Music Recommendation Using Facial Recognition

Prof. D.A. Shinde[1], Akshada Borge[2], Sumedh Gawali[3], Aditya Nimbhore[4], Akash Nikam[5], Sarvesh Sharma[6]

Professor, Department of Information Technology, PVGCOE & SSDIOM, Nashik, India
Guide, Department of Information Technology, PVGCOE & SSDIOM, Nashik, India

Abstract
Music plays a vital role in our everyday life. Life without music cannot be imagined. Music changes our mood; Whatever our mood might be, the only thing we do in all of our moods is to listen to music. We also listen to music when working, driving, travelling and even when reading a comic or a story. Music can induce a clear emotional response in its listeners. The pitch and rhythm of the music are managed in the areas of the brain that deal with emotions and mood. Thus, music plays an important role in enhancing our mood. As elders have said "Face is the Index of the Mind", the mood of a person can be known by looking at the face of the person. The abstract of this system/ project is to build an automated system that builds playlists and plays the songs according to the mood of the user by directly discerning the facial emotions of the user. This model requires a camera to capture the face of the user and then the mood of the user is recognized by CNNs. Then the playlist is recommended to the user based on the discerned "Mood" of the user. This disposes of the tedious and monotonous task of physically gathering tunes into various records and helps in creating a suitable playlist dependent on a person's passionate highlights. Hence, the proposed system can be used to build a music recommendation system based on the facial emotion gestures of the user.

***Keywords*: Facial expression, CNN, Facial emotion gestures, Agile methodology.**

I. INTRODUCTION

Recent studies have shown that listening to music can alter how people feel, act, and think. In a study on the motivations behind music consumption, researchers found that connecting emotions and moods is one of music's important functions. The positive effects of music on mood and self awareness are significant. According to research, there is a significant correlation between a person's emotional state and musical tastes [1]. Additionally, musical characteristics like rhythm, tone, timbre, and meter interact with the brain areas in charge of controlling emotions and mood [2]. Personal interaction is an essential part of our daily lives. It enables us to closely examine human behavior, which includes important elements like body language, voice, facial expressions, and emotions [3]. Emotion detection is incorporated into many applications in today's technologically advanced society, including security systems and smart cards. This useful application not only recognizes user moods but also creates playlists that are suited to those feelings. For instance, the system will play cheerful, mood lifting music when someone is feeling down, while it will provide a mix of various musical genres to reinforce positive sensations when someone is feeling upbeat [4]. This system's soundtrack is composed of Hindi songs. It uses Hear-cascade technique, which has an accuracy rate of about 92.10% and is a key technology in this context, for face expression identification. Thanks to advancements in digital signal processing and feature extraction techniques, the field of automated emotion identification in multimedia, which includes music and movies, is developing quickly. This system has the potential to be very important in many applications, such as music entertainment and human computer interaction systems. We suggest creating a face expression-based recommendation system that would provide better user experiences and tailored recommendations.

II. LITERATURE SURVEY

"David Matsumoto" and "Hyi Sung Hwang" published a paper titled "Reading facial expressions of emotion" Emotions are an incredibly important aspect of human life and basic research on emotions of the past few decades has produced several discoveries that have led to important real world applications. This article described two of those discoveries – the universality of facial expressions of emotionand the existence of micro expressions – because of their importance to and novelty in psychology. The paper discussed how those discoveries create programs that teach people how to read facial expressions of emotion, as well as recent research that has validated those training programs and documented their efficacy. "Akshobhya Rao BV" and "Fathima RameeshaAsokan" published a paper titled Emotion Based Music Player (Emotify). Music is a major form of entertainment. Through the advent of technology, much focus has been given to the optimization of manual labor. There are still many traditional music players who need songs to be selected and arranged manually. User, the playlist needs to be generated and modified for every mood which takes time. Some of the music players have advanced features, such as lyrics and assisting the user by suggesting similar tracks. "DegerAyata" and "Yusuf Yuslun" published a paper titled "Emotion Based Music Recommendation". Most of the existing music recommendation systems use collaborative or content based recommendation engines. However, the music choice of a user is not only dependent to the historical preferences or music contents. But also dependent on the mood of that user.
This paper proposes an emotion based music recommendation
framework that learns the emotion of a user from the signals obtained via wearable physiological sensors.

"Asha Sugave" and "Sahil Mulani" published a paper titled "Emotion Recognition from Audio- Visual Data". Emotion Recognition Systems is used to identifying the emotions of humans with their accuracy. This paper using Audio-visual Data to recognizing emotion. This emotion recognition system automatically identifies the human emotional states from his or her voice and face images. An audiovisual emotion recognition system is used to develop uses fusion algorithm. In this system firstly separate emotion recognition systems that use voice and facial expressions were tested separately.

Various methodologies have been proposed to classify the behavioural and emotional state of the user. Mase et al. focused on using movements of facial muscles while Tian et al. [8] attempted to recognize Actions Units (AU) developed by Ekman and Friesen in 1978 using permanent and transient facial features. With evolving methodologies, the use of Convolutional Neural Networks (CNNs) for

emotion recognition has become increasingly popular [9]. Music has also been classified using lyrical analysis [6], [10]. While this tokenized method is relatively easier to implement, on its own it is not suitable to classify songs accurately. Another obvious concern with this method is the language barrier which restricts classification to a single language. Another method for music mood classification is using acoustic features like tempo, pitch and rhythm to identify the sentiment conveyed by the song. This method involves extractinga set of features and using those feature vectors to find patterns characteristic to a specific mood [7], [21]. In this section, we study the usage of convolutional neural networks (CNNs) in the context of emotion recognition [11], [12]. CNNs are known to simulate the human brain when analyzing visuals; however, given the computational requirements and complexity of a CNN, optimizing a network for efficient computation is necessary. Thus, a CNN is implemented to construct a computational model which successfully classifies emotion into 4 moods, namely, happy, sad, angry and neutral, with an accuracy of 90.23%.

There are different ways of emotional analysis by using facial expression, gestures, body movement, speech etc. Many research has been conducted with different approaches for detecting and classifying the physiological behavioural and emotional status expressed on the face by the users. The facial digital image is preprocessed and subjected to different algorithms for feature extraction and classification. In 1978, Ekman and Friesen developed Action Units (AU) by using transient and permanent facial features[1]. Their

work mainly focused on establishing a dependency between movement of facial muscles and expressed emotions due to variations of positive and negative emotional triggers. The framework Facial Action Coding System targeted about 44 action units on the face which were detected and measured for their intensity. Many feature-based algorithms are designed to extract human emotions as recommended by Ekman[2]. In paper [4], the author proposes a geometric based approach to recognize facial expression. Here, the features are extracted based on the movement of facial landmarks which signifythe location of feature points such as eyes, eyebrows, lip corners etc. A feature vector is derived out of the distances between these feature points. Depending on the change in emotional conditions, the distances also change which is tracked with respect to the image of a neutral state. Further, the emotions are classified using SVM (Support Vector Machine) and RBFNN (Radial Basis Function Neural Network) with distance vectors as input to the classifier. In paper [5], both geometric based approach and appearance based feature extraction using Gabor wavelet coefficients is performed using two layer perceptron. Classificationof music based on lyrical analysis is easy but not accurate. The major challenge in this approach is the language barrier that restricts classification of tokens belonging to a single language. Alternatively, emotional features andsentiment conveyed in the music can be classified using acoustic related attributes like pitch, rhythm and tempo [6]. This approach aims at extracting and defining feature vectors corresponding to characteristics of a specific mood. For example, a feature of fast tempo corresponds to an angry emotional state. So, in order to effectively recommend the songs, it is required for the system not only to know the emotions of the user but also know the mood conveyed by the song.

III. HARDWARE AND SOFTWARE REQUIREMENT

Table 1. Hardware Requirements

SR. NO	Hardware	Description
1	Processor	Intel i3 or AMD
2	Speed	1.1GHz
3	Hard Disk Space	40 GB
4	RAM	8 GB

Table 2. Software Requirements

SR. No	Software	Description
1	Operating System	Windows / Linux / MacOS
2	IDE	Vs Code
3	Processor	Intel i3
4	Coding Language	Python

[B]Counting of Object

IV. ARCHITECTURE

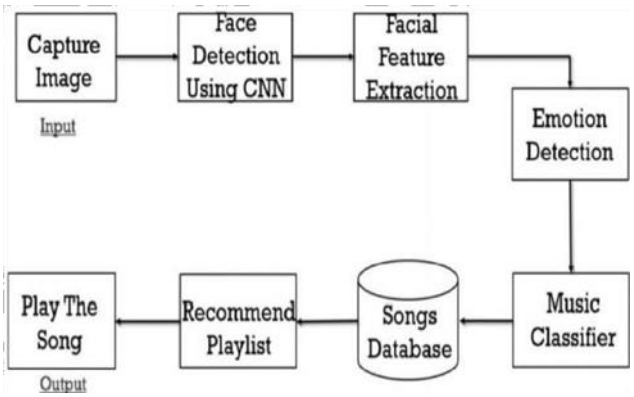

Fig. System Architecture

V. METHODOLOGY

On the start of the system there will be two options for the counting of objects, one will be Generic which will calculate all the objects in the frame and the other one will be specific which will only count specific object as directed by the user. There are mainly two functionalities [A] Object Detection, [B] Counting of Objects.

[A] Object Detection

In the first process, Real-time Capturing is the first step. This is done using Webcam or any other live camera. When we annotate an image in Real-time, we are adding metadata to a dataset. Each image in dataset must be thoughtfully and accurately preprocessed to train an AI system to recognize objects similar to the way a human can. Image preprocessing is a process of dropping of data before it is used to enhance the further performance. The preprocessing is followed by feature extraction which transforms raw data into numerical features that can be processed while preserving the information in the original dataset.

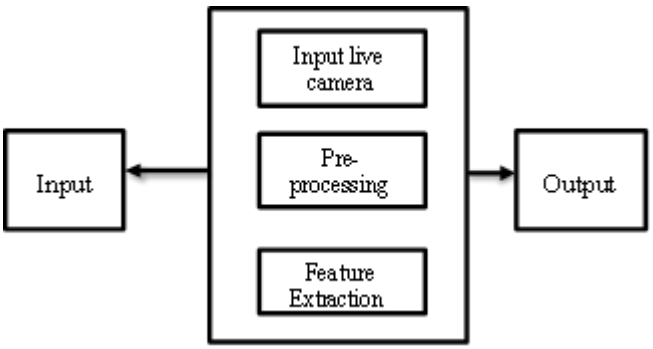

Figure 1: Object detection process

This process starts with segmentation where the image is segmented into multiple image segments; it reduces the complexity of the image and enables further analysis of each image segment. This follows by classification by CNN algorithm which uses neural networks to provide real-time object detection. The object is detected and counted and the result is displayed to the user.

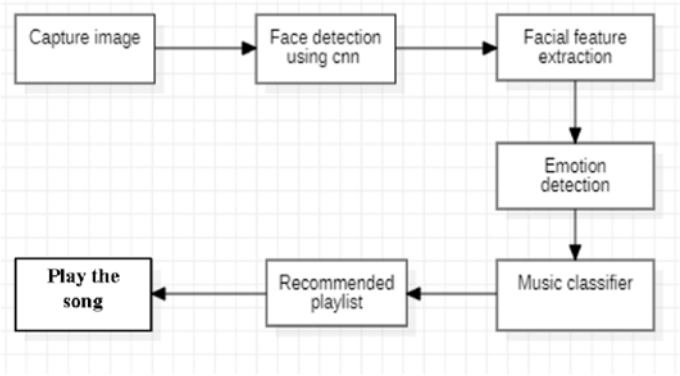

Figure 2: Flow of Process

VII. SUMMARY

This paper discusses four methods being used in the project. The table below gives the limitation of each method.

Method Used	Limitation
Generic	It cannot instantiate generic types with primitive types. It cannot create instances of type parameters.
Specific	Issues with research samples and selection. Limited access to data. Time Constraint
CNN Algorithm	Issues with research samples and selection. Limited access to data. Time Constraint
Agile Methodology	Poor Resource planning. Limited documentation. Fragmented output.

CONCLUSION

The results obtained above are very promising. The high accuracy and quick response time of the application makes it suitable for most practical purposes. The music classification module in particular, performs significantly well; it achieves high accuracy in the "angry" category, while also performing appreciably well in the "happy" and "calm" categories. Thus, EMP reduces user efforts for generating playlists. by efficiently mapping the user's emotion to the correct song class with an overall accuracy of 97.69%, it achieves optimistic results for the four moods studied. We also recognize the room for improvement. It would be interesting to analyze how the system performs when all seven basic emotions are taken into consideration; additional songs from different languages and regions can also be added to make the recommendation system more robust. User preferences can be collected to improve the overall system using collaborative filtering. We plan to address these issues in a future work.

REFERENCES

[1] Swathi Swaminathan and E. Glenn Schellenberg, "Current emotion research in music psychology," Emotion Review, vol. 7, no. 2, pp. 189–197, Apr. 2015.

[2] "How music changes your mood", Examined Existence. [Online].
Available:http://examinedexistence.com/howmusic-changes-yourmood/. Accessed: Jan. 13, 2017.

[3] Kyogu Lee and Minsu Cho, "Mood Classification from Musical Audio Using User Group-dependent Models."

[4] Daniel Wolff, Tillman Weyde, and Andrew MacFarlane, "Culture-aware Music Recommendation."

[5] Mirim Lee and Jun-Dong Cho, "Logmusic: context-based social music recommendation service on mobile device," Ubicomp'14 Adjunct, Seattle, WA, USA, Sep. 13–17, 2014.

[6] D.Gossi and M.H.Gunes,"Lyric-based music recommendation," in Studies in computational intelligence. Springer Nature, pp. 301–310, 2016.

[7] Bo Shao, Dingding Wang, Tao Li, and Mitsunori Ogihara,"Music recommendation based on acoustic features and access patterns," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 8, Nov. 2009.

[8] Ying-li Tian, T. Kanade, and J. Cohn, "Recognizing lower. Face action units for facial expression analysis," in Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Mar. 2000, pp. 484–490.

[9] Gil Levi and Tal Hassner, "Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns."

[10] E. E. P. Myint and M. Pwint, "An approach for mulitlabel music mood classification," in 2010 2nd International Conference on Signal Processing Systems, Dalian, 2010, pp. V1-290-V1-294.

[11] Peter Burkert, Felix Trier, Muhammad Zeshan Afzal, Andreas Dengel, and Marcus Liwicki, "DeXpression: Deep Convolutional Neural Network for Expression Recognition.

[12] Ujjwalkarn, "An intuitive explanation of Convolutional neural networks," the data science blog, 2016. [Online]. Available: https://ujjwalkarn.me/2016/08/11/intuitiveex planatio n-convnets/. Accessed: Jan. 13, 2017.