and good validation accuracy to ensure the model performs well on unseen data and can make accurate predictions

The following table shows a comparison between different machine learning ,deep learning algorithms and our proposed system.

TABLE I. ACCURACY COMPARISON

S.no	Models	Accuracy
1.	Flask+CNN [Proposed Work]	70.16%
2.	Using chatbot+ Tone Analyzer [1]	80%
3.	VGG+SVM [13]	66.31%
4.	CNN+SVM [16]	68.20%
5.	Search RBF-SVM [5]	70.12%
6.	Bag of Words [7]	67.40%
7.	GoogleNet [15]	65.20%
8.	MobileNet + OpenCV[4]	71.20%

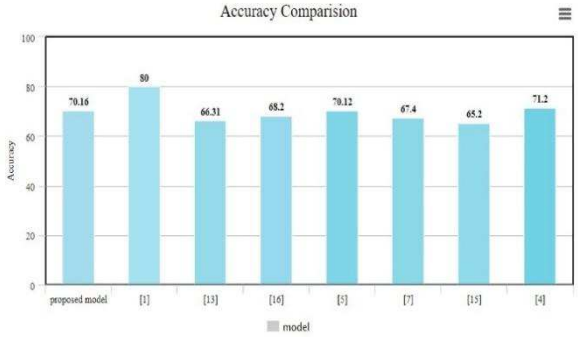

Fig. 13. Bar Graph Representing Accuracy Comparison

V. CONCLUSION

Even though the human emotions are hard to study as the facial expression for a particular emotion varies from person to person, by providing the model with datasets that identify the important facial expression, a machine learning model can be taught to predict the emotion of a person. Our model has a 70% accuracy rate for recognizing seven moods: anger, disgust, fear, happy, sad, surprise, and neutral. An increase in the number of epochs, which lengthens the model's training process, may enhance the accuracy factor. Presently the model has been trained with 10 epochs and produces an accuracy in the range of 69.6% - 70.16% accuracy. Our dataset contains an uneven number of images for each image causing misclassification of the emotion. For example sometimes our model is misclassifying “disgust” as “angry” because our dataset contains predominately more number of angry facial expression images and fewer disgust facial expression images and the facial features (eyebrows and cheeks) for both the emotions are similar.

VI. FUTURE SCOPE

This model can be more effective by deploying it as a mobile application as most of the users enjoy music on their mobile phones. Using other parameters like heart rate or body temperature, which may also be thought of as a future implementation to this model, it is possible to identify strong emotions like fear, disgust, and any other negative emotions. The main future implementation to our paper is, our dataset predominately containing few emotional expressions than other emotions causing misclassification of emotions this needs to rectified by maintain an constant number of sample images of for all the emotions which needs to be collected additionally from real-time faces through show casing of our paper at paper expos, large public gathering’s like malls data can be collected. Future potential uses for our paper could include suggesting TV shows, films, and web series based on mood detection.

ACKNOWLEDGMENT

We would like to thank our paper guide Prof.Bawani Sankar Panigrahi for the valuable support and provisioning of their valuable time to successfully conclude the paper. We also like to extend our gratitude to Dr.M.A.Jabbar.,Head of Department, Department of Computer Science and Engineering (AI&ML) for providing insightful feedback and unwavering guidance. Their expertise and mentorship have played a vital role in shaping the direction of this research.

REFERENCES

[1] Music Recommender System Using ChatBot Shivam Sakore1 , Pratik Jagdale2 , Mansi Borawake3 , Ankita Khandalkar4 1, 2, 3, 4Dept. of Computer Engineering PDEA’S College of Engineering Pune, India.

[2] An Emotional Recommender System for music Vincenzo Moscato, Antonio Picariello and Giancarlo Sperl´ı.

[3] CHAT BOT SONG RECOMMENDER SYSTEM Prof. Suvarna Bahir*1, Amaan Shaikh*2, Bhushan Patil*3, Tejas Sonawane*4 *1Guide, Sinhgad Academy of Engineering Pune, India. *2,3,4Student, Sinhgad Academy of Engineering Pune, India.

[4] Mood Based Music Recommendation System, June 2021 https://www.researchgate.net/publication/352780489_Mood_based_music_recommendation_system

[5] A Music Recommendation System Based on Acoustic Features and User Personalities Rui Cheng1(&) and Boyang Tang2 1 University of Arizona, Tucson, AZ, USA ruicheng@email.arizona.edu 2 Technology University of Delft, Delft, The Netherlands B.Tang@student.tudelft.nl.

[6] M. Albanese, A. dAcierno, V. Moscato, F. Persia, and A. Picariello, “A multimedia recommender system,” ACM Transactions on Internet Technology (TOIT), vol. 13, no. 1, p. 3, 2013.

[7] R. T. Ionescu, M. Popescu, and C. Grozea, “Local Learning to Improve Bag of Visual Words Model for Facial Expression Recognition,” Work. challenges Represent. Learn. ICML, 2013.

[8] P. Aiswarya, Manish and P. Mangalraj, ”Emotion recognition by inclusionof age and gender parameters with a novel hierarchical approach using deep learning,” 2020 Advanced Communication Technologies and Signal Processing (ACTS), Silchar, India, 2020, pp. 1-6.

[9] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Papers. 632. "Facial Emotion Recognition Using Machine Learning" by Nitisha Raut (sjsu.edu).

[10] 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

[11] 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

[12] F. Ricci, L. Rokach, and B. Shapira, “Recommender systems: introduction and challenges,” in Recommender systems handbook. Springer, 2015, pp. 1–34.

[13] R. GUETARI, A. CHETOUANI, H. TABIA and N. KHLIFA, ”Real time emotion recognition in video stream, using B-CNN and F-CNN,” 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, 2020, pp. 1-6.

[14] K. Zhang, Y. Li, J. Wang, E. Cambria and X. Li, ”Real-Time Video Emotion Recognition Based on Reinforcement Learning and Domain Knowl- edge,” in IEEE Transactions on Circuits and Systems for Video Technology,vol. 32, no. 3, pp. 1034-1047, March 2022.

[15] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, “Deep learning ap- proaches for facial emotion recognition: A case study on FER-2013,”in Smart Innovation, Systems and Technologies, 2018, vol. 85, doi: 10.1007/978-3-319-66790-4.1.

[16] Y. Tang, “Deep learning using linear support vector machines,” arXiv, vol. 1306.0239, 2013.