IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11319

Sensible Portable Player

Jayesh Chauhan[1], Sneha Nagdeve[2], Rajani Meshram[3], Saloni Pillewan[4],

Dr. (Miss.) Uma Thakur[5]

Priyadarshini College of Engineering, Nagpur, Research Scholar, Computer Science and Engineering Department, India [1, 2, 3, 4]

Professor at Nagpur's Priyadarshini College of Engineering's Computer Science and Engineering Department, India [5]

Abstract: Although each individual human has a unique face, their expressions tell us the same tale and play an important role in determining an individual's emotions and behaviour. Music is the purest form of art and a medium of expression, and it is seen to have a stronger emotional connection. It has a unique capacity to make one feel better. This project system aims to create an effective music player that is based on the user's emotion and employs facial recognition algorithms to do so. The extracted facial traits will generate a system, minimising the effort and time required to accomplish so manually. A camera is used to record facial data. Deep learning algorithms are used by the emotion module to determine the exact mood associated with a given phrase. For real-time film, the mood detection module in the system has an accuracy of over 80%, while static images have an accuracy of 95 to 100%. As a result, it produces greater precision in terms of time and performance.

Keywords: Computer Vision, Deep Learning Techniques, Face Recognition, Emotion and Mood Detection, Mood Extraction Module, Computer Vision

I. INTRODUCTION

Human emotions must be expressed and identified in communication systems. Humans are capable of expressing and recognising emotions. Computers utilise image analysis or sensors to recognise human emotions. We deal with many individuals in our daily lives and in our professional lives, either directly or indirectly through phone calls, and it is sometimes vital for people to be aware of the current emotions of the person with whom they are talking. The various sorts of human emotions include surprise, fear, fury, happiness, sadness, disgust, and neutral. When it comes to communicating emotions, facial movement and speech tone are crucial. The energy in the emission of words is TOLD by the physicality and tone of the face, which can be adjusted to transmit various emotions. These variations in signals, as well as the information perceived by any other sensory ORGAN, are easily recognised by humans. This research looks at how images, sensors, and speech may be used to capture emotions. Music is a crucial means of enjoyment for music lovers and listeners, and it may occasionally even provide a therapeutic method. Where words fail, music talks, and as a result, it can transform a PERSON'S negative emotion into a pleasant mood simultaneously and gradually. Gestures, voice, facial expressions, and body language are all examples of how emotions can be expressed. We use facial expressions to help the system figure out how the user is feeling. Using the camera on the mobile device, we may capture the user's facial expression. There is a plethora of emotion detection systems that use collected photos to assess the emotion. In this application, we use neural networks to recognise emotions.

II. SYSTEM REQUIREMENTS

The minimum prerequisites for developing this application are as follows:

1. Hardware requirements

- PROCESSOR: 2 GHz
- RAM: 1 GB

2. Browser

- Chrome 51 or higher
- Firefox 47 or higher
- Opera 37

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11319

● Edge 105 3. Database

● Firebase

● NoSQL 4.

API: Affective Emotion Recognition API Page Style

III. REVIEW OF LITERATURE

There are a variety of applications that have facilities and services for creating music playlists or playing a certain song, and this procedure requires all human labour. To classify human emotional states of behaviour, numerous strategies and approaches have been suggested and developed. Using complicated methodologies like Viola and Jones, the proposed approaches have only concentrated on a few of the basic emotions.

The publications listed above provide a general overview of what methods have been taken and demonstrate techniques and algorithms such as machine learning with support vector mechanisms for recording facial expression and producing a playlist for the user.

Several research papers giving a brief about the idea are:

[1] According to the authors, recording, playing, processing, and managing digital audio is simple. Because of its widespread use, gadgets for handling it are inexpensive, allowing more people to record and play music and voice. Furthermore, the internet has simplified the process of obtaining recorded sounds. As a result, the number of people who own recorded music has significantly expanded. Most current audio players compress audio files and store them in internal memory. Because storage costs have steadily decreased, the amount of music that will be stored has exploded. If each song is saved in compressed format and contains 5 Mbytes, a player with 16 Gbytes of memory may carry around 3,200 songs. It's tough to effectively organise such enormous amounts of music. People tend to listen to a small selection of favourite songs over and over again, while others are unjustly disregarded. Affection is a method for organising music collections that we created. Affection gathers pieces of music that reflect similar feelings and assigns a symbol to each category. These symbols make it simple for listeners to choose music that matches their mood.

[2] The authors of this research claim that music is omnipresent throughout our lives. People hear music in many situations, either actively or passively, and consciously or unconsciously feel it as a kind of emotion expression. In this paper, we describe a new location and emotion aware web-based interactive music system. Its goal is to provide the user's favourite music while also keeping track of their whereabouts and emotions. The method starts by offering advice based on professional expertise. If the user doesn't like the recommendation, he or she can ignore it and chose the music on his or her own. During the process of learning music preferences, the user's interactions with the system, current location, and emotion are logged. As a result, the system may adjust to the user's current musical tastes. Furthermore, the more the user interacts with the system, the more personalised music is created for him or her.

IV. METHODOLOGY

4.1 Face Recognition Module

Face detection is the process of detecting a face from an image or video input. Face recognition technologies come in a variety of flavours. Face recognition is done using the Viola Jones algorithm. The primary steps in Viola Jones' algorithm are as follows:

A. HAAR feature

Some elements of the face are represented by HAAR features. Convolution kernels are used to identify the presence of a feature in an image, and Har features are similar to them. Each feature has a single value, This is calculated by subtracting the total number of pixels in the black rectangle from the total number of pixels in the white rectangle. the total of pixels in the white rectangle The black parts are changed with plus ones, whereas the white regions are replaced with minus ones in this feature..

B. Integral image

Every time the window advances in HAAR feature computation, all pixels in the black and white zones must be added up. It's a time-consuming technique with an integrated picture solution. It saves time by summing all pixels beneath a

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11319

rectangle with only four integral image corner values instead of summing all pixels under a rectangle with all four corner values. To obtain the value of any pixel, just sum the values of the pixels on the top and left.

C. Adaboost

The Viola Jones technique evaluates features in every given picture using a 24*24 window as the basic window. We'd have to calculate 160,000+ features in this window if we took into consideration all available feature characteristics like location size and type, which is nearly impossible. So, the main concept is to get rid of a lot of features that are redundant or ineffective, leaving just the ones that are highly useful. Adaboost's version eliminates 160 thousand features, reducing the amount of features we need to assess to a few thousand. The retrieved features from Adaboost are poor classifiers. Adaboost builds a linear combination of weak classifiers.

D. Cascading

Viola Jones' face detection technology is based on scanning the detector across the image multiple times, each time with a different size each time. Despite the fact that a picture should contain one or more faces, it is evident that a significant fraction of the examined sub windows are negatives. As a result, the algorithm should prioritise removing non-faces as early as possible. A single strong classifier constructed from a linear combination of all the best features is not suitable for assessing on each window due to the computation expense.

4.2 Module for Extracting Facial Features

For feature extraction, CNN is utilised. We must employ datasets including photos of joyful, angry, sad, and neutral emotions to train the system for the emotion recognition module. CNN has the unique ability to discover characteristics from dataset pictures for model creation via automated learning. To put it another way, CNN is capable of self-learning characteristics. CNN can internally represent a two-dimensional image. This is represented as a three-dimensional matrix, on which operations for training and testing are carried out. In certain other neural networks, such as fully connected networks, all nodes in one layer are connected to nodes in the next layer. Each link has a weight connected with it. As a result, the computational complexity will increase. Even yet, nodes in one layer are only linked to valid nodes in the following layer in CNN. The computational complexity will be lowered as a consequence. This has a lot of layers for training and testing input images. The last layer is fully integrated and includes a classification job, allowing pictures to be categorised according to their emotional content. The observed emotions should fit into one of four categories: furious, joyful, sad, or neutral. The entire dataset will be split in half before entering the CNN. The vast bulk of it will be utilised for training, with the remaining 20% being used for testing. This model will then be put to the test. The correctness of the system may be determined during the testing time by verifying whether the photos are accurately categorised Increasing the number of epochs or the number of pictures in the dataset can enhance accuracy. The input will be accepted by the neural network's convolution layer. A procedure known as filtering takes place at the convolution layer. Filtering is the math that goes into matching. First, the feature and the picture patch must be aligned. After that, multiply each picture pixel by the feature pixel that corresponds to it. Add them all up and divide by the feature's total number of pixels. After filtering, one image becomes a stack of images in convolution. Nonlinear processes can be implemented using the ReLU, tanh, and sigmoid functions. In ReLU, all negative values are transformed to zero, while positive ones are left alone. When the device is turned off the dripping A slight gradient is possible using ReLU. As a consequence, no information will go unnoticed. This also resolves the issue of the dying ReLU. ReLU outperforms the other two non-linear functions. The pooling layer will be placed after the convolution layer. This is done to make the convolution layer's picture stack smaller. As a result, the total number of parameters will be reduced [20]. Here you may choose a window size (usually 2 or 3). Then pick a stride (usually 2). A stride is the number of pixels that shift across the input matrix. When the stride is one, we move the filters one pixel at a time; when the stride is two, we change the filters two pixels at a time, and so on. The window is then moved across the photos that have been filtered. There are three different forms of pooling. The three types of pooling are maximum pooling, average pooling, and sum pooling. From a corrected feature map, max pooling finds the biggest element. Average pooling is the process of taking the average of the components in the window. Sum pooling refers to the sum of all elements in a feature map. Additional convolution and pooling layers can be added until the requisite precision is achieved. After the pooling layer, the matrix is flattened to a vector and sent to the fully connected layer. The goal is to convert a two-dimensional feature matrix into a feature vector that may be used to train a neural network or classifier. When the layers are fully linked, the vector elements merge to form models. Finally, data is classified using activation functions such as the SoftMax or Sigmoid functions.

4.3 Module for Emotion Detection and Song Classification

According on the identified emotion, the neural network classifier assigns one of four emotion labels: joyful, angry, sad, or neutral.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11319

V. IMPLEMENTATION

The suggested method is based on an automatic music recommendation system that plays music based on the person's mood or present feeling. When the app is opened, the person's photo is taken, and the current emotion is collected and identified. According to the information provided by the image, the music is played in connection to the feeling. The music on your phone has already been split into four groups: cheerful, sad, angry, and neutral. The freshly uploaded songs are also automatically sorted into moods. The system is made up of three components: facial expression recognition, song emotion recognition, and system integration. The modules for recognising face expressions and recognising aural emotions are mutually incompatible. As a result, the system integration module connects two modules in order to discover the best match for the identified emotion.

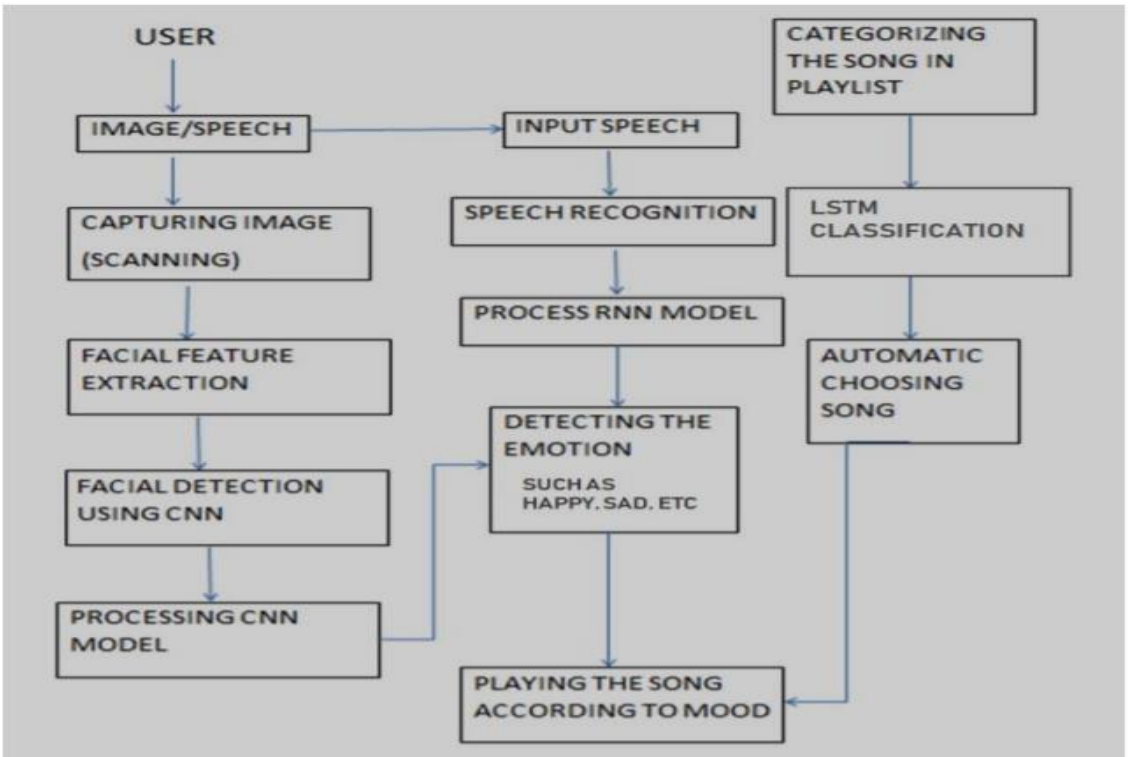
1.0 PROPOSED SYSTEM WORK

The suggested system can detect the user's facial expressions and extract facial landmarks based on those expressions, which can then be categorised to determine the user's sentiment. The user will be shown music that matches their emotions when the emotion has been recognised.

This section depicts the application's design and functioning phases. The user may access their personalised play-lists and play songs depending on their emotions using the Emotion-Based Music Player, which is loaded on a mobile device. The application's overview is seen in Figure 1.1.

1. Phase of login/sign-up: In order to save personal information, users must create a profile. If the user already has an account, they may log in to access their personalised play-lists and music. Once a user checks in, their profile is maintained on the app until they log out explicitly. The system takes the user's input (category and interest level) when they upload songs.

2. After the authentication step, the programme will ask the user for permission to access media and images, as well as access the camera to capture the user's image.



3. Affdex API: The programme sends the image to the Affdex SDK after it has been captured. The image is processed there, and the programme receives the image response.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11319

4. Emo-phase: During this phase, the programme gets visual data and determines the emotion based on a set of criteria. To get the emotion play-list, this emotion is given to the database.

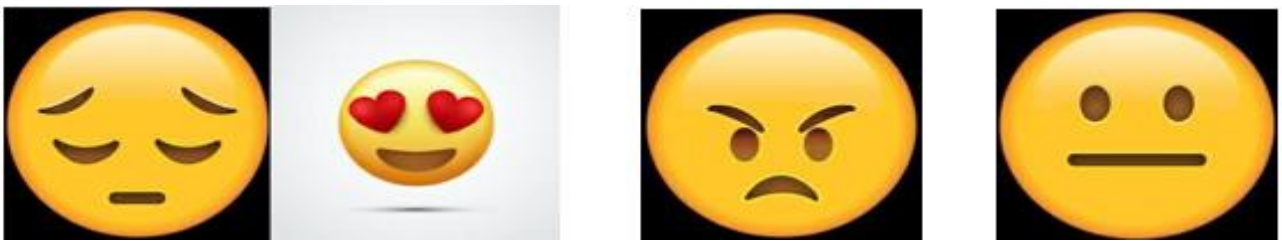
5. The music are arranged using the EMO algorithm at this phase, and the user can play any song from the list provided. The user may add, remove, and alter songs at any time in the programme, as well as change the song's category and interest level. The programme also has a suggestion tab, which alerts the user to songs that are infrequently played.

6. The creation of Emotion-Based Music Player employs a MERN stack method. Everything in the MERN stack is written in JavaScript and follows the MVC architecture. Both the server-side and client-side executions are done in JavaScript to increase the application's performance.

7. Emotion Based Music Player is a fantastic app for music listeners with a smartphone and internet connectivity. Anyone who creates a profile on the system has access to the application. The programme has been designed to meet the following user requirements:

a. Creating an account or signing up, signing in
b. Adding songs
c. Removing songs
d. Updating songs
e. Personalized playlist & Recommendations.
f. Capturing Emotions using a camera.

- **Detected emotions:**



Sad Happy Angry Neutral

From above fig., We have chosen these 4 Emotions based on which we may classify emotion directory for playing song.

VI. EXPERIMENT RESULTS AND ANALYSIS

This research presents a music recommendation system that extracts the user's image, which is taken via a camera connected to the computing platform. The collected frame of the image from the webcam feed is then transformed to a grayscale image to increase the performance is the name of the classifier that is used to identify the face in a photograph after it has been taken. After the conversion, the image is delivered to the classifier algorithm, which uses feature extraction to classify the image. extraction The face of a person can be extracted from the frame of a web camera broadcast using procedures. Individual aspects of the face are retrieved and transmitted to the trained network to detect the emotion conveyed by the user once the face has been extracted.

The User's Instructions Have Been Explained In this case, the users were given instructions on how to do the prediction of the expressed emotion, which yielded the following findings. When the inner mood is sad yet the facial display is pleasant, it can result in a failure situation. Table 1 shows the values, and the outcome is displayed in;

Fig. Instructions Explained to the User

User	Emotion	Facial Expression	Accuracy
1	Happy	Happy	100
2	Sad	Happy	0
3	Happy	Happy	100
4	Sad	Sad	100

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 3, March 2022

DOI: 10.17148/IJARCCE.2022.11319

CONCLUSION

Human emotion identification via facial expressions has a wide range of practical applications. It eliminates the time-consuming task of selecting the appropriate song for each occasion based on the individual's mood. This paper provides an overview of numerous methodologies and approaches that have been proposed and developed to identify human emotional states of behaviour when playing music, as well as an abstract picture of the proposed system that we will build...

ACKNOWLEDGMENT

We'd like to express our sincere gratitude to Dr. Uma Thakur for her invaluable guidance and continuous support. We also want to thank Dr. S.A. Dr. Leena Patil, the HOD of the Computer Science and Engineering Department, and Dr. Dhale, the principal of Priyadarshini College of Engineering Nilesh Shelke, the Project In-charge, for their kind cooperation, valuable guidance, and constant motivation, as well as for providing the necessary infrastructure and all the facilities for the project's development. We would also like to extend our gratitude to all of the faculty members in the department and institution. as well as all non-teaching personnel, for their assistance throughout the project. Finally, a big thank you to the researchers whose study pointed us in the right direction.

References

V. Patchava, P. Jain, R. Lomte, P. Shakthi, H. B. Kandala, “Sentiment Based Music Play System”.

1. Viral Prasad and Aurobind V. Iyer, "Emotion Based Mood Enhancing Music Recommendation," in 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information, and Communication Technology (RTEICT), India, pp. 1573-1577, May 19-20, 2017.
2. Harshala Chaudhari, Amrapali Waghmare, Reshma Ganjewar International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 10, May 2015, Dr. Abhijit Banubakode A media player that is controlled by Human Emotions.
3. "Facial Expression Recognition based on Local Region-Specific Features and SVM" by Deepak Ghimire, Sung Wan Jeong, Joonwhoan Lee, and Sang Hyun Park in Multimedia Tools and Applications, Vol.76, Issue 6, pp. 7803–7821, March 2017
4. Mary Duenwald (2005). The Physiology of Facial Expressions. Retrieved on October 9 2012 from.
5. Frijda, N.H. (1986): The emotions. New York: Cambridge University Press.
6. Maria M. Ruxanda1, Bee Yong Chua, Alexandros Nanopoulos, Christian S. Jensen. (2007): In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Emotion-Based Music Retrieval on a Well-Reduced Audio Feature Space was published (ICASSP): 181-184.
7. Eva Cerezo1, Isabelle Hupont, Critina Manresa, Javier Varona, Sandra Baldassarri, Francisco J. Perales, Eva Cerezo1, Isabelle Hupont, Critina Manresa, Javier Varona, Sandra Baldassarri, Francisco J. Perales, Eva Cerezo1, Isabelle Hu. Perales, and Francisco J. Seron. (2007): Real-Time Facial Expression Recognition for Natural Interaction: In J. Martí et al. (Eds.): IbPRIA 2007, Part II, LNCS 4478, (pp. 40–47). Springer-Verlag Berlin Heidelberg.
8. Ekman, P. & Friesen, W. V. (1969): The repertoire of nonverbal behavior: Categories, origins, usage, and coding: Semiotica, Vol. 1: 49–98.
9. Stock Photography [Image]. Retrieved on October 9 2012 from http://www.dreamstime.com.

