International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Music Implication and Suggestion System with Emotion Detection

Mayank Sharma
Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India
ms5194@srmist.edu.in
RA1911003011027

Rohan Chhatwal
Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India
rc9574@srmist.edu.in
RA1911003011029

Dr. K. Geetha
Assistant professor
Computer Science and Engineering SRM Institute of Science and Technology Chennai, India
geethak5@srmist.edu.in

Abstract*— In the past, choosing music based on one's mood required human interaction. However, advancements in computer vision technology now allow for the automation of this process. This article presents a real-time system for facial feature extraction and emotion recognition, which involves comparing selected facial features to a face database to identify emotions from images. This field of research is active in image processing and human-computer interaction. The system employs a convolutional neural network to detect and classify human emotions from dynamic facial expressions in real-time. The model is trained on the FER dataset, which contains over 30,000 facial RGB images depicting different expressions. Expression-based music players utilize this technology to analyze data and generate playlists based on the user's facial expressions. This feature enhances the traditional music player experience.

Keywords*— Emotion Recognition, Face Detection, Music recommendation, Deep Learning, Supervised Learning, Machine Learning etc.

I. INTRODUCTION

Music is a powerful force in people's lives, providing a respite from their everyday routines and linking directly to their emotions and feelings. Emotions are subjective and private physiological and mental states, characterized by a range of behaviors, thoughts, and feelings, and are integral to computer vision research. Emotions often facilitate interactions and mediate between people, and comprehending them adds context to seemingly puzzling social communications. Facial expressions provide a simpler and more practical way to recognize emotions, with seven universal emotions - anger, disgust, fear, happiness, sadness, surprise, and contempt - identified across different cultures. Recognizing emotions based on facial expressions is a hot topic in various fields that provides solutions to various challenges. Music enthusiasts often face the frustrating experience of not finding songs that match their mood or situation. Despite advancements in multimedia and technology, such as variable playback speed, genre classification, and volume modulation, users still have to manually browse through playlists to select songs based on their current mood and behavior. Thus, there is a need for a system that can automate the process of selecting music based on a person's mood and alleviate the burden of manual browsing.

II. LITERATURE SURVEY

S Metilda Florence and M Uma (2020) put forth a research paper titled "Emotional Detection and Music Recommendation System based on User Facial Expression" proposing an innovative system that identifies a user's emotional state through facial expressions and recommends music that aligns with their mood. By analyzing facial landmarks, the system can classify the user's emotional state and suggest songs that best express their current mood. This can be especially beneficial for users looking to alleviate stress levels by listening to music without having to spend time searching for suitable tunes. The proposed architecture comprises three components: the Emotion-Audio extraction module, Audio extraction module, and Emotion-Emotion extraction module. However, the accuracy of this system may be limited by the smaller picture dataset employed, requiring well-lit environments and a minimum image quality of 320p for the classifier to produce reliable results. In real-world settings, handcrafted features may lack generalizability, posing challenges for the system's practical application.

In 2021, H. Immanuel James, J. James Anto Arnold, J. Maria Masilla Ruban, and M. Tamilarasan presented a paper titled "Emotion Based Music Recommendation," which proposed a method for creating emotion-based music players that analyze facial expressions to generate playlists. This approach aims to minimize the tedious work of manually organizing music into distinct lists by producing a suitable playlist based on an individual's emotional traits. A linear classifier is employed for face detection, and a facial landmark map of a given face image is constructed based on the pixel intensity values indexed of each point using regression trees trained with a gradient boosting approach.
Emotions are categorized into four types: happy, angry, sad, and surprised, and are classified using a multiclass SVM Classifier. However, the paper highlights the limitations of the proposed system due to the limited availability of images in the image dataset, which makes it currently unable to accurately record all emotions. Additionally, handcrafted characteristics are often insufficiently generalizable in real-world scenarios.

Ali Mollahosseini, Behzad Hasani and Mohammad H. Mahoor (2017) introduced "AffectNet," a database for facial expression, valence, and arousal computing. This database contains over 1 million facial images that were obtained from three major search engines using emotion-related keywords in six different languages. The images were manually annotated to indicate the existence of seven facial expressions and the strength of valence and arousal in about half of them. Two baselines were proposed for categorizing pictures in the categorical model and predicting the value of valence and arousal in the continuous domain of the dimensional model. However, there were certain limitations, such as the fact that the VGG16 model only improved on AlexNet by replacing larger kernel-sized filters with multiple smaller ones, and the AffectNet database did not contain very powerful samples, i.e., those with arousal of 1 or -1 and valence of 1 or -1.

A. Maintaining the Integrity of the Specifications

In contrast to traditional methods of finding new music, recommendation systems do not require any additional effort from the user. Once a user creates playlists and listens to their preferred music, they simply need to install the application and register. The more information the system has about the user's preferences, the better it is at providing personalized music suggestions. As the accuracy of the algorithms improves, users can expect to experience an increase in their enjoyment of music and their overall satisfaction with the streaming service.

B. Objectives of Proposed System

The main aim of this project is to achieve two objectives, which are emotion recognition and music recommendation based on facial emotion. To accomplish this, the user takes an image that is fine-tuned for feature extraction using image processing techniques.

The emotion recognition phase involves detecting the user's image through the camera and classifying it into one of the seven universal expressions, such as happiness, sadness, anger, surprise, disgust, fear, and neutrality, which are labelled in the FER2013 dataset. To train the system, convolutional neural networks (CNNs) were used, which have shown high efficiency in image processing. The dataset was split into training and test datasets and then trained on the training set, without any feature extraction process before feeding it into CNN.

The music recommendation phase takes place after facial emotion detection, where each song is played from the playlist based on the user's mood, such as sad, angry, chilled-out party, or happy.

III. PROPOSED WORK

A collaborative filtering system aims to predict a rating or preference from a user. The objectives of this approach are:

- Our project aims to create a music recommendation engine by utilizing a dataset sourced from Kaggle that includes numerous song titles, artists, and lyrics. Additionally, we plan to scrape data from LyricsFreak to enhance the quality of our dataset

- To build a music recommendation system using collaborative filtering, we will utilize the Song Dataset, which is an open-source collection of audio attributes and metadata for over one million recordings from various popular music genres.

As digital audio formats have proliferated, the effective organization and search for music have become increasingly important. Although there have been successful advancements in music information retrieval (MIR) methods over the last decade, music recommender systems are still in their early stages. This study aims to provide a comprehensive overview of a broad framework and state-of-the-art methods for music recommendation. Collaborative filtering (CF) and content-based models (CBM) are two common algorithms that have proven effective. Two user-centric methods, the context-based model and the emotion-based model, have gained popularity due to the difficulty of locating songs in the long tail and the powerful emotional connotations in music.

We provide an overview of the fundamentals of music recommendation, such as user modelling, item profiling, and matching algorithms. We discuss six distinct types of recommendation algorithms and highlight four potential issues that may arise in the user experience. Despite advancements in music recommendation, subjective recommendation systems are still in their infancy, and we suggest a paradigm shift towards intrinsic motivation by drawing from research in human behaviour, physical education, and music psychology.

Our project involves analysing a compact music database to deliver personalised song suggestions based on a user's past listening history and the songs they have played. We will leverage tools like NumPy and Pandas to achieve this. We will employ Cosine similarity in addition to Count Vectorizer. Furthermore, we intend to create a front-end that displays recommended songs once a specific piece of music has been processed.

Deep learning is gaining popularity in the field of music recommendation systems, where deep neural networks are utilized to learn sequential patterns of music items such as songs or artists based on audio signals or metadata. DL is also used to extract latent features of music items from audio signals. Instead of using latent item variables, sequential music recommendation employs sequence models of music items, like automated playlist continuation.

This system explores the complexities of recommendation systems in the music domain and provides an overview of the latest advances in using deep learning techniques for music recommendation. The presentation is structured around various parameters, such as the type of neural network, data sources, recommendation techniques (collaborative, content-based, or hybrid), and the task at hand (standard or sequential music recommendation). Furthermore, the presentation covers major challenges in music recommendation systems and how they are linked to recent advancements in deep learning.

To solve the information retrieval issue of music recommendation, we aim to create a system that considers the actual content of the songs. Unlike other systems, our approach prioritizes acoustic similarity between musical compositions as the key metric. In this study, we compare and contrast two approaches for building a content-based music recommendation system. The first method involves analyzing auditory properties, which is a commonly used approach. The second approach involves using deep learning and computer vision techniques to enhance the performance of the recommendation system.

IV. COMPARISON OF EXISTING METHODS WITH MERITS AND DEMERITS

A. Advantages

- Because the recommendation process is less complex in its initial stages and does not require any additional user information, it is easier to provide suggestion services.

- The primary advantage of this system is its capability to recognize prevalent music preferences among a particular social group and subsequently endorse such music and artists to its members. Many products, not limited to artistic ones, employ it as their default item-to-item recommendation mechanism.

- Individuals with particular interests can utilize the recommendation feature of the service.

- It can recommend unique or unconventional tracks.

- Content attributes can be utilized to produce a collection of suggested readings.
- The approach can be put into practice with minimal exertion.

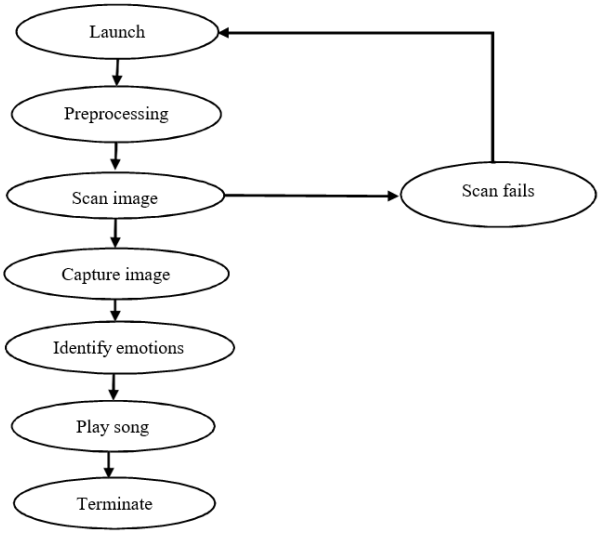

Fig1. System Methodology

B. Disadvantages

- Converting material into valuable characteristics for evaluation by the system can be difficult.
- To perform content-based recommendations, it is crucial to have well-defined structural characteristics of the material and the user's preferences represented in a characteristic form.
- The content-based filtering method, which utilizes machine learning to comprehend and determine the user's preferences, is not suitable for analyzing media content such as music, movies, etc.
- Music,for instance cannot be scrutinized based on data related to its properties.Nevertheless,it is challenging to determine the quality of the suggestions made by content-based.

C. Challenges to Address

- Time taken to do that analysis is very large
- Modern equipment will be required forproper analysis
- Dataset is not properly available on the Web.

V. METHODOLOGY

This project is comprised of several steps, including data preparation, feature extraction using deep CNN, visualization, training and testing of data, identifying emotions based on facial expressions, and playing music that aligns with the detected emotion. The first step involves preparing the necessary data, while the second step involves extracting features from the data using deep CNN. Visualization is used to better understand the data and its features. Next, the data is trained and tested to improve accuracy. The system then identifies emotions based on the facial expressions of the user. Finally, based on the detected emotion, the system selects and plays music that aligns with the user's emotional state.

A. Dataset

To implement the FER system, the FER2013 dataset from Kaggle was utilized. The dataset contains 35,887 annotated images, which were divided into 3589 test images and 28709 training images. Additionally, the dataset includes 3589 private test images used for the final test during the challenge. The images in the dataset are 48x48 in size and grayscale. The dataset comprises images captured from various angles, under different lighting conditions, and at different scales. For more information about the dataset, refer to Table 1 which outlines its description.

Emotions	No. of images
Angry	4593
Happy	8989
Sad	6077
Fear	5121
Surprise	4002
Neutral	6198
Disgust	547

Table1: Description of FER dataset

B. *Facial Expression Recognition Process*

This project involves three main phases in the FER process. Firstly, in the pre-processing phase, the dataset is prepared to be compatible with the generalized algorithms that produce efficient results. Then, in the face detection phase, real-time images are captured and faces are detected. Finally, in the sentiment classification phase, a CNN algorithm is implemented to classify input images into one of seven categories. These phases are illustrated in a flowchart depicted in Figure 2.

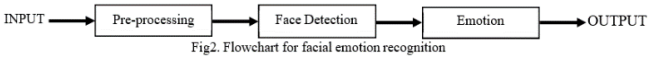

Fig2. Flowchart for facial emotion recognition

C. *Pre-processing*

To ensure accurate and efficient results from the FER algorithm, pre-processing is required since FER input images may contain noise, lighting variations, size, and color. To achieve this, several pre-processing techniques are utilized. Firstly, the image is converted to grayscale, as color images are harder to process with algorithms. Secondly, image normalization is performed to remove lighting variations and enhance facial details. Lastly, the image is resized to eliminate unnecessary parts of the image, which reduces memory requirements and computation speed. These pre-processing operations are essential in preparing the input data for the FER algorithm.

D. *Face Detection*

In the FER system, frames are first passed through a hair cascade classifier, which is a feature-based method for detecting human faces. This classifier is already implemented as a built-in function in OpenCV. Once the classifier detects the face regions, they are cropped from the original frame and converted to grayscale.

E. *Emotion Classification*

The goal of the emotion classification phase is to categorize images into one of the seven universal expressions (happiness, sadness, anger, surprise, disgust, fear, and neutrality) as defined in the FER2013 dataset, using CNNs, which have a proven track record of productivity in image processing. Prior to inputting data to the CNN, no feature extraction was conducted. The process is divided into the following steps:

1. **Data Split:** The FER2013 dataset was divided into training, public testing, and private testing sets based on the "usage" label. The training and public testing sets were used to create the model, while the private testing set was used to assess the model's performance.

2. ***Training and Model Creation:*** The neural network architecture included the following layers:

a) Convolutional Layer: Randomly instantiated learnable filters slide over the input in a convolutional layer. This operation generates a 3D volume consisting of multiple filters, also known as feature maps.

b) Max Pooling: To decrease the spatial size of the input layer and reduce the input size and computational cost, pooling layers are used.

c) Fully Connected Layer: Each neuron in the previous layer is connected to an output neuron in a fully connected layer. The final output layer size corresponds to the number of bins into which the input image is classified.

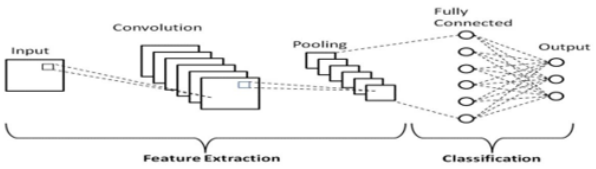

Fig3. CNN model layers.

d) Stack Normalization: Stack normalization accelerates the training process by adjusting the mean activation closer to 0 and the activation standard deviation closer to 1.

In the model evaluation phase, the model created during the training phase was tested on a validation set of 3589 images. To classify emotions in real-time captured images, the technique of transfer learning can be used. The pre-trained weights and values from the model generated during the training process can be utilized for implementing new facial expression recognition problems. As a result, FER can process real-time images faster as the model already contains the necessary weights.

F. *Feature Extraction*

In this stage, the system computes all the features that are extracted to determine the location of the eyes, mouth, and nose on the face of an individual. This computation is used to detect any facial movements.

G. *Emotions Detection and Music Recommendation*

The CNN classifier is used to detect three basic emotions,

Happy, Neutral, and Sad, by applying it to the extracted features. Once the emotion is detected, a corresponding song from the playlist is played based on the user's mood. The mood can be categorized as sad, angry, party, relaxed, or happy.

VI. CONCLUSION

Facial features play a significant role in emotion recognition, and using Convolutional Neural Networks (CNN) is a viable solution for detecting emotions. As facial expression recognition technology becomes more accurate, it can be applied in various domains, including innovation and improving the accuracy of events. In this paper, we introduce an emotion detection model that suggests music based on the user's mood.

REFERENCES

[1] https://towardsdatascience.com/the-keys-building-collaborative- filtering-music-recommender-65ec3900d19f

[2] https://musicoveryb2b.mystrikingly.com//blog/music - recommendation-systems-at-work#:~:text=It%20is%20powerful%20because%20it,only%20i n%20the%20creative%20industry.

[3] https://towardsdatascience.com/the-abc-of-building-a-music- recommender-system-part-i-230e99da9cad

[4] https://www.sciencedirect.com/science/article/pii/S1877050919 310646

[5] https://link.springer.com/article/10.1007/s10844-005-0319-3

[6] https://www.kdnuggets.com/2019/11/content-based-recommender-using-natural-language-processing-nlp.html

[7] https://www.researchgate.net/figure/Comparative-analysis-of- recommendation-techniques-on-the-basis-of- Accuracy_fig3_274096470

[8] Florence, S. Metilda, and M. Uma. "Emotional Detection and Music Recommendation System based on User Facial Expression." IOP Conference Series: Materials Science and Engineering. Vol. 912. No. 6. IOP Publishing, 2020.

[9] Vinay p, Raj p, Bhargav S.K., et al. “Facial Expression Based Music Recommendation System” 2021 International Journal of Advanced Research in Computer and Communication Engineering, DOI: 10.17148/IJARCCE.2021.10682

[10] Zeng Z, Pantic M, Roisman GI and Huang TS 2008 A survey of affect recognition methods Audio, visual, and spontaneous expressions IEEE transactions on pattern analysis and machine intelligence

[11] James, H. Immanuel, et al. "EMOTION BASED MUSIC RECOMMENDATION SYSTEM." EMOTION 6.03(2019).

[12] 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

[13] Hui-Po Wang, Fang-Yu Shih,“Detect and Transmit Emotions in Online Chat using Affective Computing”, National Tsing Hua University, Hsin Chu, Taiwan,2020.

[14] Yading Song, Simon Dixon, Marcus Pearce, “EVALUATION OF MUSICAL FEATURES FOR EMOTION CLASSIFICATION”, University of London, ISMR, 2012.

[15] Aayush Bhardwaj ; Ankit Gupta ; Pallav Jain ; Asha Rani ; Jyoti Yadav "Classification of human emotions from EEG signals using SVM and LDA Classifiers", 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)