Music Recommendation based on emotion detection using facial expressions and speech recognition

M. Roshan Aditya
Computer Science and Engineering
Vellore Institute of Technology, Chennai Campus.
Chennai, India.
roshanaditya23@gmail.com

Sravanakumar Sathish
Electronic and *Computer and Engineering*
Vellore Institute of Technology, Chennai Campus.
Chennai, India.
sravanakumarsathish02@gmail.com

Abstract— **In this paper, a novel framework for music selection is presented that incorporates speech analysis and facial expression-based emotion recognition. The technology is intended to provide personalized music recommendations that correspond with the user's current emotional condition. One convolutional neural network (CNN) model that has been trained on real-time data streams facilitates emotion recognition. The Speech Recognition library is used for speech recognition, whereas the OpenCV library is used for facial expression analysis. Unlike traditional approaches, the OpenAI API is used to generate music suggestions by utilizing sophisticated language models. The suggested framework provides a tailored and interesting user experience by smoothly merging AI-driven music recommendation with emotion identification.**

Keywords—CNN, OpenCV, Speech Recognition, OpenAI API.

INTRODUCTION

Music is an essential part of human culture and has the power to elicit a wide range of feelings in those who listen to it. Music is a powerful tool for expressing emotions and has been shown to have therapeutic benefits in addition to being just for fun. Music suggestion algorithms have become essential tools for customizing the music listening experience as technology develops. These systems make use of a variety of strategies, including context-aware, content-based, and collaborative filtering, to recommend music that aligns with users' listening patterns and tastes. The user's present emotional state is a critical factor that many sophisticated music recommendation systems fail to take into account.

Researchers working to understand human emotions and how they affect behavior are most advanced in their understanding of emotion detection. Facial expressions and voice analysis are two of the many readily available and non-invasive techniques for detecting emotions among the many other techniques. Through the examination of elements such as facial asymmetry, skin color changes, and facial muscle movements, facial expressions reveal an individual's emotional state. Speech recognition, on the other hand, uses a speaker's subtle variations in pitch, tone, and rhythm to determine how they are feeling.

In order to provide users with personalised music recommendations that are tailored to their emotional landscape, this study presents a novel approach to music recommendation that leverages the capabilities of emotion recognition based on speech and facial expression analysis. A single convolutional neural network (CNN) model, which is skilled at training on real-time data streams to predict the user's emotional state in real-time, lies at the heart of this approach. Accurate voice analysis is made possible by the voice identification library, and robust facial expression identification is made possible by the use of the OpenCV library. Using the OpenAI API, music suggestions are obtained in a non-traditional manner by utilizing sophisticated language models. This revolutionary combination of AI-powered music recommendation and emotion detection signals a paradigm change by providing users with a personalized and engaging audio experience.

This work aims to highlight the revolutionary potential of such a system in enhancing the listening experience to music and highlights the viability of utilising artificial intelligence and machine learning techniques in music recommendation systems. The suggested approach, which places emphasis on the deep emotional connection between music and emotions, aims to provide personalised music recommendations based on individual emotional states, which makes it a unique and valuable addition to the field of music recommendation systems.

I. LITERATURE SURVEY

(1) Londhe RR and Pawar DV's paper concentrated on the study of changes in the crooks of the face and the intensities of the corresponding pixels. The author used Artificial Neural Networks (ANN), which was used to classify the feelings. The author also proposed colorful approaches for a playlist. (2) Kabani H, Khan S, Khan O and Tadvi S's paper proposed two significant orders for facial point birth, which included Appearance- grounded point birth and geometric grounded point birth, which included birth of some essential points of the face similar as mouth, eyes, and eyebrows. This sytem's computational time taken is1.000 second which is veritably lower therefore helping in achieving a better real time performance and effectiveness.(3) Gupte A, Naganarayanan A and Krishnan M's paper determines the mindset of the user by using facial expression Humans frequently express their feeling by their expressions, hand gestures, and by raising the voice of tone but substantially humans express their passions by their face. Emotion- grounded music player reduces the time

complexity of the user. The image of the user is captured by the web camera, and the images are saved. The images are first converted from RGB to binary format. This process of representing the data is called a point- point discovery system. This process can also be done by using Haar Cascade technology handed by Open CV. The music player is developed by using a java program. It manages the database and plays the song according to the mood of the user. (4) Hashizume,Y., Li,L., & Toda, T's paper proposed a music similarity computation system fastening on individual necessary sound sources. The proposed system learns similarity criteria using deep metric literacy with a triplet loss. trials showed that it was possible to learn the similarity criteria fastening on different musical instruments. (5) ParulTambe, YashBagadia, Taher Khalil and Noor UlAin Shaikh's paper proposed an idea which automated the relations between the users and music player, which learned all the preferences, feelings and conditioning of auser and gave song selection as a result. The colorful facial expressions of users were recorded by the device to determine the emotion of the user to prognosticate the genre of the music. (6) Vivek JD, Gokilavani A, Kavitha S, Lakshmanan S and Karthik S's paper proposed an emotion- grounded music player using image processing This showed how colorful algorithms and ways that were suggested by different authors in their exploration could be used for connecting the music player along with mortal feelings. It has therefore helped in reducing the sweats of user in creating and managing playlist and furnishing an excellent experience to the music listeners by bringing them the most suitable song according to the user's his/ her current expression. (7) Dureha A's paper came up with an algorithm that gives a list of songs from the user's playlist in agreement with the user's emotion. The algorithm which was designed was concentrated on having lower computational time and also therefore reduces the cost included in using colorful tackle. The main idea was to insulate the feelings into five orders i.e., Joy, sad, wrathfulness, surprise and fear also handed a largely accurate audio information reclamation approach that uprooted applicable information from an audio signal in lower time.(8) Turnbull,D.R., McQuillan,S., Crabtree,V., Hunter,J., & Zhang, S's paper tried to measure fashionability bias in three state- ofart recommender system models(e.g., SLIM,Multi-VAE, WRMF) and on three marketable music streaming services (Spotify, Amazon Music, YouTube). It is found that the most accurate model (SLIM) also has the utmost fashionability bias while less accurate models have lower fashionability bias.(9) Ren,Y., Lu,J., Beletchi,A., Huang,Y., Karmanov,I., Fontijne,D.,. & Xu, H's paper paper proposed a new algorithm is proposed to fete the facial emotion, which included three stagespre-processing, point birth, and bracket. The first part describes colorful stages in image processing include preprocessing, filtering used to prize colorful facial features. The alternate part optimized the eye and lip cirque characteristics, and in the third part, the eye and lip optimal parameters were used to classify the feelings. The attained results showed that the speed of facial recognition was far better than other usual approaches. (10) Rabashette,M.D., Tale,M.R., Hinge,M.A., Padale,M.K., Chavan,M.R., & Deshmukh, N's paper concentrated on creating a system that fetches the emotion of the user using a camera and also automates the result using the emotion discovery algorithm. This algorithm captures the mood of the user after every decided interval of time as the mood of the user may not be the same after some time; it may or may not change. The proposed algorithm on an average advised estimation takes around0.95-1.05 sec to induce an emotion- grounded music system, which was better than former being algorithms and reduces the cost of designing. (11) Florence,S.M., & Uma, M's paper came up with an algorithm which first captures the users picture in order to descry the user's mood and also according to the mood, an applicable song from the playlist of the user is shown matching the user's demand. (12) Sakore,S., Jagdale,P., Borawake,M., & Khandalkar, A's paper enforced a chatbot that recommends music grounded on the user's textbook tone. In this design, the main thing is to reliably determine a user's mood grounded on their textbook tone with an operation that can be installed in user's desktop. When a song is broken down into quantifiable rudiments like meter, harmony and timbre, it can be matched to specific mood orders according to anticipated data. (13) Yang, J's paper studies a substantiated song recommendation algorithm grounded on oral features. The specific work includes three corridors. originally, the spectrum feature extraction and observe feature extraction of songs. Secondly, grounded on the song of CNN bracket, it sets different types of song bracket. Thirdly, a substantiated song recommendation system was grounded on CNN bracket. Also, the reasons why the bracket of CNN isn't suitable for direct song recommendation is anatomized, and also, a recommendation system grounded on song scrap bracket is proposed .(14) Arkachaari,V., Rao,P.M., & Darapaneni, N's paper explores a new frame to influence the raga- rasa association in Indian classical Music to make an intelligent classifier and its operation in music recommendation system grounded on user's current mood and the mood they aspire to be in. (15) Rozhevskii,D., Zhu,J., & Zhao, B's paper finds a way to integrate users' particular traits and their current emotional state into a single music recommendation system with both cooperative and happy- grounded filtering. They seeked to relate the personality and the current emotional state of the listener to the audio features in order to make an emotion- and-personality aware MRS. (16) Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J's paper provides a thorough analysis of studies on artificial intelligence-based music creation, hoping to close the gap left by earlier assessments that concentrated on particular issues and solutions. Using the PRISMA approach, the authors analyze the extent and trends of this study, noting the growing interest from computer scientists and artists alike, as well as the participation of significant corporations in the sector. From a technical and melodic standpoint, they examine several generation designs and point out areas that need more investigation. All things considered; the report offers insightful information about the state of AI-based music generation today. (17) In order to improve writing effectiveness and quality, Pokhrel, S., & Banjade, S. R's survey investigates the incorporation of the OpenAI language model, specifically GPT-3, into a content generating tool. Using the large corpus of text data that the model was trained on, the application helps users create high-quality content for a variety of platforms, such as blog posts and Facebook advertising. It performs better than conventional rule-based chatbots by using a recurrent neural network (RNN) architecture to provide predictions and insights into language usage and structure that are more accurate. With the use of sophisticated machine

learning algorithms, the tool's dashboard provides an extensive range of features catered to various content kinds, guaranteeing flawless grammar, error-free authoring, and audience-specific customization. It also helps with search engine optimization, which expands the content's audience and boosts traffic.

II. PROPOSED SYSTEM

The proposed system, as shown in figure 1, presents a novel framework for choosing music, combining speech analysis with emotion detection based on facial expressions to provide user-specific music choices that are in line with their present emotional state. With the help of a single convolutional neural network (CNN) model, as shown in figure 2, that was trained on live data streams, the system is able to identify user emotions with high accuracy. The Speech Recognition library makes speech recognition easier, and the OpenCV library makes face expression analysis easier. Unlike traditional approaches, the OpenAI API uses sophisticated language models to provide music recommendations. By fusing AI-driven music selection with emotion recognition, this innovative framework offers a personalised and captivating user experience.

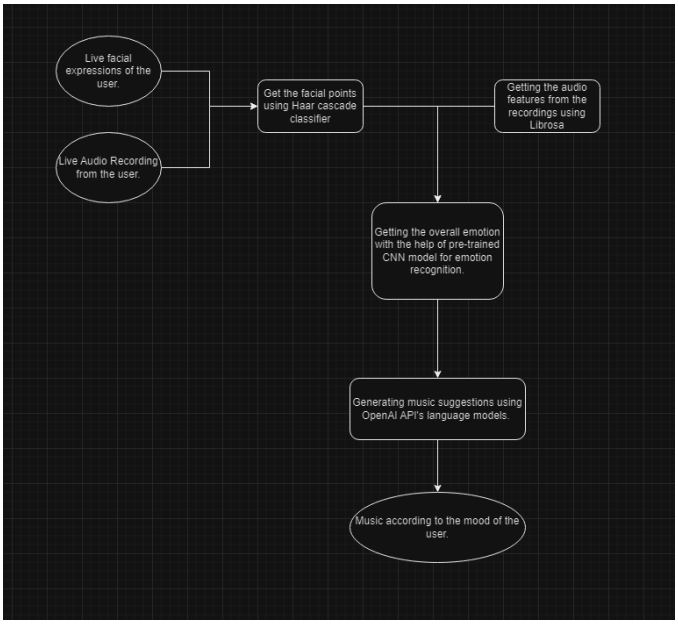

Figure 1: Flowchart of music recommendation after collecting live data.

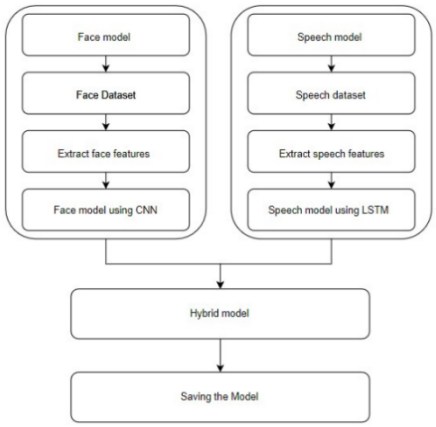

Figure 2: Flowchart of building the hybrid model.

The recommended songs are displayed in a pandas data frame. In the proposed methodology, the first step is to capture live data from the user, which consists of facial expression and speech data. The system then processes the facial expression and speech data to predict the user's emotional state as shown in figures 3, 4 and 5. The prediction is made using a single CNN model, which is trained on emotion-labelled datasets.

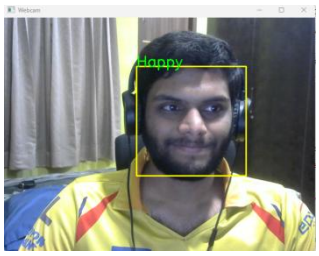

Figure 3: Emotion recognition model: detecting happy emotion after taking in live audio and live face expression.

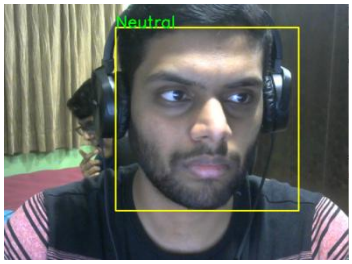

Figure 4: Emotion recognition model: detecting neutral emotion after taking in live audio and live face expression

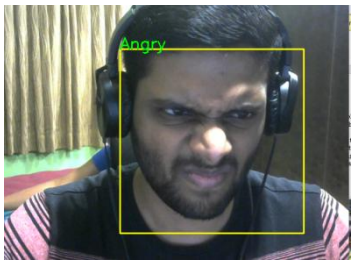

Figure 5: Emotion recognition model: detecting angry emotion after taking in live audio and live face expression

A. Music Recommendation

Unlike other approaches, the OpenAI API is used by the music recommendation system to produce music recommendations by utilising advanced language models. The proposed system guarantees a customised and engaging user experience by integrating AI-driven music selection with emotion identification in a seamless manner. The algorithm accesses a CSV file from Spotify that has music sorted by mood after determining the ultimate feeling. The algorithm suggests music based on danceability, tempo, and valence to best suit the user's mood.

In summary, our proposed system utilizes facial expression and speech recognition to predict the user's emotional state, and recommends music from the user's Spotify playlist based on the predicted emotion. The system employs a single CNN model to predict the user's emotional state. The final emotion is used to recommend music tracks from a large dataset stored In a CSV file, using a OpenAI API. The system provides a new and innovative approach to music recommendation, offering a more personalized and intuitive experience for users.

III. CONCLUSION AND FUTURE WORK

In conclusion, the proposed music recommendation system is a big step forward in using AI technology to improve user experience when listening to music. Through the OpenAI API, the system combines advanced language models with facial expression and speech analysis-based emotion identification to provide personalized music recommendations that correspond with the user's emotional state at that moment. This architecture not only improves user experience but also shows how AI-powered systems can comprehend and react to human emotions instantly. More system optimization and improvement in the future will surely result in even more precise and customized music selections.

Future study and development in this field have a number of options. First, by using sophisticated machine learning approaches to increase the precision and resilience of emotion identification algorithms, the system's performance may be further enhanced. Furthermore, the system's personalization capabilities could be improved by including more extensive music data and improving the recommendation algorithm to take into account elements other than simple mood classifications, like personal tastes and environmental clues. A deeper comprehension of the user's emotional state and more sophisticated recommendations may result from investigating other sources of emotional cues, such as user interaction patterns or biometric data. Further intriguing possibilities for this research include extending the system's interoperability with more music streaming services and investigating creative ways to incorporate user feedback for ongoing improvement.

REFERENCES

[1] Londhe RR and Pawar DV 2012 Analysis of facial expression and recognition based on statistical approach International Journal of Soft Computing and Engineering 2

[2] Kabani H, Khan S, Khan O and Tadvi S 2015 Emotion based music player International Journal of Engineering Research and General Science 3 750-6

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

[4] Hashizume, Y., Li, L., & Toda, T. (2022). Music Similarity Calculation of Individual Instrumental Sounds Using Metric Learning. arXiv preprint arXiv:2211.07863.

[5] ParulTambe, YashBagadia, Taher Khalil and Noor UlAin Shaikh 2015 Advanced Music Player with Integrated Face Recognition Mechanism International Journal of Advanced Research in Computer Science and Software Engineering 5

[6] Vivek JD, Gokilavani A, Kavitha S, Lakshmanan S and Karthik S 2017 A novel emotion recognition-based mind and soul-relaxing system In2017 International Conference on Innovations in Information, Embedded and Communication Systems 1-5 IEEE

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

[8] Turnbull, D. R., McQuillan, S., Crabtree, V., Hunter, J., & Zhang, S. (2022). Exploring Popularity Bias in Music Recommendation Models and Commercial Steaming Services. arXiv preprint arXiv:2208.09517.

[9] A. habibzad, ninavin and Mir kamalMirnia A new algorithm to classify face emotions through eye and lip feature by using particle swarm optimization 2012 4th International Conference on Computer Modeling and Simulation

[10] Rabashette, M. D., Tale, M. R., Hinge, M. A., Padale, M. K., Chavan, M. R., & Deshmukh, N. (2016). Emotion Based Music System. Int. J. Emerg. Trends Sci. Technol.

[11] Florence, S. M., & Uma, M. (2020, August). Emotional detection and music recommendation system based on user facial expression. In IOP Conference Series: Materials Science and Engineering (Vol. 912, No. 6, p. 062007). IOP Publishing.

[12] Sakore, S., Jagdale, P., Borawake, M., & Khandalkar, A. Music Recommender System Using ChatBot.

[13] Yang, J. (2022). Personalized Song Recommendation System Based on Vocal Characteristics. Mathematical Problems in Engineering, 2022.

[14] Arkachaari, V., Rao, P. M., & Darapaneni, N. (2022). Bi-Sampling Approach to Classify Music Mood leveraging Raga-Rasa Association in Indian Classical Music. arXiv preprint arXiv:2203.06583.

[15] Rozhevskii, D., Zhu, J., & Zhao, B. (2022). Psychologically-Inspired Music Recommendation System. arXiv preprint arXiv:22

[16] Civit, M., Civit-Masot, J., Cuadrado, F., & Escalona, M. J. (2022). A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. Expert Systems with Applications, 118190.

[17] Pokhrel, S., & Banjade, S. R. (2023). AI Content Generation Technology based on Open AI Language Model. Journal of Artificial Intelligence and Capsule Networks, 5(4), 534-548.