

MUSIC RECOMMENDATION SYSTEM USING SENTIMENT ANALYSIS

Pranjal Patil, Prachi Kaladeep, Sudhisha Zare , Dr. Ketaki Naik

Department of Information Technology, Bharati Vidyapeeth's
College of Engineering for Women, Pune, India

Abstract: *Recent studies have already shown that humans respond and react positively to music, and that music has a significant impact on human brain activity. People nowadays frequently prefer to listen to music as a source of entertainment based on their emotions and aspirations. This project concentrates on a system that recommends songs to users based on their emotional state. Computer vision components are used in this system to determine the user's emotion based on facial expressions. Once the emotion is identified, the system recommends a piece of music for that sentimentality, saving the user a significant amount of time over manually selecting and playing songs. It diminishes the time and effort required to manually search for music out of a list based on a person's current state of mind. The CNN algorithm and the Euclidean Distance classifier are used to detect a person's expressions by extracting facial features.*

Keywords: *CNN, Euclidean Distance classifier, API, face detection, Convolution, Feature Extraction, Image Acquisition, Pre-processing, Bias*

1. INTRODUCTION

Facial expressions are a natural way to communicate emotions, and these emotions can be used in leisure and Human Machine Interface (HMI) domains in modern environment. With technological advancements, various music players are implemented with features such as reversing the media, fast forwarding it, and streaming playback with multicast streams. Despite the fact that these features meet the user's basic needs, one must manually search for a song from a large set of songs based on the current situation and mood. We suggest a CNN-based method for recommending music based on expressions.

2. PROBLEM STATEMENT

In sentiment classification, music is categorized using a CNN-based approach by analyzing multi-modal sentimental data received by facial expressions and body language and semantic analysis of the user's textual/ verbal interactions, thereby intensifying the system's decision on recognized emotions in real-time. For example, innovative sentiment analysis can be applied to emotional states such as "infuriated," "sorrowful," or "joyful."

3. OBJECTIVE

The major goal of this effort is to create an advanced system that can quickly recognize facial expressions and then play a soundtrack based on that particular expression. Joy, Sorrow, Rage, Loathing, Anxiety, Surprised, and Neutrality are the seven widely recognized emotions. As a result, recommendations can be made based on the user's emotional and physiological status, which is mostly gathered through facial expressions, gestures, pulse rate, movement, speech/text interactions, and so on.

4. SCOPE

We created this software for desktop use. We are automating music classification to make it easier to identify relevant data such as trends, popular genres, and performers. Identifying music genres is the initial step in this strategy.

5. LITERATURE REVIEW

The review [1] goes into detail about how the system is unable to appropriately suggest existing products to a new user. According to the study [2], the emotion was mapped using the Valence-Arousal Co-ordinate System, and songs were suggested using the algorithm. According to the research paper [3,] music is recommended for use with the proposed system. The results are presented in two directions: numerical distributions and graphics. However, the algorithm is unable to effectively recommend a new item to an existing user. The study article [4] is scrutinized. The Keras library was used to calculate accuracy.

The network is initially trained for 60 epochs, with an accuracy of roughly 63 percent for those 60 epochs. The article [5] was tested on dynamic photos for face expression detection to attain real-time performance. The SVM classifier with the combined feature set produced the best overall performance, according to review [6]. (31.3 percent).

6. PROPOSED SYSTEM

The computer's automatic analysis and interpretation of music is a new option in the field of music information retrieval. We begin by logging onto the system. We catch the face or image in the camera during the capture phase. In this case, we use a camera to recognize

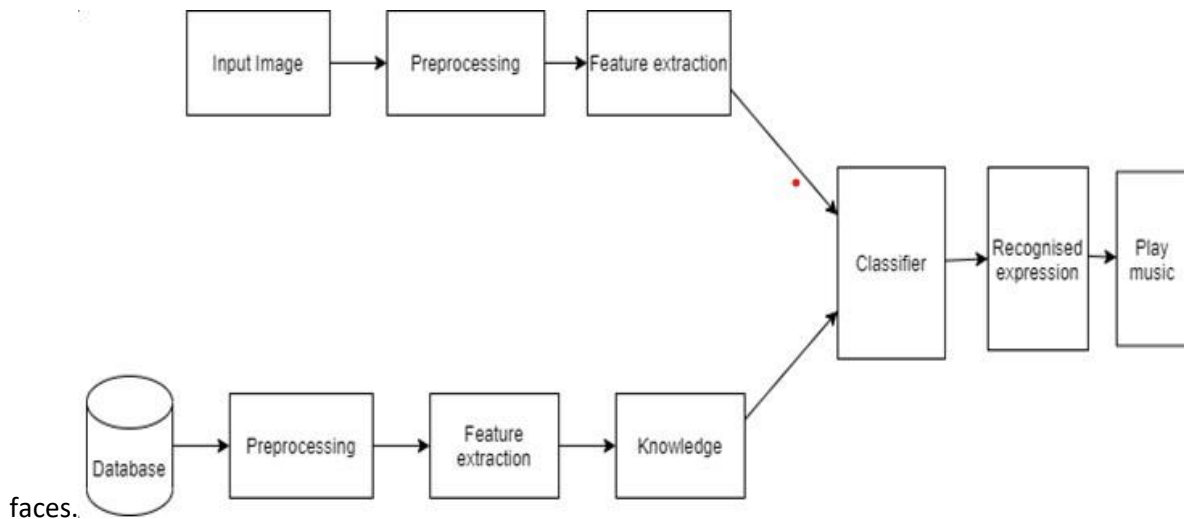


Fig.1. System Architecture: Music Recommendation using Sentiment Analysis

The basic purpose of the face detection technique is to identify the face that is exhibiting an emotion such as happiness, sadness, anger, and so on. In the following block, which is API. This API enables the two programs to communicate with one another. It is the interaction between the capture phase and the main phase. During the data storage step, we saved the photos that define various emotions. This data, along with incoming information from the capture phase, is compared in the main phase, and automatically played songs for these specific stages are projected in the displaying phase.

The system will use a simultaneous advanced search feature with which the user can search for a particular sound track on three various platforms. like: YouTube, Gana, Google. This gives user a flexibility to choose that music in either video or audio mode.

The CNN algorithm is used to classify the facial emotion as well as to get a count of the emotions occurring in specific category or class during the facial expression detection phase. This helps to know and improve the overall system accuracy.

7. PROPOSED OUTCOME

Using Python, create software that recognizes user emotion based on face expression.

- Incorporate the Python code into the web service and play music depending on the user's face expression.

- To provide high-quality entertainment experience for consumers.

8. METHODOLOGY

CONVOLUTIONAL NEURAL NETWORK (CNN)

A Convolutional Neural Network (CNN) is a type of neural network that is specialized in data processing with a grid-like architecture, as in a photograph. It employs a technique known as Convolution. Convolution is a mathematical operation performed on two functions that results in a third function that expresses how the shape of one is transformed by the other.

The first layer usually extracts basic features such as horizontal or diagonal edges. This output is passed on to the next layer which detects more complex features such as corners or combinational edges. As we progress further into the network, it can recognize more complex features such as objects, faces, and so on. The classification layer generates a series of confidence ratings (numbers ranging from 0 to 1) based on the activation map of the final convolution layer, indicating how probable the picture is to fit to a "class."

There are mainly five modules in the system:

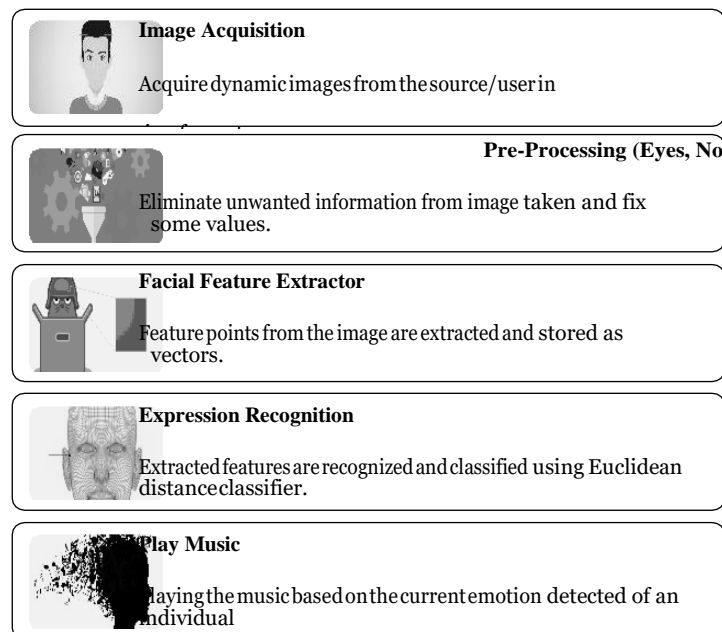


Fig.2 Modules in the System

Image Acquisition

The initial objective in every computer vision technique is to obtain the images from the sources. These photos can be captured with a camera or by using standard datasets that are freely available online. Photographs should be saved in .jpg format. The images under consideration here are user- dependent, or dynamic images. The amount of sample training photos taken into account here.

Pre-processing

In Pre-processing is mostly used to remove irrelevant data from the obtained image and to fix specific values so that the value remains constant throughout. All photos are transformed from RGB to Grayscale and scaled to 256*256 pixels during the pre-processing phase. Images in .jpg format are being considered; any other formats will not be considered for further processing. Throughout pre-processing, the region of interest is regarded to be the eyes, nose, and mouth. The cascade object detector, which employs the Jones-Viola algorithm, detects it.

Facial Feature Extraction

Now next stage after pre-processing is feature extraction. Even during training and testing phases, the extracted facial features are saved as usable information in the form of vectors. Mouth, forehead, eyes, skin complexion, cheek and chin dimples, eyebrows, nose, and wrinkles on the face are all examples of facial features. Because they convey the most appealing expressions, the eyes, nose, mouth, and forehead are considered for feature extraction in this study. The presence of wrinkles on the brow or an open mouth indicates that the person is either shocked or afraid. However, it is impossible to depict a user's complexion. CNN approach is often used to extract facial characteristics.

Expression Recognition

Euclidean distance classifier is used to recognize and classify a person's expressions. It finds the closest match for the test data in the training data set and so provides a better match for the currently detected expression. The Euclidean distance, which is given by "(3.1)," is essentially the distance between two points. It is derived from the mean of the training dataset's Eigen faces. The training images are labelled with expressions such as joyful, sad, fear, surprise, rage, disgust, and neutral, corresponding to varying distances from the mean image. When the Euclidean distance between the Eigen faces of the test picture and the mean image equals the distance between the Eigen faces of the training dataset, the expression is classed and named according to the labelled trained images. The closer the

match, the smaller the distance value obtained. If the distance value is large enough for an image, the system must be taught for that specific person.

Play Music

The final and most crucial component of this system is the playback of music based on an individual's present emotion. Once the person's facial expression has been classified, the person's emotional state is identified. A variety of songs from various genres referring variety of emotions are collected and placed on the list. Each emotion category has a number of songs. When the person's expression is categorized using the CNN algorithm, music from that class is played.

Convolutional neural networks are comprised of a number of layers of artificial neurons. Artificial neurons are mathematical functions which compute the weighted sum of many inputs and output one activation value, which is an approximate replica of their biological counterparts. When you feed an image into a ConvNet, every layer generates many activation functions, which are then passed on to the next layer. The basic workflow of a neuron is depicted in the graphic following graphic

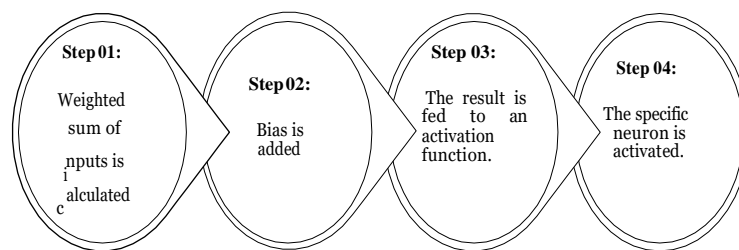


Fig. 3: Basic workflow of a neuron

Haar Cascade

Haar Cascade is an Object Detection Algorithm which is used to recognize facial images or real-time video. The algorithm employs the edge or line detection features proposed by Viola and Jones in their 2001 paper "Rapid Object Detection Using a Boosted Cascade of Simple Features." To train on, the algorithm is given a large number of positive images with faces and a large number of negative images with no faces. The model developed as a result of training is available on the Open CVGitHub repository. <https://github.com/opencv/opencv/tree/master/data/haarcascades>.

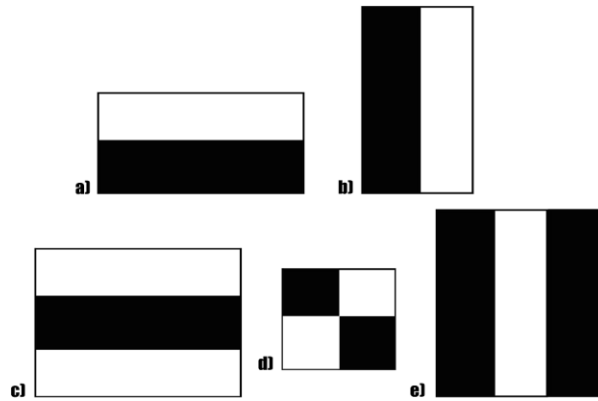


Fig.4: Viola and Jones' original research paper included a specimen of these Haar features

These are broadly classified into three categories based on the feature sought by each. The first set of two rectangle features is in charge of determining whether the edges are horizontal or vertical (as shown above). The second set of three rectangle features is in charge of determining whether a lighter region is surrounded by darker regions on either side or vice versa. The third set of four rectangle features is in charge of detecting changes in pixel intensities across diagonals.

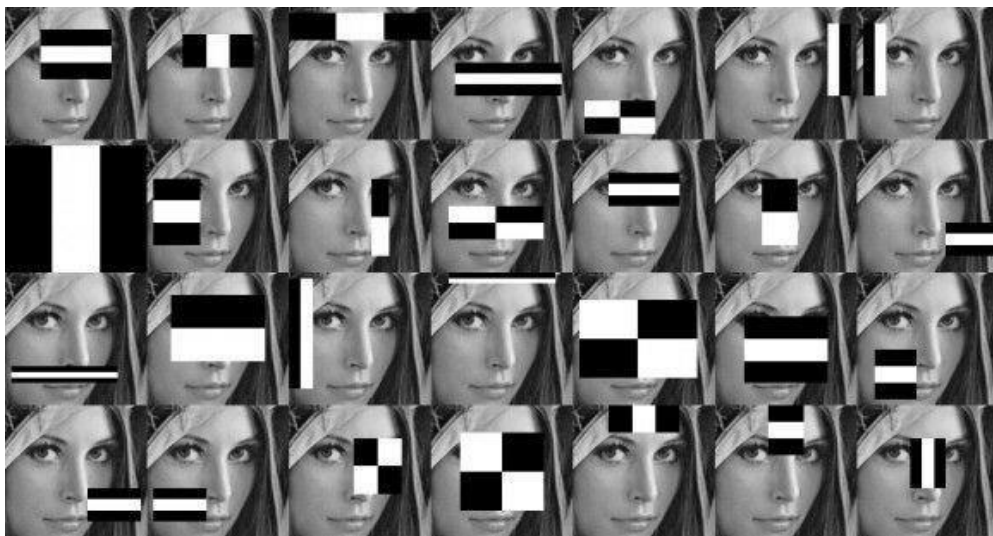


Fig.5: A general illustration of training Haar classifier

The goal is to calculate the sum of all image pixels in the darkened area of the haar feature and the total of all image pixels in the whitened area of the haar feature. Then figure out the difference of both.

If the image contains an edge separating black pixels on the right from white pixels on the left, the haar value will be closer to 1. That is, if the haar value is close to 1, we say that

there's an edge intercepted. There is no edge in the preceding example because the haar value is far from 1.

9. RESULTS and DISCUSSION

The proposed work presents a facial expression detection system that plays a piece of music based on the detected expression. The CNN these expressions. Real photos, i.e., user-dependent images, are captured in this work using the in-built camera. The final result demonstrates that the accuracy level reached is up to 95% technique is used to extract features, and the Euclidean distance classifier is used to classify.

10. FUTURE SCOPE

This approach can be applied in a variety of other fields, including security (classifying persons within an organization based on a pre-defined database) and domestic security (recommending movies, series, and so forth). Productivity - Improving productivity in businesses for their employees and stakeholders to support the appropriate person, at the right time for lead creation by addressing the employee and state of mind, and so on. This module may be applied in multiple applications with minor modifications as needed.

11. CONCLUSION

The haar classifier and CNN models' final result demonstrates that the accuracy level reached is up to 95%. The software successfully recognizes users' emotion based on face expression. Python code is incorporated into the web service giving the feature of play music depending on the user's face expression.

This project is an attempt to build and deploy intelligent entertainment platform for all music lovers, incorporating the new age technologies.

ACKNOWLEDGEMENTS

With an immense pleasure and satisfaction, we are presenting this Project as part of our curriculum of BE (Information Technology Engineering) – 2015 course of Savitribai Phule Pune University, Pune in the academic year 2021-22. We wish to express our sincere gratitude towards all those who have extended their support right from the stage this topic was conceived. We are profoundly grateful to Prof. Dr. K. A. Malgi Project Guide, for her expert guidance and continuous encouragement throughout to see that each work rights its target since its commencement to its completion. We would like to extend our thanks to our panel members Prof. A. D. Khairkar and Prof. M. A. Rane for their valuable suggestions and feedback throughout the project work. We are also grateful to Prof. Dr. D. A. Godse, (HOD),

Department of Information Technology Engineering, for her support and encouragement. Finally, we are also grateful to Honorable Mr. S. R. Patil, Principal, Bharati Vidyapeeth's College of Engineering for Women, Pune, for his support and guidance that has helped us to expand our horizons of thought and expression. We would also thanks to our family for their support and motivation throughout this process.

REFERENCES

- [1] Yu, Ziyang, Mengda Zhao, Yilin Wu, Peizhuo Liu, and Hexu Chen. "Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition." In *2020 39th Chinese Control Conference (CCC)*, pp. 7257-7263. IEEE, 2020.
- [2] Nathan, Karthik Subramanian, Manasi Arun, and Megala S. Kannan. "EMOSIC—an emotion-based music player for Android." In *2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 371-276. IEEE, 2017.
- [3] Kittimathaveenan, Kajornsak, Chanathip Pongskul, and Salisa Mahatanarat. "Music recommendation based on color." In *2020 6th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pp. 1-4. IEEE, 2020
- [4] Krupa, K. S., G. Ambara, Kartikey Rai, and Sahil Choudhury. "Emotion aware Smart Music Recommender System using Two Level CNN." In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 1322-1327. IEEE, 2020.
- [5] Kamble, Sushmita G., and A. H. Kulkarni. "Facial expression-based music player." In *2016 International Conference on Advances in Computing, Communications and Informatics (IKriCACCI)*, pp. 561-566. IEEE, 2016.
- [6] Przybysz, André L., Ricardo Corassa, Carolina L. Dos Santos, and Carlos N. Silla. "Latin music mood classification using cifras." In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1682-1686. IEEE, 2015.
- [7] Liang, Beici, and Minwei Gu. "Music genre classification using transfer learning." In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 392-393. IEEE, 2020.
- [8] Lee, Doohwan, Kohki Shibahara, Takayuki Kobayashi, Takayuki Mizuno, Hidehiko Takara, Akihide Sano, Hiroto Kawakami, Tadao Nakagawa, and Yutaka Miyamoto. "A sparsity managed adaptive MIMO equalization for few-mode fiber transmission with various differential mode delays." *Journal of Lightwave Technology* 34, no. 8 (2016): 1754-1761
- [9] Cherubin, Stefano, Clara Borrelli, Massimiliano Zanoni, Michele Buccoli, Augusto Sarti, and Stefano Tubaro. "Three-dimensional mapping of high-level music features for music browsing." In *2019 International Workshop on Multilayer Music Representation and Processing (MMRP)*, pp. 19-26. IEEE, 2019.
- [10] Hsu, Wen-Yu, Lan-Li Hsieh, Yu-Huei Su, Mei-Ju Su, Li Su, Mei-Chih Chen, and Hong-Ting Chan. "Establishment of a Music Care System for the Elderly in a Long-term Care Facility." In *2019 E-Health and Bioengineering Conference (EHB)*, pp. 1-4. IEEE, 2019.
- [11] Ayata, Deger, Yusuf Yaslan, and Mustafa E. Kamasak. "Emotion based music recommendation system using wearable physiological sensors." *IEEE transactions on consumer electronics* 64, no. 2 (2018): 196-203

- [12] Rosa, Renata L., Demsteneso Z. Rodriguez, and Graça Bressan. "Music recommendation system based on user's sentiments extracted from social networks." *IEEE Transactions on Consumer Electronics* 61, no. 3 (2015): 359-367.
- [13] Moscato, Vincenzo, Antonio Picariello, and Giancarlo Sperli. "An emotional recommender system for music." *IEEE Intelligent Systems* 36, no. 5 (2020): 57-68.
- [14] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I-I. IEEE, 2001 (alalek, 2013)
- [15] Jones, Paula, Paul Viola, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In *University of Rochester. Charles Rich*. 2001.
- [16] Cuimei, Li, Qi Zhiliang, Jia Nan, and Wu Jianhua. "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers." In *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, pp. 483-487. IEEE, 2017.