Facial Recognition and Mood Detection and Recommend Songs

Harsh Bansal, Aditya Joshi, Archit Sharma, Preet Arnav Singh, Ms. Harashaleen Kaur
CSE-AIML, Chandigarh University Mohali, India

Abstract--When faced with a large selection of music, it can be challenging for individuals to decide what to listen to. Fortunately, there are various recommendation systems that can help users find what they're looking for. With this music recommendation system we aim to provide users with options that match their preferences. By analyzing a user's facial expression and emotions, we can gain a better understanding of their current mental and emotional state. This information can be leveraged to offer music recommendations that are tailored to each individual's preferences.

Keywords--Facial recognition, mood detection, artificial intelligence, personalized experience, music streaming services, emotional state, physiological cues, accuracy, potential risks, ethical implications, CNN (Convolutional Neural Network), ANN (Artificial Neural Network), Grayscale.

I INTRODUCTION

Facial recognition and mood detection are two rapidly advancing areas of artificial intelligence with the potential to revolutionize numerous industries. These technologies use unique facial features and physiological cues to identify and authenticate individuals and analyze emotional states, respectively.

Facial recognition technology utilizes algorithms to analyze facial features like the distance between the eyes, the shape of the nose, and facial contours to identify individuals. While this technology has been used in various sectors like security, banking, and law enforcement, concerns about privacy and security breaches have also been raised.

Mood detection involves analyzing a person's facial expressions, voice, and physiological cues to identify their emotional state. Mood detection technology can improve mental health interventions, such as detecting signs of depression or anxiety, and can also personalize marketing and advertising content.

One interesting application of these technologies is recommending music based on a person's mood. AI algorithms can analyze a person's facial expressions, voice, and physiological cues to suggest songs that match their mood. For instance, if someone is feeling happy, they might be recommended upbeat, cheerful music, while someone who is feeling sad might be recommended more somber, reflective tunes. Music streaming services have already implemented this technology to provide a more personalized experience to their users.

However, there are concerns about the accuracy of mood detection technology, as it may not always accurately reflect a person's emotional state. Also, the potential for misuse of these technologies, such as using mood detection to manipulate consumer behavior, needs to be addressed.

In conclusion, facial recognition and mood detection technologies have the potential to transform many industries. Recommending music based on a person's mood is just one example of the exciting applications of these technologies. However, it is crucial to consider the ethical implications of using these technologies and address potential risks and limitations to ensure their responsible use.

II LITERATURE SURVEY

In recent years, several studies have explored the use of artificial intelligence and image processing techniques to develop emotion-based music players. Londhe et al. [1] used ANN -Artificial Neural Networks for classifying emotions based on changes in facial curvatures and corresponding pixel intensities. Additionally, they proposed playlist-generation approaches. Zheng et al. [2] identified two main categories for facial feature extraction based on

appearance and geometry. Nikhil et al. [3] developed a system that determines a user's emotional state using facial expressions and plays music accordingly. To address the issue of playing songs randomly, they proposed a music player based on emotions that automates the process of selecting appropriate songs. Zeng et al. [4] reviewed various audio/visual computing methods for affect recognition, including approaches to handle audio and video recordings in affective states, and presented a prototype of emotion categories. Tambe et al. [5] proposed a method that learns user preferences, emotions, and activities to suggest song selections based on recorded facial expressions. Jha et al. [6] developed a music player based on emotions using image processing techniques to suggest suitable songs based on the current expression of the user. Anukritine et al. [7] developed an algorithm to suggest songs from the playlist of the user that is based on their emotions and proposes an approach to retrieve audio information that extracts information from audio signals in less time. Aditya et al. [8] developed an Android based application that uses image processing technique in order to analyze and suggest songs to the user based on their mood. Habibzad et al.
[9] presented a new algorithm for facial emotion recognition that contains three stages: Pre-processing, Feature extraction, and Classification. Prof Deshmukh and her team aimed to develop a system that detects the user's emotions using a camera and applies an emotion detection algorithm to automate the results, which takes an average of 0.95–1.05 seconds to generate a music system based on emotions. Liu et al. [10] developed a system that uses BCI - Brain- Computer Interfaces in order to monitor the cognitive state of a person, but requires contin. In recent years, several studies have explored the usage of Artificial Intelligence and image processing techniques to develop emotion-based music players. Swati Vaid et al. [11] reviewed EEG (electroencephalography) and its potential for estimating the emotion of a person, but noted that this method is not portable or cost-effective. In recent years, several studies have shown the use of artificial intelligence and image processing techniques to develop emotion-based music players. Ankita Mahadik et al. [12] proposed a mood-based music player system using MobileNet model with Keras which performs real time mood detection and suggests songs as per the detected mood. Chen et al. [13] proposed a method using deep learning which uses a combination of CNNs and Recurrent Neural Networks (RNNs) for real-time emotion recognition. He et al. [14] developed a model which combines facial expression analysis with collective filtering to recommend songs in real-time. The model uses facial expressions to deduce the user's mood and collective filtering techniques to recommend songs based on the mood profile. Li et al. [15] proposed a system that uses multimodal approach that combine facial expressions with audio features that gets extracted from songs.

III.DATA GATHERING AND DATA SOURCE

A survey was carried out among music listeners, which included questions about their song preferences based on different emotional states. The survey focused on three specific parameters: the types of songs that respondents preferred to listen to when feeling happy, the kinds of songs they would choose when feeling sad or lonely, and the types of music they preferred when feeling angry.

Facial recognition classifiers are trained using the CK+ dataset and the HELEN dataset. The CK+ dataset comprises some facial action coded sequences from individuals, and the subjects are labeled according to the expression they are conveying. Although the dataset captures images of the subjects' neutral expression to their most intense emotion, only the initial and final images are utilized for the analysis and network training. Meanwhile, the HELEN dataset consists of approximately 200 photos, each of which is accompanied by 164 landmark positions that are employed to train the classifier.

IV PROPOSED ARCHITECTURE-

The proposed system is designed to identify the facial expressions of the user and extract facial landmarks from those expressions. These landmarks are then used to categorize the user's sentiment. Once the emotion is determined, the system will present songs that match the user's emotions.

V METHODOLOGY

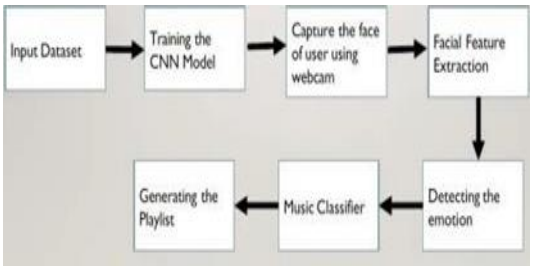

Fig.1. Block Diagram

A. Emotion Extraction Module

The proposed system incorporates an emotion extraction module that captures the user's image using a camera or webcam. To enhance the classifier's accuracy in identifying the face, the captured image is transformed into a grayscale image. The image is then analyzed by a classifier algorithm that employs feature extraction methods to extract distinctive facial features and detect the user's emotions. The classifier is trained on a dataset like CK+, which enables it to recognize emotions based on facial landmarks. When presented with a new set of images, the classifier can detect facial landmarks and provide their coordinates based on the training it has received. By following this process, the system can determine the user's emotions and suggest songs that match their mood.

CNN Convolutional Neural Network: It is a subtype of ANN (Artificial Neural Network) that is specifically designed for image processing and feature extraction. CNNs have become popular in the field of computer vision and are widely used in various applications like image classification. CNNs also demonstrates the success in processing natural language for text classification. Convolutional Neural Network is a flexible tool that can be used in various ways. For example, it can be used in preprocessing input images for face detection and recognition, where it can extract the facial features and reduce the dimensionality of the images. Additionally, CNN can be utilized to classify a user's facial expressions into various categories of emotions, such as happy, sad, angry, and so on. Finally, CNN can also be used to match a user's mood with the mood of the songs, where a similarity measure or distance metric can be used. It contains three primary layers: Convolutional layers, Pooling layers, and Fully-Connected layers. The Convolutional layer are responsible for identifying features in the input image through the use of filters or kernels. The Pooling layer then simplify the feature maps by reducing their size and complexity. Finally, the Fully-connected layer perform the classifications based on the features that have been extracted from the input image.

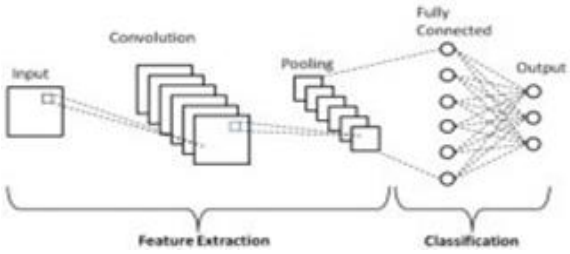

Fig.2. CNN Layers

Grayscale: Grayscale is a term used to describe a series of gray shades ranging from white to black that represent the level of light intensity in an image or display. Grayscale images don't have any color information, but only contain data about the brightness level of an image. Besides, grayscale can also be used to convert color images into black-and-white images by applying a weighted average of the color channels. Grayscale is a useful tool in image processing for various purposes. For instance, it can be used to preprocess input images for face detection and recognition, thereby reducing the computational complexity and noise in the images. Additionally, grayscale can also be utilized to extract features from a user's facial expressions, such as the position and shape of the mouth, eyes, and eyebrows, which can indicate their mood. Finally, grayscale can be used to create a visual representation of the mood of a song or user, where darker shades represent negative emotions and lighter shades represent positive emotions.

B. Audio Extraction Module

Once the emotion extracted from the image, the user is provided with a list of songs based on the identified emotion. The user can listen any song from the list. The frequency with which the user listens to the songs determines their display order in the list. PHP (Hypertext Preprocessor), MySQL, HTML, CSS, and JavaScript are the web technologies used in developing this module.

C. Emotion - Audio Integration Module

The system saves the emotions that are being extracted and displays the songs based on those emotions on a web page. For instance, if the detected

emotion is classified as happy, then songs are displayed to the user from the happy database. The same can be done for all types of moods.

VII.EXPERIMENT RESULTS & ANALYSIS RESULTS

The presented work proposes a music recommendation system that captures users with a camera mounted to a computing platform. The collected image is converted to grayscale to enhance the performance of the facial recognition classifier. Based on facial landmarks, the classifier system extracts the user's face from the webcam feed and determines their moods. The HELEN dataset, consisting of over 2,000 images, is used to train the classifier to detect facial landmarks on new and unknown images. The CK+ dataset is used to train the network to recognize the user's emotions. Once the system detects the user's emotion, the music player selects suitable music that fits the user's mood, reducing the user's time spent searching for music.

The ultimate goal of the system is the enhancement of the user's pleasure and to lighten their mood. To evaluate its robustness, the system needs to be tested under various lighting conditions. The system is capable of capturing new images of the user to update its training dataset and improve its accuracy. The proposed system utilizes facial landmarks and has been tested in different scenarios, demonstrating an accuracy of over 80 percent for most test cases in emotion classification.

VII RESULTS ANALYSIS

A. Experiment Results-

The consumers are provided with clear instructions on how to execute the emotion prediction process. These instructions ensure that the customers are aware of the necessary steps to carry out the prediction accurately. Following these instructions leads to successful emotion detection in most cases. However, in some cases where the internal emotion is sad but the facial expression appears satisfied, the prediction may fail, as depicted in Figure 3. Thus, the instructions are crucial for ensuring the accurate prediction of emotions. They are explained in detail to the user to avoid any confusion or misunderstanding.

Fig.3. Happy emotion detection

VIII EXPERIMENT RESULTS

The users were not provided with any instructions in this scenario, leading to failures in recognizing the inner emotions or matching them with the facial expressions. Some cases even resulted in emotions matching the user's facial expressions. Figure 4 illustrates the outcome of this situation where instructions were not explained to the user.

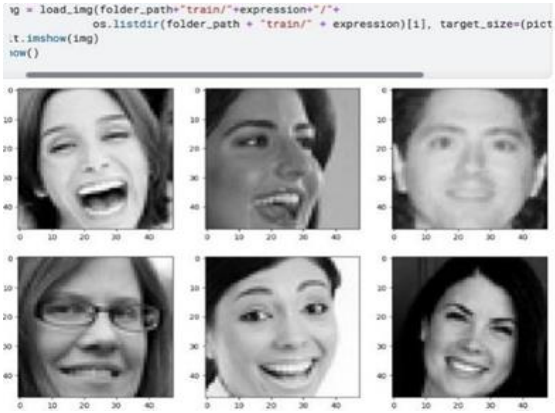

Fig.4. Anger/Disgust emotion detection

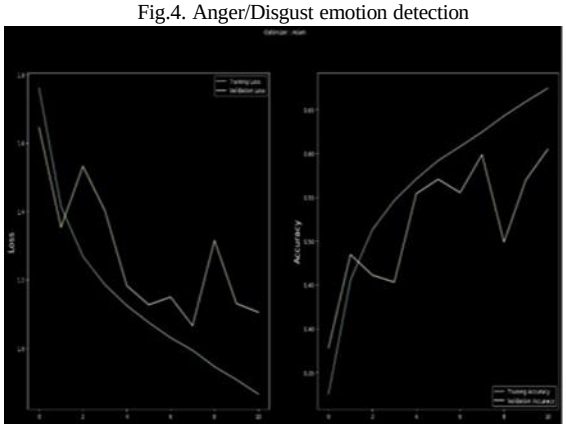

Fig.5. Accuracy Graph

IX CONCLUSION & FUTURE WORK-

Facial expressions have long been a means of nonverbal communication, and it has become crucial to recognize and research the emotions from facial expressions. The use of algorithms such as image processing have been increased to solve the problem of emotion detection. Researchers are exploring different features and processing methods in order to improve the accuracy of emotion classification. Image processing algorithms are of great importance in both medical science and human science, and new methods are being developed regularly to extract emotions of the user and use them for treatment.

Emotion recognition has become important in all the aspects of life. A robust algorithm that can accurately classify emotions can advance the industry significantly. Facial expression recognition can be used in fields such as psychology, sociology, criminology, and marketing, among others. Additionally, it can be applied in healthcare settings to help doctors diagnose and treat patients with mental health disorders.

The use of facial landmarks has emerged as an effective way to detect emotions. Researchers have developed different facial landmarks schemes that extract the key features of a face, such as the position of the eyes, nose, mouth, and eyebrows. These landmarks are then used to identify facial expressions and classify emotions.

Several systems have been developed that use facial landmarks to recognize emotions. These systems have been tested in real-time environments and successfully captured the emotions of users. However, to assess the robustness of these systems, they should be evaluated under different lighting conditions.

To improve the accuracy of emotion recognition systems, certain considerations need to be taken into account. One such consideration is the limited availability of images in the dataset used by the system. The dataset must have enough images of different emotions to ensure that the system can recognize emotions accurately.

Another consideration is the importance of well- lit images to obtain accurate results. Proper lighting is necessary to capture the facial features accurately and ensure that the system can recognize emotions correctly. Finally, the quality of the image should be at least 320 p to ensure that the system can predict the user's emotion accurately.

In conclusion, facial expression recognition has become an important area of research with a numerous applications. Image processing algorithms have been used to develop systems that accurately recognize emotions. The use of facial landmarks has emerged as an effective way to detect emotions. However, to improve the accuracy of emotion recognition systems, the limited availability of images in the dataset, well- lit images, and image quality should be taken into account. By addressing these issues, emotion recognition systems can provide more reliable results and advance the industry significantly.

REFERENCE

[1] Dureha, Anukriti. "An accurate algorithm for generating a music playlist based on facial expressions." International Journal of Computer Applications 100.9 (2014): 33-39.

[2] Yan, R., Yang, M., Zheng, Q., Wang, D., & Peng, C. (2023). Facial expression recognition based on hybrid geometry-appearance and dynamic-still feature fusion. Multimedia Tools and Applications, 82(2), 2663- 2688.

[3] Zhang Z, Lyons M, Schuster M, Akamatsu S. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. InProceedings Third IEEE International Conference on Automatic face and gesture recognition 1998 Apr 14 (pp. 454-459). IEEE. S. R. N. S. M. A. H. Akhand, "Facial Emotion Recognition Using Transfer Learning in the Deep CNN," MDPI, vol. 10, no. 9, 2021.

[4] Zhang Z, Lyons M, Schuster M, Akamatsu S. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In Proceedings Third IEEE International Conference on Automatic face and gesture recognition 1998 Apr 14 (pp. 454-459). IEEE.

[5] Sharma, V. P., Gaded, A. S., Chaudhary, D., Kumar, S., & Sharma, S. (2021, September). Emotion-based music recommendation system. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends

and Future Directions) (ICRITO) (pp. 1-5).

[6] IEEE. R. Shyam, "Enhanced Object Detection with Deep Convolutional Neural Networks," International Journal of All Research Education and Scientific Methods (IJARESM), vol. 9, no. 7, pp. 27-36, 2021

[7] Gunasekar, M., A. Panneerselvam, V. Sneharathna, M. Suganneshan, and K. Logeswaran. "Improved Facial Emotion Recognition using Yolo and DeepFace for Music suggestion." In 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1124-1127. IEEE, 2022.

[8] B Ko**B**rief Review of Facial Emotion Recognition Based on Visual Information Sensors, volume 18, issue 2Posted: 2018

[9] Singh, R., Gehlot, A. and Joshi, A., 2022, November. Study on Facial Recognition to Detect Mood and Suggest Songs. In 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC) (pp. 1078-1084). IEEE. M Slaeney BabyEars: A recognition system fore motive vocalizations Speech Common, volume 39, issue 3-4, p. 367 – 384 Posted: 2003

[10] Song M, You M, Li N, Chen C. A robust multimodal approach for emotion recognition. Neurocomputing. 2008 Jun 1;71(10-12):1913-20.

[11] Ekman, P. & Keltner, D. (1997). Universal facial expressions of emotion: An old controversy and new findings. In Segerstråle, U. C. & Molnár, P. (Eds.), Nonverbal communication: Where nature meets culture (pp. 27-46). Mahwah, NJ: Lawrence Erlbaum Associates.

[12] Matsumoto, D. & Kupperbusch, C. Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience. Asian Journal of Social Psychology (4), pp. 113-131 (2001)

[13] I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[14] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J. Ambadar, Z. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. IEEE Computer Society Conference CVPRW (2010)

[15] Zhang, Z. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. International Journal of Patten Recognition and Artificial Intelligence 13 (6):893-911 (1999).

[16] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998).

[17] Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput. 27(6):803– 816 (2009).

[18] Carcagnì P, Del Coco M, Leo M, Distante C. Facial expression recognition and histograms of oriented gradients: a comprehensive study. Springer Plus. 4:645. (2015).