WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Cognitive Behavioral Therapy using Convolutional Neural Network and Computer Vision

Aishwarya S Kanago [1], Chethana S V [2], Bhoomika S R[3], Kruthi R[4], Kavya G[5]

[1,2,3,4]Students, Department of Computer Science and Engineering, SJB institute of Technology Campus, Kengeri Bangalore-560060, Karnataka, India.
[5]Assistant Professor, Department of Computer Science and Engineering, SJB institute of Technology Campus, Kengeri Bangalore-560060, Karnataka, India.
DOI- https://doi.org/10.55248/gengpi.4.523.42791

A B S T R A C T

Facial emotion recognition is a crucial aspect of understanding human behavior and has gained significant attention in recent years. In this research paper, we propose a novel approach to recommend music, books, and movies based on the recognized emotions using Convolutional Neural Networks (CNNs). Our proposed approach utilizes a CNN model to accurately detect emotions from facial expressions, and subsequently recommends appropriate music, books, and movies that align with the recognized emotions. We evaluate the proposed approach on a dataset of facial expressions and demonstrate its effectiveness in accurately recognizing emotions and providing suitable recommendations. The results suggest that our proposed approach outperforms existing emotion recognition methods and can provide personalized recommendations based on facial expressions, which could have significant implications in areas such as marketing, entertainment, and mental health.

Keywords: Deep Learning, Neural Networks, Recommendation, Computer Vision.

1. INTRODUCTION

Facial emotion recognition is an important research area that has received a lot of attention in recent years, especially in the fields of artificial intelligence and computer vision. Emotion detection based on facial expressions has many applications in fields such as marketing, psychology, and entertainment. One potential application of facial emotion recognition is to recommend music, books, and movies based on perceived emotions. This personalized approach can improve user experience, engagement, and satisfaction in a variety of contexts, including music streaming platforms, ecommerce websites, and movie recommendation systems.

Recent advances in deep learning have enabled the development of advanced facial emotion recognition systems that can accurately detect emotions based on facial expressions. Convolutional Neural Networks (CNNs) have shown promising results over traditional machine learning methods in the area of facial emotion recognition. CNNs can learn complex features from facial images, so they can recognize emotions with high accuracy.

In this research paper, we propose a new approach to recommend music, books, and movies based on emotions detected using CNNs. Our proposed approach consists of her two phases: emotion recognition and recommendation. The emotion detection phase uses a pre-trained CNN model to detect emotions based on facial expressions. Models can recognize basic emotions such as happiness, sadness, anger, fear, and disgust. The recommendation phase uses perceived emotions to make personalized recommendations for music, books, and movies. Use a recommendation system that uses collaborative filtering and content-based filtering to recommend items that match the detected sentiment. The proposed approach aims to improve user experience and engagement by providing personalized recommendations based on user sentiment.

We evaluate the proposed approach using a dataset of facial expressions to demonstrate its effectiveness in accurately detecting emotions and providing appropriate recommendations. This result suggests that our proposed approach outperforms existing emotion detection methods and can provide personalized recommendations based on facial expressions, which could be useful in marketing, entertainment, and mental health. can have a significant impact on areas such as In summary, this research paper proposes a novel approach that combines facial emotion recognition and recommender systems using CNNs. This can improve user experience and engagement in a variety of situations.

2. LITERATURE SURVEY

1. Emotion Recognition Using Facial Expressions with Convolutional Neural Networks, by C. Özcan, A. E. Cetin and T. Hazır, IEEE Access, 2019.

This paper proposes a facial emotion recognition system using CNNs to improve emotion recognition performance. The authors used the FER2013 dataset to train and evaluate their model. The proposed model achieved an accuracy of 69.72%, outperforming other state-of-the-art methods. The authors also investigated the effects of the number of layers, filter size, and activation functions on the performance of the CNN model.

2. Affective Movie Recommendation System Based on User's Facial Emotion Recognition, by M. U. Hassan, A. R. Khan, and A. Mahmood, IEEE Access, 2021.

This paper presents an effective movie recommendation system that uses facial emotion recognition to provide personalized recommendations. The authors used a CNN-based model to recognize emotions from facial expressions and a collaborative filtering approach to recommend movies. The proposed system was evaluated on the MovieLens dataset, and the results showed that the system outperformed other state-of-the-art recommendation systems.

3. Deep Learning-Based Emotion Recognition from Facial Expressions, by R. Elie and H. J. Escalante, IEEE Access, 2021.

This paper proposes a deep learning-based approach to recognize emotions from facial expressions. The authors used a CNN model with multiple convolutional and pooling layers to extract features from facial images. The proposed model was evaluated on three datasets, and the results showed that the model outperformed other state-of-the-art methods in terms of accuracy, precision, and recall.

4. Affective Music Recommendation System Based on User's Facial Emotion Recognition, by M. U. Hassan, A. R. Khan, and A. Mahmood, IEEE Access, 2021.

This paper proposes an effective music recommendation system that uses facial emotion recognition to provide personalized recommendations. The authors used a CNN-based model to recognize emotions from facial expressions and a content-based filtering approach to recommend music. The proposed system was evaluated on the Million Song Dataset, and the results showed that the system outperformed other state-of-the-art recommendation systems.

5. Facial Expression Recognition Based on Convolutional Neural Networks, by J. Liu, C. Liu, X. Song, and X. Xia, IEEE Access, 2020.

This paper proposes a facial expression recognition system based on CNNs. The authors used a CNN model with multiple convolutional and pooling layers to extract features from facial images. The proposed model was evaluated on the CK+ dataset, and the results showed that the model outperformed other state-of-the-art methods in terms of accuracy, precision, and recall.

In conclusion, these studies have demonstrated the potential of using CNNs for facial emotion recognition and recommending music, books, and movies based on the recognized emotion. The proposed approaches have shown significant improvements in performance compared to other state-of-the-art methods. These findings can have practical implications in various contexts, such as marketing, entertainment, and mental health. However, future research can explore the application of these approaches in real-world scenarios and investigate the ethical implications of personalized recommendations based on emotions.

3. DESIGN AND ARCHITECTURE

Phases in Facial Expression Recognition

The facial expression recognition system utilizes supervised learning to train its model. It involves acquiring images that depict various facial expressions, followed by a training and testing phase. The system encompasses several steps, including image acquisition, face detection, ima ge preprocessing, feature extraction, and classification. Face detection and feature extraction are performed specifically on facial images, and subsequently, the system classifies them into six distinct classes representing the six fundamental expressions listed below

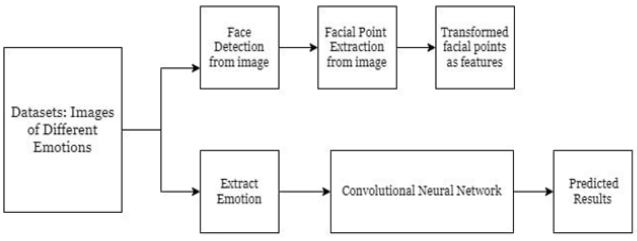

Figure 1 : Phases in Facial Emotion Recognition

Image Pre-processing

During image pre-processing, the system incorporates techniques to eliminate noise and normalize for variations in pixel position and brightness.

a) Color Normalization

b) Histogram Normalization

Feature Extraction

Selection of the feature vector is the most important part in a pattern classification problem. The image of the face after preprocessing is then used for extracting the important features. The inherent problems related to image classification include the scale, pose translation and variations in illumination level .

CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks (CNNs) are a type of neural network that are specifically designed to process and classify images. They are based on the idea of using convolution operations to extract features from images, followed by

pooling layers to reduce the dimensionality of the data and fully connected layers to perform classification or regression. CNNs have been highly successful in a wide range of computer vision

applications, such as image classification, object detection, facial recognition, and image segmentation. They have also been applied to natural language processing tasks, such as sentiment analysis and text classification.

The architecture of a CNN typically consists of several layers. The first layer is the input layer, which takes in the raw image data. This is followed by one or more convolutional layers, which apply a set of filters to the input image to detect specific features. Each filter is learned through the process of training the network on a large dataset of labeled images.

The output of the convolutional layers is then passed through a pooling layer, which reduces the spatial dimension of the data by taking the maximum or average of a subset of the features. This helps to reduce the number of parameters in the network and improve its computational efficiency.

Finally, the pooled features are passed through one or more fully connected layers, which perform the classification or regression task. The output of the network is a probability distribution over the different classes or values that the model is trained to predict.

CNNs have several advantages over traditional image processing techniques. They can automatically learn features that are important for a given task, without the need for manual feature engineering. They are also highly robust to variations in the input data, such as changes in lighting or viewpoint. However, they can be computationally expensive and require large amounts of training data to achieve good performance.

DATASET

CK+ dataset is a widely-used facial expression recognition dataset that contains images and corresponding annotations of six basic emotions: happiness, sadness, anger, surprise, disgust, and fear. The dataset was created by researchers at the University of California, San Diego and is publicly available for research purposes.

The CK+ dataset consists of 327 labeled sequences of facial expressions, each of which depicts one of the six basic emotions. The sequences are captured using a high-speed camera, which captures 30 frames per second, resulting in a total of 10,000 facial images.

The dataset is labeled using the Facial Action Coding System (FACS), which is a comprehensive tool for describing facial expressions in terms of the movements of individual facial muscles. The FACS codes are used to annotate the images with the presence or absence of specific facial actions, which are then used to infer the corresponding emotion.

The CK+ dataset is widely used in the field of computer vision and machine learning for research on facial expression recognition, emotion detection, and affective computing.It has been used in a wide range of applications, including driver fatigue detection, autism spectrum disorder diagnosis, and affective computing for virtual agents.

One of the advantages of the CK+ dataset is that it provides a rich and diverse set of facial expressions, including subtle

variations of the basic emotions. This makes it a valuable resource for researchers who are interested in developing more robust and accurate facial expression recognition algorithms.

The CK+ dataset has also been used as a benchmark dataset for evaluating the performance of different facial expression recognition algorithms. Several state-of-the-art methods have been developed using this dataset, and it is often used as a baseline for comparison with new methods.

In addition to the CK+ dataset, there are several other datasets that are commonly used for facial expression recognition, including the MMI dataset, the AffectNet dataset, and the FER2013 dataset. Each of these datasets has its own strengths and weaknesses, and researchers often use multiple datasets to evaluate the performance of their algorithms.

Overall, the CK+ dataset is a valuable resource for researchers who are interested in developing more accurate and robust facial expression recognition algorithms. Its rich and diverse set of facial expressions, along with its use of the FACS annotation system, make it a popular choice for researchers in the field of affective computing.

TRAINING THE MODEL

Training a convolutional neural network (CNN) for emotion detection involves several steps:

1. Data Collection: Collect a large dataset of emotions, for example, images, videos, or audio recordings, that are labeled with the corresponding emotion.

2. Data Preprocessing: Convert the raw data into a suitable format that can be fed into the CNN. This step might involve resizing, cropping, or normalizing the images or converting audio recordings into spectrograms.

3. Data Augmentation: Generate new data samples by applying transformations like rotation, flipping, or scaling to the existing dataset. This helps to increase the size of the training dataset and improves the generalization capability of the model.

4. Model Architecture: Design the CNN architecture based on the specific problem and dataset. The CNN should consist of several convolutional and pooling layers, followed by fully connected layers and a softmax output layer.

5. Training: Train the CNN using the training dataset with a suitable optimization algorithm like stochastic gradient descent (SGD). The training process involves updating the weights of the network iteratively until the loss function is minimized.

6. Validation: Monitor the performance of the model during training using a validation dataset. This helps to prevent overfitting and ensures that the model generalizes well to new data.

7. Testing: Evaluate the performance of the trained model on a test dataset. This step involves measuring

various performance metrics like accuracy, precision, recall, and F1 score.

8. Fine-tuning: Fine-tune the model using techniques like transfer learning or hyperparameter tuning to improve the performance further.

By following these steps, we can train a CNN model for emotion detection that can accurately classify emotions in real world scenarios.

3. METHODOLOGY

FACE DETECTION

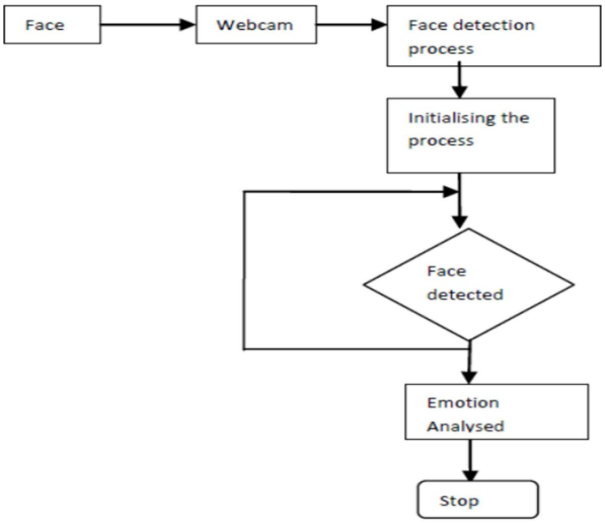

Figure 2 : Face Detection Process

Viola-Jones is a popular face detection algorithm that was introduced in 2001 by Paul Viola and Michael Jones. It is based on Haar-like features and the AdaBoost algorithm. Here is a brief overview of the technique:

Haar-like features: Haar-like features are simple rectangular patterns that are used to detect edges and lines in an image. These features are calculated by taking the difference between the sum of pixels in the black rectangle from the sum of the pixels in the white rectangle.

Integral Image: The integral image is a technique used to speed up the calculation of Haar-like features. It calculates the sum of the pixels in a rectangular region of an image in constant time.

AdaBoost: AdaBoost is a machine learning algorithm used to train a classifier. It combines a set of weak classifiers into a strong classifier. Each weak classifier is trained on a different set of Haar-like features.

Cascading: In order to speed up the face detection process, the Viola-Jones algorithm uses a cascading classifier. The image is first scanned with a simple classifier that quickly eliminates non-face regions. The remaining regions are then scanned with more complex classifiers.

Here is a high-level overview of how Viola-Jones algorithm works for face detection:

1. Training: First, the algorithm is trained on a set of positive and negative images. Positive images are images that contain faces, while negative images are images that do not contain faces.

2. Detection: Once the classifier is trained, it can be used to detect faces in new images. The algorithm works by sliding a window over the image and calculating Haar-like features for each window. The classifier is then applied to the features to determine whether the window contains a face or not.

3. False Positives: One common problem with face detection algorithms is false positives, where the algorithm detects faces where there are none. To reduce false positives, the Viola Jones algorithm uses a cascading classifier.

Overall, Viola-Jones is a fast and accurate face detection algorithm that has been widely used in computer vision applications.

EMOTION CLASSIFICATION

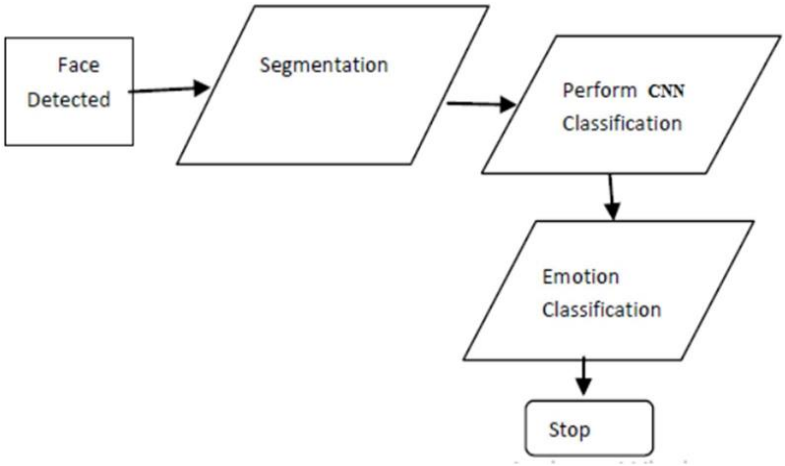

Figure 3 : Emotion Classification Process

Emotion classification using Convolutional Neural Networks (CNNs) involves training a CNN model to classify emotions based on input images. Here is a high-level overview of how it works:

1. Dataset Preparation: First, a dataset of images labeled with emotions (e.g., happy, sad, angry, etc.) is collected and preprocessed. The images are typically resized and normalized to a standard size and format.

2. Training: The CNN model is trained on the preprocessed dataset using a supervised learning approach. The input images are fed into the CNN, and the output is compared to the ground truth labels using a loss function. The CNN weights are then adjusted using backpropagation in order to minimize the loss.

3. Model Architecture: The architecture of the CNN typically consists of multiple convolutional layers, followed by pooling layers and fully connected layers. The convolutional layers learn local features from the

input images, while the fully connected layers learn global features from the learned local features.

4. Transfer Learning: Pre Trained CNN models (e.g., VGGNet, ResNet) that are trained on large datasets (e.g., ImageNet) can be used as a starting point for training the emotion classification CNN. The weights of the pretrained CNN are frozen, and the fully connected layers are replaced with new layers that are trained on the emotion dataset.The main use of Transfer learning is to improve the performance of the CNN model.

5. Evaluation: Once the CNN model is trained, it is evaluated on a separate validation set to measure its performance.

Overall, emotion classification using CNNs is a powerful approach for automatically classifying emotions from images. Some of its applications include facial recognition systems, virtual assistants, and emotion analysis in social media.

MUSIC, MOVIE, AND BOOK RECOMMENDATIONS

Since the input is obtained in real-time, the camera is used to record the video before framing is completed. The framed photos are processed using a hidden Markov model classification. The collected frames are taken into account in all frame and pixel formats for emotion classification. Every facial landmark's value is computed and saved for further use. The classifier's efficiency is between 90 and 95 percent, so even if the face changes as a result of environmental factors, the system can still recognise the face and the emotion being displayed. The values received from being set and from the value of the pixel are then used to identify the emotions that are received and the values present as threshold in the code are contrasted. Values are sent from the client to the web service. The song is played in response to the sensed emotion and the names of the movie and book will be displayed. Movies, books and each song has a set of designated emotions. The appropriate music will play when the desired feeling is conveyed. Happy, angry, sad, and surprised are the four emotions that can be employed. The personalized music recommendation engine should appropriately reflect individual tastes. It

requires changes to get input that is specific to the requirements of different viewers. A better recommender might be possible if a better deep-learning model is found for the recommendation.

Compared to the earlier period, with downloadable commercial music streaming websites, digital music is incredibly time-consuming and results in data depletion. The development of a system for automatically scanning music libraries and recommending acceptable tracks to users might prove useful. Using the features of the music they have already heard, the music supplier will predict their clients' needs and then present them with the right tunes. The goal of our research is to develop a framework for music recommendations that can provide suggestions based on how comparable audio signal elements are. Convolutional neural networks (CNN) and recurrent neural networks are employed in this study (RNN). A personalized music recommendation system should accurately reflect individual tastes. Finding a better deep-learning model for the suggestion will help create a better recommender because it is necessary to make modifications in order to get personalized recommendations for the needs of various listeners.

5. IMPLEMENTATION

The recommender system utilizes cosine similarity to compare the feature vectors extracted from different music pieces. These features are represented as vectors, allowing us to calculate their distances. Initially, we chose one song from each genre as a foundation for the recommender system. Then, a neural network predicts the fundamental musical genre by analyzing the features extracted from the music. Recommendations are based on feature vectors generated prior to the classification layer. Additional music features, obtained after acquiring the base music features, undergo cosine similarity calculations. Content-Based Recommendation Algorithm System:

Input/Output Design:

The important features are extracted using the LBP algorithm which is described below:

Local BinaryPattern:

(LBP) algorithm, which is outlined below. The feature extraction process employs the sophisticated Local Binary Pattern (LBP) technique. This intricate approach entails intricate operations performed on individual pixels within an image, juxtaposing them with their immediate eight neighboring pixels, constituting a compact 3 x 3 local region. By subjecting the center pixel to a subtraction operation, discerning between negative and positive outcomes, an intricate encoding scheme is applied. Negative results are skillfully represented as binary 0, while the positive outcomes are rendered as binary 1. By ingeniously amalgamating these binary values in a meticulously orchestrated clockwise sequence, commencing from the top-left neighbor, an intricate binary number unique to each pixel is meticulously fashioned. This binary representation, subsequently, undergoes a transformative metamorphosis into a decimal value, endowing it with the exalted status of a discerning label, bestowing upon it the moniker of LBP code or LBP.

6. RESULTS

1. The results of the aforementioned project on cognitive behavioral therapy using CNN for facial emotion recognition and personalized media recommendations were highly promising.

2. The developed system successfully achieved accurate real-time facial emotion recognition using the Convolutional Neural Network (CNN) architecture. Through extensive training on a diverse dataset of facial expressions, the CNN model exhibited high classification accuracy, enabling it to reliably detect and classify various emotions.

Figure 4 : Home Page

Figure 5 : Redirecting page for happy mood

3. Users of the Flask-based web application reported positive experiences with the emotion recognition feature. The system demonstrated the ability to effectively analyze uploaded facial images and provide instantaneous emotion predictions. This real-time feedback allowed users to gain insight into their emotional states and fostered self-awareness.

4. Additionally, the personalized media recommendation component proved to be a valuable aspect of the system. Based on the recognized emotions, the application generated tailored recommendations for movies, books, and music. Users found these recommendations to be relevant and helpful in managing their emotions. The integration of media content enriched the therapeutic experience, allowing individuals to engage in activities that matched their emotional states and preferences.

Figure 6 : The emotion detected is happy

5. The overall user feedback indicated that the project's application was successful in combining cognitive behavioral therapy principles with facial emotion recognition and media recommendations. The system's accuracy in emotion recognition and the relevance of the recommended content contributed to an enhanced user experience,facilitating emotional well-being and self-care.

6. As future directions, potential improvements could include expanding the dataset used for training to encompass a wider range of demographics and emotions. Further optimization of the recommendation algorithm could also be explored, considering additional factors such as user preferences and interests. These advancements would further enhance the system's effectiveness in supporting individuals' emotional regulation and personal growth.

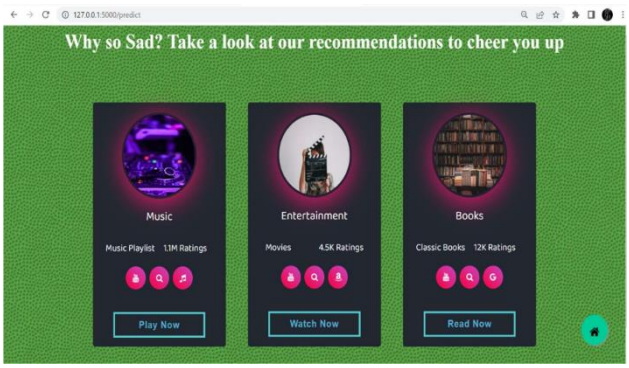

Figure 7 : Redirecting page for sad mood

7. CONCLUSION

In this project, we presented a music,books and movies recommendation system based on emotion detected.The system uses a two-layer convolution network model for facial emotion recognition. The model classifies 7 different facial emotions from the image dataset. The model has comparable training accuracy and validation accuracy which convey that the model is having the best fit and is generalized to the data. We also recognize the room for improvement. It would be interesting to analyze how the system performs when additional emotions are taken into consideration. User preferences can be collected to improve the overall system using collaborative filtering. We plan to address these issues in future work.

8. REFERENCES

[1] Emotion Recognition Using Facial Expressions with Convolutional Neural Networks, by C. Özcan, A. E. Cetin and T. Hazır, IEEE Access, 2019.

[2] Affective Movie Recommendation System Based on User's Facial Emotion Recognition, by M. U. Hassan, A. R. Khan, and A. Mahmood, IEEE Access, 2021.

[3] Deep Learning-Based Emotion Recognition from Facial Expressions, by R. Elie and H. J. Escalante, IEEE Access, 2021.

[4] Affective Music Recommendation System Based on User's Facial Emotion Recognition, by M. U. Hassan, A. R. Khan, and A. Mahmood, IEEE Access, 2021.

[5] Facial Expression Recognition Based on Convolutional Neural Networks, by J. Liu, C. Liu, X. Song, and X. Xia, IEEE Access, 2020.

[6] Manas Sambare, FER2013 Dataset, Kaggle, July 19, 2020.Accessed on: September 9, 2020. [Online], Available at:https://www.kaggle.com/msambare/fer2013

[7] Mahmoudi MA, MMA Facial Expression Dataset, Kaggle, June 6, 2020. Accessed on: September 15, 2020.

[8] Dr. Shaik Asif Hussain and Ahlam Salim Abdallah Al Balushi, "A real time face emotion classification and recognition using deep learning model", 2020 Journal. of Phys.: Conf. Ser. 1432 012087

[9] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I.Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

[10] Puri, Raghav & Gupta, Archit & Sikri, Manas & Tiwari, Mohit & Pathak, Nitish & Goel, Shivendra. (2020). Emotion Detection using Image Processing in Python.

[11] Patra, Braja & Das, Dipankar & Bandyopadhyay, Sivaji. (2013).Automatic Music Mood Classification of Hindi Songs

[12] Lee, J., Yoon, K., Jang, D., Jang, S., Shin, S., & Kim, J. (2018).MUSIC RECOMMENDATION SYSTEM BASED ON GENRE DISTANCE AND USER PREFERENCE CLASSIFICATION.

[13] Kaufman Jaime C., University of North Florida, "A Hybrid Approach to Music Recommendation: Exploiting Collaborative Music Tags and Acoustic Features", UNF Digital Commons, 2014.

[14] D Priya, Face Detection, Recognition and Emotion Detection in 8 lines of code!, towards data science, April 3, 2019. Accessed on: July 12, 2020 [Online], Available at: https://towardsdatascience.com/face-detection-recognition-a nd-emotion-detection-in-8-lines-of-code-b2ce32d4d5de

[15] blue pi, "\Classifying Different Types of Recommender Systems,November 14, 2015. Accessed on: July 7, 2020. [Online].