WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Mental Health Assistance based on Emotion Recognition using Machine Learning

Madhushri Wadki[1], Tejas Biradar[2], Shikhar Shetty[3], Smita Patted[4], Ms. Rajeshwari Kisan[5]

[1]Dept. of Computer Science Engineering S. G. Balekundri Institute of Technology. Belagavi, Karnataka, India. madhushri.wadki2001@gmail.com
[2]Dept. of Computer Science Engineering S. G. Balekundri Institute of Technology. Belagavi, Karnataka, India. tbiradar49@gmail.com
[3]Dept. of Computer Science Engineering S. G. Balekundri Institute of Technology. Belagavi, Karnataka, India. shettyshikhar13@gmail.com
[4]Dept. of Computer Science Engineering S. G. Balekundri Institute of Technology. Belagavi, Karnataka, India. pattedsmita19@gmail.com
[5]Assistant Professor, Dept. of Computer Science Engineering S. G. Balekundri Institute of Technology. Belagavi, Karnataka, India. rajeshwari.kisan@sgbit.edu.in

ABSTRACT—

Facial Expressions are an essential feature of non- verbal communication, as we move towards digitization, the human computer interactions play a vital role. The emotional changes result in changes in the expressions. This paper elaborates development of Deep Convolutional Neural Network Model using tf.keras for building and training Deep Learning Model. The aim is to classify facial images into one of the face detection classifiers using open CV and one of its classifiers for drawing the boundary box around the face to detect the correct expression. For training the CNN models we have used 48x48 grey–scale images from Kaggle's ICMP 2013-Facial Expression Recognition (FER) dataset. Depending on the facial expression detected we will develop a music and video recommendation system which can be used to recommend music and videos to the users depending on the facial expression or emotion detected. The flask app is developed which is used to load the face image and to perform facial emotion detection and suggest the music and video playlist depending on the emotion or expression detected.

Keywords— Facial emotion recognition, Music and video recommendation, Deep learning, Convolutional Neural Network (CNN), Flask web application framework.

I. Introduction

Face detection and emotion recognition are prominent subjects in the security domain, offering solutions to a range of challenges. These challenges primarily arise from the uncontrolled settings in which facial images are captured, including factors like diverse poses, varying lighting conditions, and a wide array of expressions for face recognition. Similarly, emotion recognition systems must contend with different sound frequencies. To facilitate the comparison of facial features in any face and emotion detection system, the database assumes a crucial role.

The exploration of human emotional responses to visual stimuli like photographs and films, commonly known as visual sentiment analysis, has emerged as a captivating yet challenging field. Its objective is to comprehend the higher-level information conveyed by visual data. The progress of powerful algorithms in computer vision has played a significant role in the success of current models. Many existing models aim to address this challenge by either proposing robust features or employing more advanced models. Notably, the suggested inputs mainly focus on visual elements extracted from the entire image or video, whereas the importance of local areas, which we believe significantly contributes to the emotional reactions of humans towards the entire image, has been relatively neglected. Utilizing image recognition techniques, the system identifies individuals in photos, analyzes their emotions, and selects emotion-related music based on their emotional states.

Humans often convey their emotions through facial expressions, making them a valuable indicator of their emotional state. Music and videos have long been recognized for their ability to influence and modify an individual's mood. By capturing and recognizing the emotions expressed by a person, and subsequently displaying suitable songs and videos that align with their mood, it is possible to create a calming and pleasing effect. This project focuses on capturing a person's emotions through their facial expressions. A music player is developed to leverage the webcam interface available on computing systems. The software captures the user's image and utilizes image segmentation and processing techniques to extract facial features, enabling the detection of the expressed emotion. The primary goal of the project is to uplift the user's mood by playing songs that cater to their emotional needs, as determined by the captured facial image. Facial expression recognition has been a long-standing and effective means of analyzing human expression since ancient times.

Facial expressions are widely regarded as the most effective method for analyzing and understanding the emotions, feelings, and thoughts conveyed by another person. Additionally, altering one's mood can play a crucial role in overcoming challenges such as depression and sadness. Through expression

analysis, it becomes possible to identify potential health risks and take proactive measures to improve the user's emotional state. By leveraging the power of facial expression analysis, steps can be taken to enhance the user's mood and promote a more positive outlook on life.

II. LITERATURE REVIEW

In study [1], The system begins by detecting the presence of a human face in the captured image. Once detected, the system proceeds to localize and crop the face, enclosing it within a bounding box. The cropped image is then converted to grayscale. The next step involves locating facial key-points on the grayscale image. These key-points provide crucial information about facial features. Using the positions of these key-points, SVM (Support Vector Machine) classifiers are employed to detect the corresponding emotion expressed by the person. The detected emotion is subsequently utilized for song filtering.

The song filtration process takes into account not only the detected emotion but also six additional factors. These factors are considered based on their relationships and dependencies with each other. By analyzing these factors, the system determines the most suitable songs to match the detected emotion and the specific combination of the other relevant factors. The goal is to create a song selection process that considers multiple factors to enhance the overall music experience based on the user's emotional state.

In study [2], Images are represented as 2-dimensional arrays of pixels, where each pixel possesses unique features. While algorithms like SVM, decision trees, and Random Forest in Scikit-learn are proficient at classification tasks, they do not inherently extract relevant features from images. This is where Convolutional Neural Networks (CNNs) come into play. CNNs are specifically designed for image processing and comprise multiple layers, including the input layer, convolutional layer, pooling layer, and dense layer.

CNNs utilize the concept of convolution, which involves sliding a filter over the input image to extract meaningful features. The convolutional layer performs this operation and learns to recognize various patterns and structures in the image data. The pooling layer reduces the dimensionality of the feature maps generated by the convolutional layer, making the network more efficient. Finally, the dense layer, also known as the fully connected layer, takes the high-level features extracted by the previous layers and performs the final classification.

MobileNet is an example of a CNN model specifically designed for mobile and embedded vision applications. It combines deep neural networks with streamlined architectures that exhibit low latency. These characteristics make MobileNet suitable for applications that require real-time processing and low-latency responses, even on resource-constrained devices.

In study [3], Real-time face recognition was performed using either a webcam or a mobile camera. The system captured 30 frames per second, which were then used as input for the CNN LeNet architecture. The LeNet model extracted features from the input images, enabling the recognition of facial emotions. This demonstrates the capability of efficiently identifying emotions even in challenging conditions such as poor lighting and variations in facial appearance caused by environmental changes.

To enhance the user's mood, a curated list of songs was compiled and associated with various emotions. Based on the detected facial expressions, the system recommends songs that align with the expressed emotion. This personalized song recommendation approach adds to the overall user experience, creating a more immersive and enjoyable interaction with the system.

In study [4], The primary objective of this system is to detect human emotions in order to develop an emotion-based music player. The system starts by capturing the person's face through the webcam, and this continuous stream of images is converted into frames. Through pre-processing techniques, the facial expressions are further transformed into a sequence of Action Units, which represent specific facial movements or expressions.

Next, feature extraction is performed on the Action Units to extract relevant information for emotion recognition. The emotions are then classified into categories such as Happy, Angry, Sad, and Surprise based on the extracted features. The system integrates web services to facilitate the transfer of emotions detected from the facial expressions. Based on the detected emotions, the music player selects and plays music that aligns with the expressed emotional state of the person.

By leveraging facial expression analysis and emotion recognition techniques, the system provides a personalized and immersive music experience tailored to the user's emotional state.

In study [5], This paper introduces a novel approach that combines deep learning, specifically a convolutional neural network (CNN), with an expression recognition model. The method leverages image processing techniques and integrates them with a music recommendation algorithm to provide music recommendations based on the detected mood of the person.

To ensure robust recognition performance, standardized and unified standards are applied to the image pre-processing stage. These standards eliminate variations in image size, posture, brightness, and darkness that may affect recognition accuracy. The model architecture includes multiple convolutional units, and the parameters of each unit are optimized using the backpropagation algorithm.

To create the music dataset, playlists from major music websites are crawled, and manual annotations are added. This dataset is then utilized to recommend music that aligns with the individual's mood, as inferred from their facial expressions. The results demonstrate an expanded scope of application for the image processing outcomes.

Overall, this research contributes to the development of a deep learning-based system that effectively combines image processing, expression recognition, and music recommendation, offering a promising approach to enhancing the music experience based on individual emotions.

III. METHODOLOGY

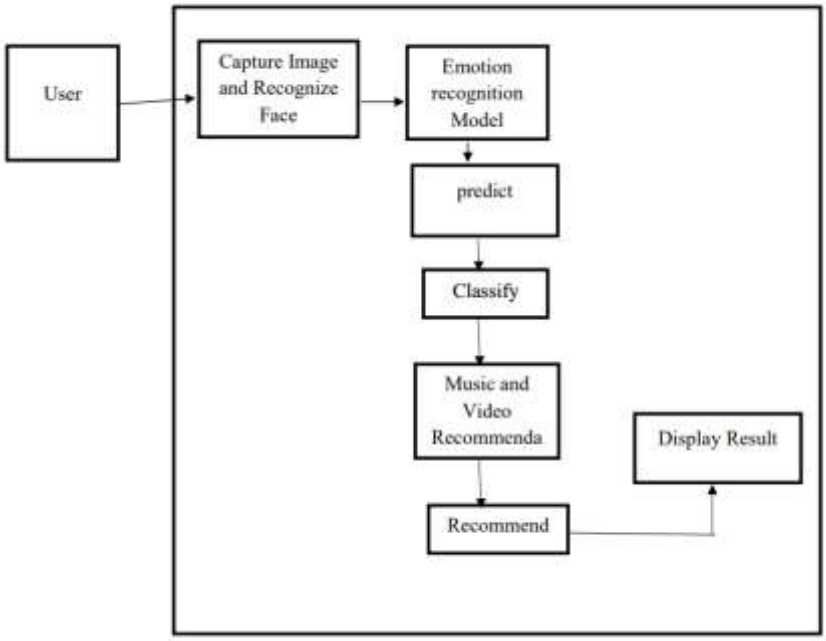

Fig 1. Architecture Diagram

The methodology for implementation of the Music and Video recommendation system based on facial mood detection is divided into following modules which are implemented through the phase of the project. The entire project implementation is divided into different modules:

1. The development of the facial mood recognition system involves training a deep learning model to detect and classify different expressions and moods exhibited by a person's face. Here is a detailed description of the process:

1. Dataset collection: A diverse dataset is collected that includes images of faces displaying various emotions such as happiness, sadness, anger, surprise, etc.

- The dataset should include a wide range of individuals, diverse facial features, and different lighting conditions to ensure robustness of the trained model.

2. Data augmentation: To increase the diversity and size of the dataset, data augmentation techniques are applied.

- Augmentation techniques may include rotation, scaling, flipping, cropping, and adding noise to the images.

3. Model architecture selection: A suitable deep learning architecture for facial mood recognition is chosen, such as Convolutional Neural Networks (CNNs).

- Popular CNN architectures like VGG, ResNet, or custom architectures can be considered based on the complexity of the problem and available computational resources.

4. Training the model: The collected and augmented dataset is split into training and validation sets.

- During training, the model learns to extract relevant features from facial images and predict the corresponding emotion labels.

- The optimization process involves backpropagation and gradient descent to update the model's weights and minimize the loss function.

5. Batch normalization: Batch normalization is applied to normalize the input features and speed up the training process.

- It helps in reducing internal covariate shift and makes the model more stable during training.

6. Performance analysis: The trained model is evaluated using the validation set to assess its performance and generalization ability.

By following these steps, the facial mood recognition system is trained to accurately detect and classify emotions based on facial expressions.

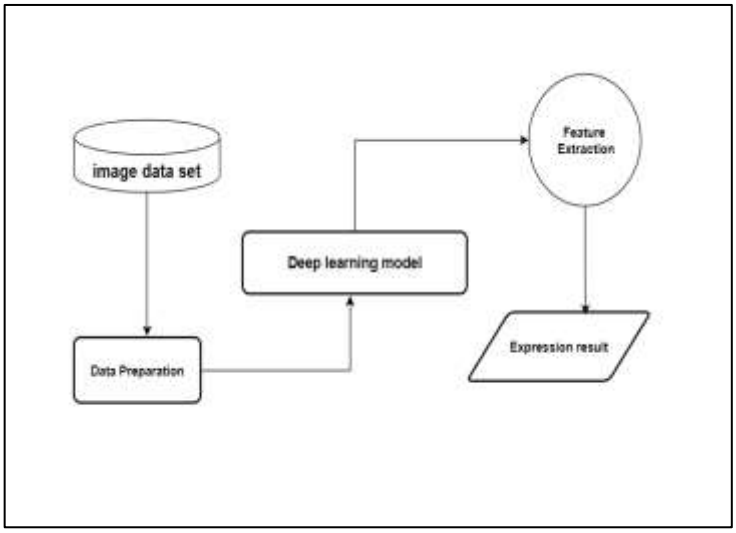

Fig 2. Data Flow Diagram(0)

2. The Music and Video recommendation system is designed to provide relevant music and video suggestions based on the output of the emotion recognition model. Here is a detailed description of how the system works:

1. Face capture: The system captures the face of the person in front of the camera, typically through a browser interface.

2. Facial feature extraction: The captured face undergoes facial feature extraction, where specific facial landmarks or features are identified and extracted. These features may include facial expressions, gestures, or other visual cues.

3. Emotion recognition: The extracted facial features are then fed into the trained emotion recognition model. The model analyzes the features and classifies the facial emotion into different categories such as happiness, sadness, anger, etc. This step relies on the deep learning model trained in the facial mood recognition system.

4. Music and video recommendation: Once the facial emotion is classified, the system sends a request to the music and video recommendation system, which is developed using a Python backend. The recommendation system takes into account the detected mood/emotion.

5. Customized playlist: The music and video recommendation system selects appropriate music and videos from a customized playlist. This playlist is tailored to match the different emotions or moods identified by the facial emotion recognition model. The playlist may include a variety of songs and videos that evoke the corresponding emotions.

6. Recommendation output: The system provides the recommended music and videos to the user based on their detected facial emotion. This output can be displayed on the front-end of the application, allowing the user to enjoy content that aligns with their mood.

By integrating the emotion recognition model with the music and video recommendation system, this approach provides a personalized and interactive experience for users, ensuring that the recommended content corresponds to their current emotional state.

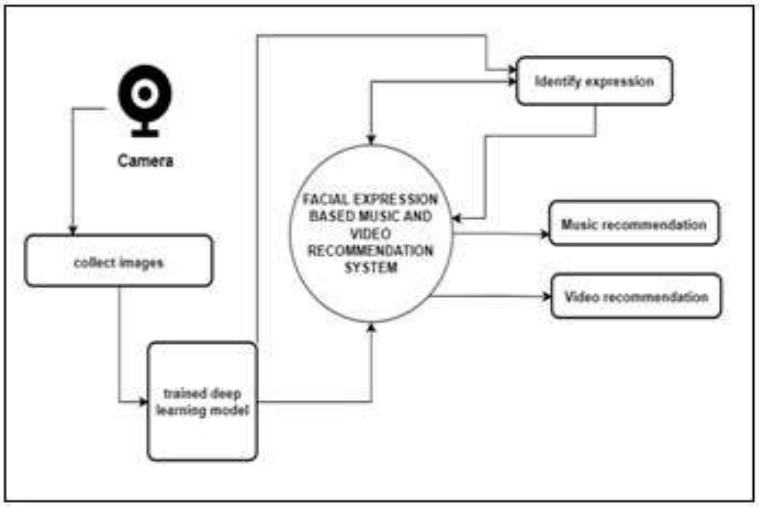

Fig 3. Data Flow Diagram(1)

3. The Flask-based Recommendation application is a web application module developed using the Flask web application framework. Its primary purpose is to generate music and video recommendations based on the detected emotions. Here is a detailed description of how this module works:

1. Web application development: The web application is developed using Flask, which is a popular Python web framework. Flask provides a lightweight and flexible environment for building web applications.

2. User interface: The front-end of the application is designed to interact with the user. It typically includes a user interface where the user can view the recommendations and control the application's functionalities.

3. Webcam integration: The application integrates with the user's webcam to capture images in real-time. The webcam is accessed through the browser, allowing the application to capture the user's face for emotion detection.

4. Emotion detection: Once the image is captured from the webcam, the application performs facial feature extraction and sends the extracted features to the trained emotion recognition model. The model detects the emotions based on the facial features and classifies them accordingly.

5. Recommendation generation: After the emotions are classified, the Flask backend sends a request to the music and video recommendation system. This request includes the detected emotions as input.

6. Customized playlist retrieval: The music and video recommendation system, which is integrated with the Flask backend, uses the detected emotions to select suitable music and videos from a customized playlist. The playlist is pre-defined and curated to match different emotions.

7. Recommendation display: The Flask backend retrieves the recommended music and videos from the playlist and sends them back to the front-end of the application. The recommendations are displayed to the user on the user interface.

8. User interaction: The user can interact with the application by selecting the recommended music or videos, controlling playback, and exploring other features offered by the application.

The Flask-based Recommendation application provides a seamless user experience by capturing real-time images from the webcam, detecting emotions, and generating personalized music and video recommendations based on the detected emotions. The application's front-end allows users to interact with the recommendations and enjoy content that aligns with their current emotional state.

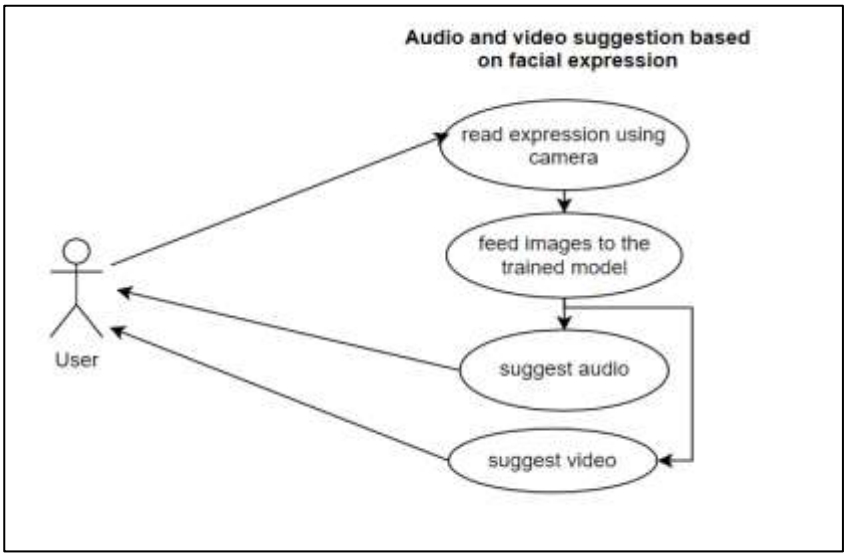

Fig 4 . Use-Case Diagram

IV. RESULT AND DISCUSSION

The proposed project aims to develop an intelligent music and video recommendation system that takes into account the user's mood or emotions. This innovative system offers a standalone application capable of playing music and videos while also providing the ability to transform one's mood from sadness to happiness through personalized recommendations. By accurately recognizing the user's emotions, the system can suggest the most suitable music to uplift their spirits and refresh their mood. Furthermore, the system is designed to continuously recommend music and videos that align with the user's desired emotional state, ensuring a sustained sense of happiness.

This project encompasses the implementation of an automatic mood and emotion recognition system integrated with a recommendation system. The end result is a user-friendly web application that can be easily accessed and launched from a single interface. The system's ability to seamlessly combine

emotion recognition, music and video playback, and personalized recommendations creates an immersive and enjoyable experience for users seeking to enhance their mood and overall well-being.

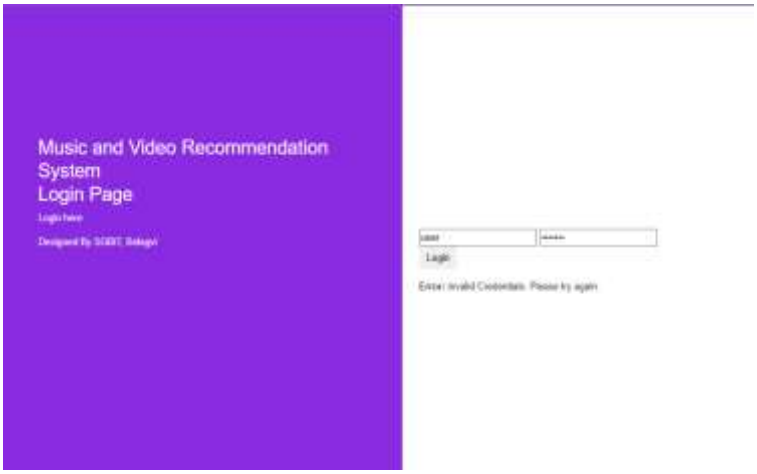

Fig 5. Login Page

To access the system or a specific application, users initiate the login process, usually through a user interface.

Users may be prompted to provide their username or any other required identification information.

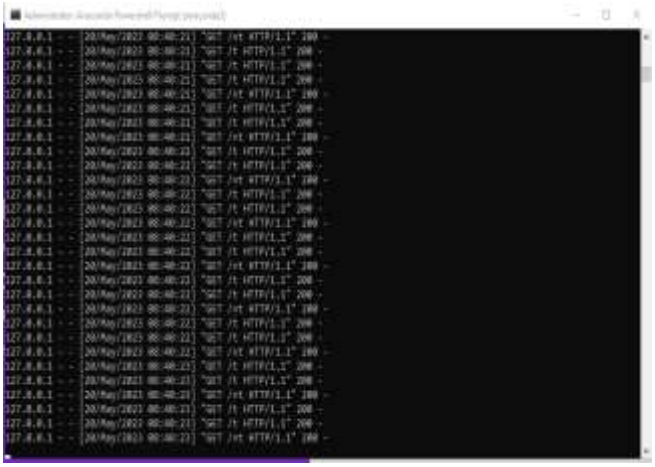

Fig 6. Execution Process

Fig 7. Music/Video Recommendation Interface

CONCLUSION

In conclusion, the project on Mental Health Assistance based on Emotion Recognition using Machine Learning offers a promising approach to address mental health issues. By utilizing machine learning techniques, specifically emotion recognition, the project demonstrates the potential to deliver personalized and contextually relevant support to individuals facing emotional distress. Through the development of an accurate emotion recognition model, the system effectively detects and classifies emotions, enabling targeted interventions and assistance. The integration of real-time emotion tracking and multimodal approaches further enhances the system's efficacy and adaptability.

The findings of this project highlight the significant role that technology can play in mental health support. By harnessing the power of machine learning and emotion recognition, individuals can receive timely and customized interventions that align with their emotional states. This can lead to improved emotional well-being, heightened self-awareness, and enhanced coping strategies.

However, it is crucial to acknowledge the limitations of the current system. Further research and development are necessary to enhance the accuracy and robustness of the emotion recognition model. Additionally, ethical considerations, privacy concerns, and user acceptance must be addressed to ensure responsible implementation of such technology.

Despite the challenges, the potential impact of this project is substantial. By bridging the gap between technology and mental health, we can provide accessible and individualized support to a wider population, ultimately contributing to improved mental well-being and a healthier society. Ongoing collaboration between technology experts and mental health professionals is vital for refining and expanding this technology to cater to the diverse needs of individuals seeking mental health assistance.

VI. FUTURE SCOPE

The project on facial emotion recognition and music/video recommendation based on emotions holds significant potential for future advancements and expansions. Firstly, the project can benefit from enhanced emotion recognition by incorporating advanced deep learning techniques and larger datasets to improve accuracy and robustness. Additionally, expanding the system to include multimodal emotion recognition, incorporating user feedback and adaptation mechanisms, and implementing context-aware recommendations can further enhance the user experience and personalization. Real-time emotion tracking in video streams, collaborative filtering for diverse recommendations, and integration with popular streaming platforms are also promising directions. Furthermore, the project's scope can extend beyond entertainment, finding applications in emotional well-being platforms and therapy sessions. Overall, the future scope of the project encompasses refining the emotion recognition model, expanding data sources and modalities, and integrating user feedback and contextual factors to provide accurate and personalized recommendations across various domains.

Acknowledgment

We thank Ms. Rajeshwari Kisan for their expertise and assistance in arranging and writing the manuscript. We also thank Mr. Sushant M for guiding us.

We express our gratitude to Dr.B. S.Halakarnimath

HOD CSE sdept. SGBIT and Dr. B. R. Patagundi, principal, SGBIT, Belagavi for funding our paper.

References

[1]. Shashwati Tidke, Ganesh Bhutkar, Dnyanal Shelke, Shivani Takale & Shraddha Sadke, “Mood-Based Song Recommendation System”, 2022.

[2]. Manoj Sabnis, Bhavesh Bhatia, Laveena Punjabi, Navin Rohra ,“Music recommendation through face recognition and emotion detection” 2022.

[3]. Adarsh Kumar Singh, Rajsonal Kaur & Devraj Sahu, “Real-Time Emotion Detection and Song Recommendation Using CNN Architecture”, 2021.

[4]. Akansha Bisht, Deepanshu Mishra, Harshit Gupta, Nancy Srivastava, Harris Kumar, “MUSIC RECOMMENDATIONS USING FACIAL EXPRESSION”, 2022

[5]. Ziyang Yu, Mengda Zhao, Yilin Wu, Peizhuo Liu, Hexu Chen,“Automatic Music Recommendation Based on Facial Micro-expression Recognition”, 2020.