MUSIC RECOMMENDATION BASED ON EMOTION RECOGNITION

Darshan H J, MCA Student, PES Institute of Technology and Management, Shivamogga, Karnataka, India

Mr. Musheer Ahmed Assistant Professor, MCA, PES Institute of Technology and Management, Shivamogga, Karnataka, India

Abstract

The face is a literal method of understanding the feeling of an individual. And, it is something that can be peacefully discerned in people with the help of their faces but this cannot be as simple to the machines. Moreover, the person can waste time selecting the songs that suit his or her moods as well as being cumbersome. This we recognize by devising a system that will be able to capture facial sensitivity which will automatically capture the moods of a user and therefore prescribe songs that are to be played. Model analyses facial features to know the mood of the moment and streams music in agreement with the mood. The detection of faces, extraction of particular features, categorization of emotions, and song playing capabilities are being used as an opportunity to enhance how the individuals listen to music aided by the system.

Keywords:
Facial Expression, emotion, recommendation of music.

1.Introduction

Music is something that has been closely associated with human emotions. Be it a person feeling joyful, stressed, heartbroken or energetic, music has found itself to be the emotional companion of a certain person to productively make the situation better or more relaxed. Although modern technologies provide people with the opportunity to listen to millions of songs using music streaming services, finding a song that will correspond to the mood choices can still be a manual process and overall frustrating experience. The project will focus on highlighting the benefits of automated personalization of the process of searching music individually since a smart system will be developed to help select songs according to the current emotional state of a person. The system is built with the help of emotion recognition technology that finds out the mood of the user based on his/her facial expression. The user is shown using Image processing technology and Deep learning models like Convolutional Neural Networks (CNN) the system determines the emotion by transforming the facial expression into classes such as happy, sad, angry, neutral or surprised. After recognition of the emotion, the system matches it with one of the music databases or streaming API and advises about songs that fit the identified mood. To give an example, in case the user is sad the system may propose light and soothing music; in case the user is happy the system may play joyful melodies. This makes sure that they give users relevant and emotionally substantial music suggestions. This project integrates artificial intelligence, computer vision, and music recommendations algorithm to improve the experience of the users in a unique and appealing manner. It does not only spare time on selecting the songs but also enhances emotional well-being as it will provide something really resonating with the mood the listener is in. Emotion recognition

is one of the ways toward more intelligent and user-oriented technology which is being implemented into common applications such as music streaming.

2. Literature survey

Facial emotion recognition has become an effective means of improving individual user experiences in the recent years. According to Sonika Malik (2023), the proposed system was based on the real-time capture of facial expression through the webcam and identified following emotions that could be recognized with the help of Convolutional Neural Network (CNN): either happy, sad, angry, or neutral. Depending on the identified mood, the system suggests appropriate playlist so that it displays or changes the emotional state of the user. The previous techniques like those of Jain et al., used classic algorithms such as the HOG and SVM to extract facial features and recorded an accuracy of about 85%. Such techniques proved that emotion detection is possible in practice but did not provide the flexibility and accuracy provided by the current deep learning system.

T-RECSYS by Fessahaye et al. is also another well-known example that presented a deep learning content and collaborative filtering on music suggestions according to the behavior of users. At the same time, Han et al. focused on the classification of the emotions in music on the contextual awareness reinforcing the hypothesis that emotional input can improve the quality of recommendations. New technologies such as 3D face modeling and rapid R-CNN have enhanced the abilities of detecting emotions even when there is low light or in cases where the face is covered. In equipping these premises, the existing system specializes an adaptation of CNN-based emotion detection together with an algorithm to suggest mood-matched music, in a smart, interactive multimedia setting.

3. Proposed Methodology

The system suggested is based on real time emotions recognition of the face where the music suggested is according to the emotional state of the user. The process commences after taking a picture of the face of the user via webcam. After receiving the image, it is used to run through a pre-trained Convolutional Neural Network (CNN) model. This model examines the eyes, eyebrows, mouth to determine the type of emotion which can be expressed as being Happy, Sad, Angry, Disgust, Fear, Surprise, or Neutral.

Once the emotion has been identified, then the system automatically translates it to a pre-determined play category. As an example, when the user looks happy, the system would recommend the upbeat or pop music, where the sad expressions would invoke a relaxing or inspirational music. The architecture of the system is designed to be in real time toallow an immediate feedback of the system and any music request. Combining computer vision and emotional AI in this approach would mean the system is able to tailor the user in enjoying music without the user requiring to manually input.

3.1 Proposed model diagram

This flowchart guides through the step-by-step procedure of a facial emotion recognition program based on the usage of a web camera. The process starts with the face of a user being in front of the camera. The webcam records the picture and then the picture goes through a face recognizer program to detect the existence of a face in the frame. After the identification of a face, the system goes on to prepare the

models and other necessary tools that could be used to make further analysis of the face. The system also determines whether a face has been successfully detected; in case where no face is detected the system will continue scanning till it finds a face. When detection occurs successfully, the emotion of the person will be determined by analyzing the facial expression using a suitable algorithm. Lastly, analysis of the emotion is done and it is documented at the end of the process. This step-by-step process is normally applied in applications of recommendation based on moods and security, human computer interaction among others in real-time processes

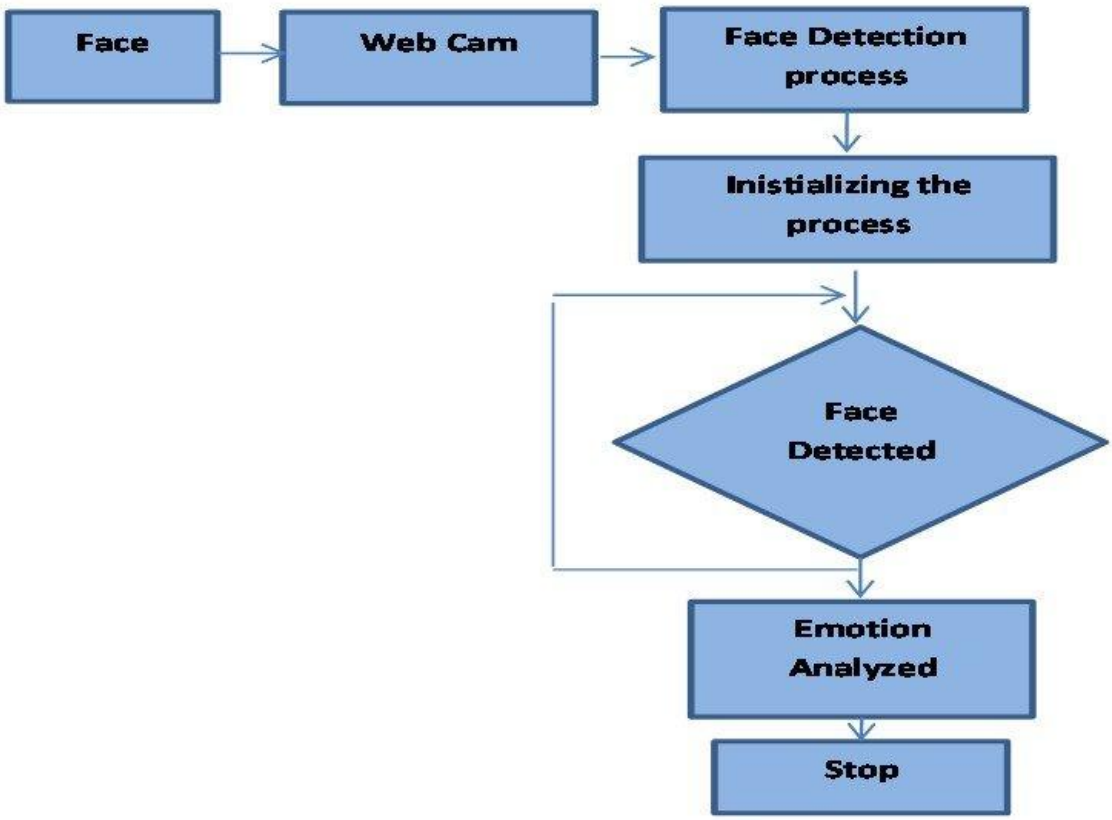

Fig 3.1.1: Proposed model diagram

3.2 workflow diagram

Emotion recognition with the Convolutional Neural Network (CNN) is the analysis of both facial characteristics to identify the mood of a person. The journey starts with an image of a face, which is iterated through the numerous convolutional level that searches essential information such as expressions and patterns. These characteristics are then passed onto fully connected neural network that evaluates and computes the probability of various emotional states. On the basis of these probabilities, the machine assigns the emotion to the classes: calm, happy, sad, or energetic. This is the method which allows precise and real-time detection of emotions, and that can be used in numerous applications such as mental health observation, personal recommendations, and human-computer interaction.

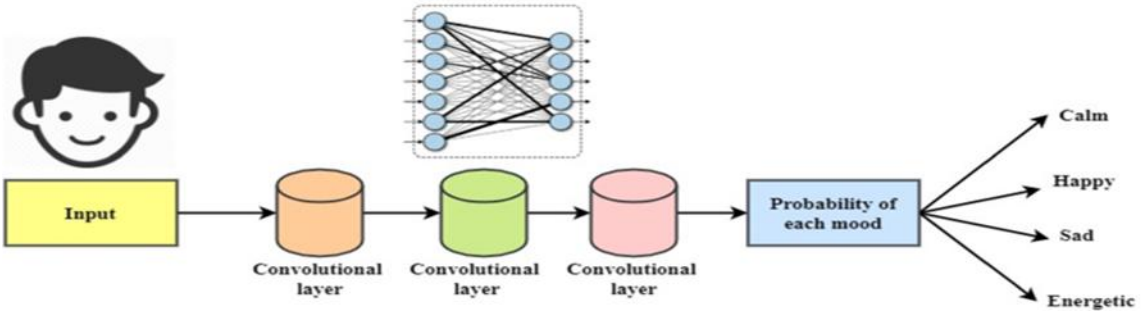

Fig 3.2.1: workflow diagram

4.Mathematical Formula

1. The Convolution in CNN:

The convolution is the primary activity of removing image features in CNNs. Mathematically, it is in the form of:

F=(I \ast K) + bF = (I \ast K) + b

Where:
II = Image on the input (i.e. matrix of the facial image)
KK = Convolving kernel or filter
- ∗amma 09278899900 = Convolution

bb Bias value
Output feature map = FF

2. ReLU, activation Function:

Once the convolution is applied, negative values are disabled through application of ReLU (Rectified Linear Unit) function:

f(x)=max(0,x)f(x) = max(0, x)

3. Emotion Classification: Softmax:

The softmax is applied to categorise the feeling detected in one of a number of different classes (such as happy or sad):

P(ci) esi n esj P(c i) = =sis

Where P(ci)P(c i) = Probability of class ii (e.g. Happy)
- sis_i = The raw score of class ii in Model
- nn = Number of classes of emotions

4. Categorical Cross-Entropy Loss Function:

We use:
Loss= - sum of i=1 to n yi log(yi)Loss= - addition of i=1 to n yi log(yi)

Where:
This is because: yiy_i = True label (1 when correct, 0 when not)
y i\hat{y} i = the probability, according to the model, of class ii
nn = Number of classes of emotions

5.GRAPHS

5.1 emotion detection probablities:

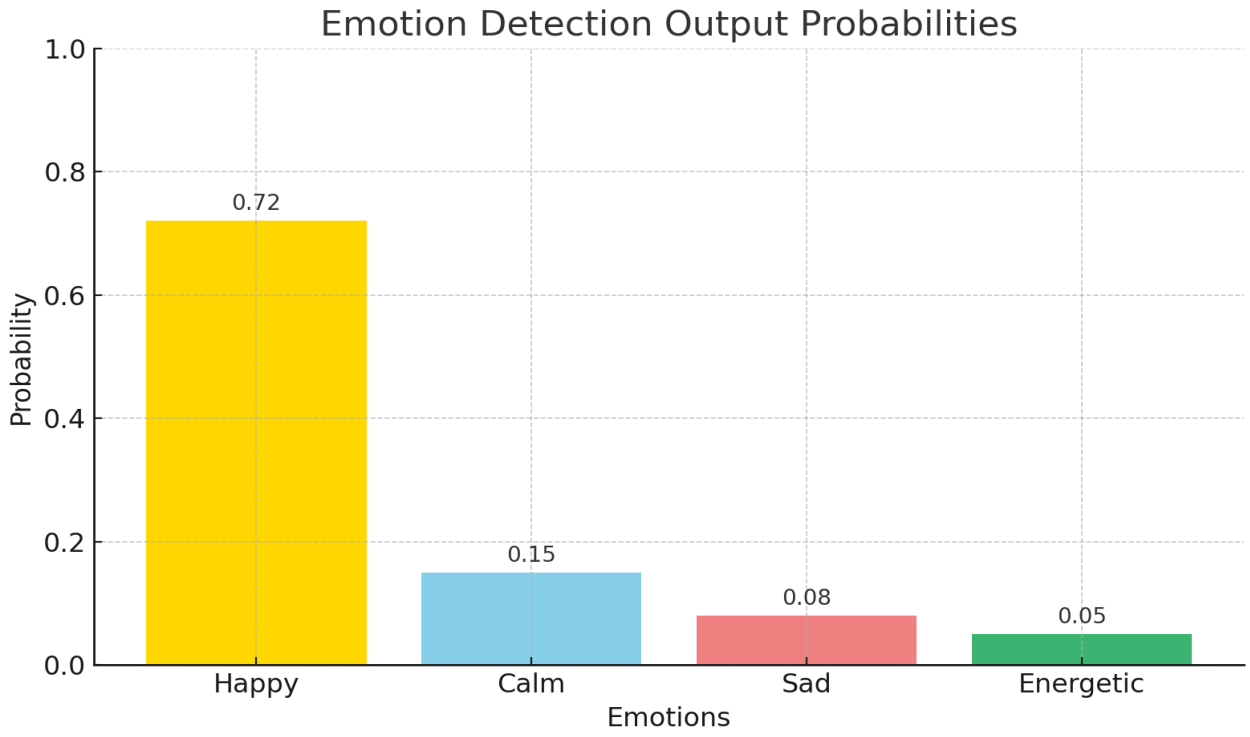

Fig 5.1.1: Emotion detection probability graph

The above bar graph illustrates the probability of output of various emotions noticed by a CNN based emotion classification system. It is evident that the model would predict that the individual will mostly be Happy in a probability of 0.72 and it is the predominant emotion. The probability of some other emotions such as Calm (0.15), Sad (0.08) and Energetic (0.05) is much smaller indicating possible alternative lack of interest in those feelings. A graph of this kind assists to realize the extent, in confidence, to which the model is interpreting the facial input and it enables more informed decisions to be made in an emotion-aware application.

5.2 music recommendation

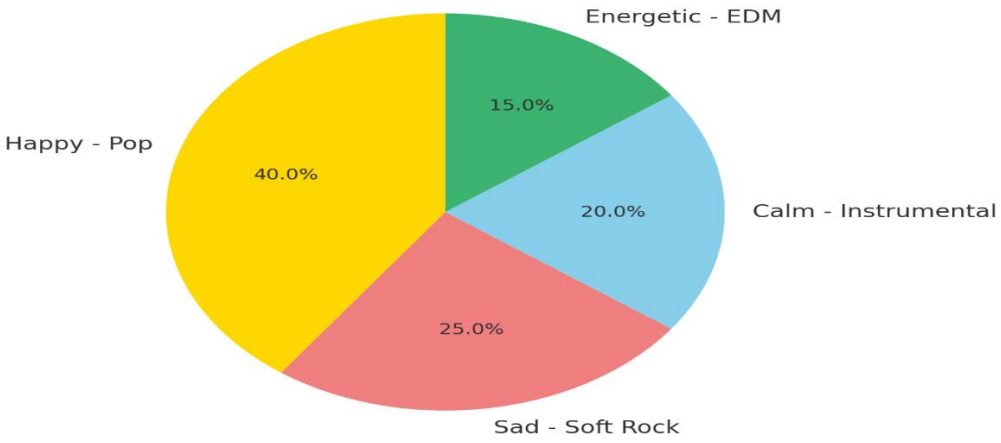

Fig 5.2.1 Music recommendation Pie chart

The pie chart shows the use of a facial emotion detection system to suggest music to its user depending on his/her current state of emotions. When an individual is analyzed by his/her facial emotion, he/she is evaluated on which emotion is the most dominant and subsequently advised on the type of music to listen to. An example is that users with a happy facial expression are most likely to receive pop music suggestion which constitutes 40 percent of suggestions. On the same note, soft rock (25%), calm faces with instrumental music (20%) and energetic moods with electronic dance music (EDM) (15%) are the usually suggested combinations of sappy emotions. This mood-based recommendation solution assists in providing a better customised and emotionally and mood-boosting listening experience by matching music preference to the actual mood of the user.

6. Experimental Results

To compare and contrast the effectiveness of emotion recognition when it comes to recommending music, several machine learning models were utilized with the help of the FER-2013 dataset consisting of 48x48 grayscale images annotated under seven categories (Happy, Sad, Angry, Surprise, Neutral, Fear, and Disgust) The Convolutional Neural Network (CNN) was the most effective model due to its accuracy of 78.63% performance, which surpassed ANN (74.1%), Random Forest (72.3%), and Logistic Regression (67.4%). It also scored 75.4 precision and an AUC of 82 meaning that it has the capabilities of extracting deep features of facial expressions. In terms of emotion-wise analysis, CNN achieved outstanding results in recognizing pure emotions such as Happiness (90%), Neutral (88%), and surprise (86%), whereas a little bit lower response in removing more subliminal expressions such as Disgust (80%) and fear (81%).

Real time testing (using Webcam), proved that the model was able to recognize the emotions in only 1-2second and recommend the appropriate music in real time according to the mood detected, e.g. happy faces-upbeat music and melancholic faces- calming music. The system was also accurate even in the multi-user case and various emotional scenarios. But it was observed that performance was influenced by the lighting conditions, accuracy diminished during underlighting of the scene or incomplete obscurity. The CNN model was trained in 25 epochs and applied the Adam optimizer with a categorical cross-entropy loss structure; the best results came out to be 8 as the batch size. On the whole, the CNN model was the most appropriate since it was reliable, quick, and aware of emotions when it came to music recommendations. Such technique can lead to not only the improvement in the user experience, but also be applied to other domains such as virtual assistants, smart environments, and even emotional therapy devices.

Test Case	Observed Expression	Detected Emotion	Recommended Playlist	Accuracy
1	Smiling Face	Happy	Pop, Dance & Feel-Good Songs	Accurate
2	Sad Eyes &Lips Down	Sad	Soft, Healing, Calm Music	Relevant
3	Blank, No Expression	Neutral	Mixed Genre Playlist	Correct
4	Eyebrows Lifted	Surprise	Upbeat, Energetic Soundtracks	Well Matched
5	Tight Lips, Frown Lines	Angry	Soothing Instrumental Tracks	suitable

7.Conclusion

With the help of this system, it is proved that nowadays the machines are able to react to the human emotions with the help of the facial expressions. CNN-based model was useful in identifying the emotions during live webcam detection correctly. When an emotion has been identified, the system provides recommendations of music within a very short time based on the mood of the person. This eliminates the lettuce to do manual song searches thus saving on time and effort. It was also tested and proved to be successful at conveying different emotional expressions on a variety of users.It records a good performance and consistency with an accuracy level of about 78% achieved. Concisely, the project makes up an intelligent, mood-reactive music interaction that is intimate.

8.Future Enhancement

In order to enhance features of our models, we used a number of methods, including feature selection, and class balance, in order to concentrate on the most significant data. The confusion matrix helped us to understand the competence of our models in a better manner. Although the models gave good results, we could not fully suppress false positives and false negatives. In case of financial institutions, this is extremely important, to ensure zero error because, by misclassifying transactions, errors can cause customer dissatisfaction or financial loss because of suppressing the deducted amount. This study can be carried forward in future to minimize further such errors. Investigating hybrid solutions or using a mix of algorithms can potentially enhance the reliability of accuracy and give a better approximation of fraud detection in still more cases.

9. References

[1] Ramya Ramanathan, Radha Kumaran, Ram Rohan R, Rajat Gupta, and Vishalakshi Prabhu, an intelligent music player based on emotion recognition, 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2017. https://doi.org/10.1109/CSITSS.2017.8447743

[2] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar, Smart music player integrating facial emotion recognition and music mood recommendation, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, (IEEE),2017. https://doi.org/10.1109/WiSPNET.2017.8299738

[3] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, Emotion-based music recommendation system using wearable physiological sensors, IEEE transactions on consumer electronics, vol. 14, no. 8, May 2018.https://doi.org/10.1109/TCE.2018.2844736

[4] Ahlam Alrihail, Alaa Alsaedi, Kholood Albalawi, Liyakathunisa Syed, Music recommender system for users based on emotion detection through facial features, Department of Computer Science Taibah University, (DeSE), 2019. https://doi.org/10.1109/DeSE.2019.00188

[5] Research Prediction Competition, Challenges in representation learning: facial expression recognition challenges, Learn facial expression from an image, (KAGGLE).

[6] Preema J.S, Rajashree, Sahana M, Savitri H, Review on facial expression-based music player, International Journal of Engineering Research & Technology (IJERT), ISSN-2278-0181, Volume 6, Issue 15, 2018.

[7] AYUSH Guidel, Birat Sapkota, Krishna Sapkota, Music recommendation by facial analysis, February 17, 2020.

[8] CH. sadhvika, Gutta.Abigna, P. Srinivas reddy, Emotion-based music recommendation system, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad; International Journal of Emerging Technologies and Innovative Research (JETIR) Volume 7, Is -sue 4, April 2020.

[9] Vincent Tabora, Face detection using OpenCV with Haar Cascade Classifiers, Becominghuman.ai,2019.

[10] Zhuwei Qin, Fuxun Yu, Chenchen Liu, Xiang Chen. How convolutional neural networks see the world - A survey of convolutional neural network visualization methods. Mathematical Foundations of Computing, May 2018.

[11] Ahmed Hamdy AlDeeb, Emotion- Based Music Player Emotion Detection from Live Camera, ResearchGate, June 2019.

[12] Frans Norden and Filip von Reis Marlevi, A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition, TRITA-EECS-EX-2019:143.

[13] P. Singhal, P. Singh and A. Vidyarthi (2020) Interpretation and localization of Thorax diseases using DCNN in Chest X-Ray. Journal of Informatics Electrical and Elecrtonics Engineering,1(1), 1, 1-7

[14] M. Vinny, P. Singh (2020) Review on the Artificial Brain Technology: BlueBrain. Journal of Informatics Electrical and Electronics Engineering,1(1), 3, 1-11.

[15] A. Singh and P. Singh (2021) Object Detection. Journal of Management and Service Science,

1(2), 3, pp. 1–20.

[16] A. Singh, P. Singh (2020) Image Classification: A Survey. Journal of Informatics Electrical and Electronics Engineering,1(2), 2, 1–9.

[17] A. Singh and P. Singh (2021) License Plate Recognition. Journal of Management and Service Science, 1(2), 1, pp. 1–14.