International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0152

e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1149 - 1156

Emotion-Based Music Player

Netravathi K S[1], Bibi Haleema N[2], Priyanka[3], Madhushree[4], Priyanka R V[5]
[1,2,3,4]Computer Science and Engineering, The National Institute of Engineering, Mysuru, Karnataka, India.
[5]Assistant Professor, Computer Science and Engineering, The National Institute of Engineering, Mysuru, Karnataka, India.
Emails: *netrakaval@gmail.com[1], sunainhaleema786@gmail.com[2], mnpriyanka2003@gmail.com[3], madhush899@gmail.com[4], priyankarv@nie.ac.in[5]*

Abstract

Songs have always been a popular medium for communicating and understanding human emotions. Reliable emotion-based categorization systems can be quite helpful to us in understanding their relevance. However, the outcome of the research on Emotion-based music classification have not been the greatest. Here, we introduce EMP, a cross-platform emotional music player that play songs in accordance with the user's feelings at the time. EMP provides intelligent mood-based music player by incorporating emotion context reasoning abilities into our adaptive music engine. EMP revolutionizes how users interact with music, fostering deeper connections between emotions and musical experiences. Our music player is composed of three modules: the emotion module, the classification module, and the queue-based module. The Emotion Module analyses a picture of the user's face and uses the VGG16 algorithm to detect their mood with a precision exceeding 95%. The Music Classification Module gets an outstanding result by utilizing aural criteria while classifying music into 7 different mood groups. The queue module plays the songs directly from the mapped folders in the order they are stored, ensuring alignment with the user's mood and preferences.

Keywords: *VGG 16 Algorithm, Emotion Context, Intelligent, EMP.*

1. Introduction

The world of music has always been an integral element of our lives and it has the power to evoke emotions and feelings that are unique to everyone. In recent years, the field of music technology has seen tremendous growth and there have been numerous advancements in the utilization of machine learning algorithms to develop intelligent music systems. One such system is the emotion-based music player, which uses VGG16 to detect emotions in music and then plays a song derived from the identified emotional state. In this project, we will explore the evolutionary process of an emotion-based music player using VGG16 for the detection of emotions using Python. The system is designed to make use of a pre-trained VGG16 to analyze facial features of the users and predict the emotion. The predicted emotion will then be utilized to select and play the most appropriate songs from a pre-defined playlist that is associated with that emotional state. The main goal of this project is to provide a personalized and emotionally engaging music experience for the user. The potential applications of this system extend far beyond just music players and could be incorporated in a range of industries, including healthcare and entertainment. Human beings exhibit diverse music preferences tailored to their varying emotional states and activities. Whether engaged in physical exertion or seeking relaxation, individuals seek out specific genres and rhythms to suit their needs. It is within this context that the concept of an emotion-based music player system emerges, offering tailored musical experiences across a spectrum of scenarios including physical labor, stress management, music therapy, and academic endeavors. We introduce an emotion-based music player system tailored to address the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0152
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1149 - 1156

intricate emotional preferences of users, playing music aligned with their emotional states [1].

1.1 Related Work

In this study, researchers propose a novel approach to music recommendation based on emotions. They leverage deep learning models to analyze user preferences and emotional responses to music, enabling more personalized recommendations. By integrating emotion recognition techniques, it can accurately capture user's mood and tailor recommendations accordingly. This paper leverages CNN which possess the capability to autonomously discern pertinent features from images, eliminates the need for manual feature crafting [2].The research introduces a system that identifies user's emotional states and recommends music tracks accordingly. By analyzing factors like tempo, pitch, and lyrics sentiment, the system tailor's recommendations to match user's current mood. Through empirical evaluation, the study showcases the effectiveness of the proposed approach in enhancing user experience and satisfaction with music recommendation services. This research underscores the importance of incorporating emotional cues into recommendation systems to provide more personalized and engaging user experiences in the realm of music streaming platforms [3]. In this work introduced a dynamic framework for music recommendations grounded in human emotions. By training song selections for distinct emotional states derived from individual listening patterns, the researchers established a personalized approach to music curation. Employing a fusion of feature extraction methodologies and machine learning algorithms, the system adeptly discerns the emotional nuances of human faces depicted in input images. Once the mood is ascertained, the system seamlessly integrates by playing music tailored to the identified emotional state, thereby enhancing user engagement and satisfaction. [4]The paper proposes an emotion-based music player system utilizing facial recognition to detect user's emotions, achieving high accuracy with SVM classification aided by PCA and a polynomial kernel. It effectively integrates Haar features and PCA for dimensionality reduction and employs SVM classification with polynomial kernels for high accuracy emotion prediction. Real-time prediction involves considering 20 samples of the user's current emotion, enabling seamless music selection based on predominant emotional states. [5]In this research paper, it utilizes deep learning mechanisms, particularly focusing on facial expression recognition. By analyzing facial traits such as expressions, color, posture, and orientation, the system automatically creates music playlists in consideration of the real-time mental state of a person. Two classifiers, Haar Cascade, CNN and SVM, are employed for emotion detection, with comparative studies conducted based on trained datasets. The model comprises face discovery and facial component extraction components, enabling the system for identifying emotions [6]. Kundeti Naga Prasanthi et al. proposed an audio player which involves Haar cascade classification for face segmentation, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for feature extraction, and Euclidean distance calculation for emotion classification. The system aims to provide a more accurate and efficient method of selecting music tailored to the user's emotional state. [7]This paper proposes a 'smart music player' system employing artificial intelligence (AI) and facial expression recognition to recommend music based on the user's mood. It employs convolutional neural networks (CNNs) for facial expression detection and analysis, categorizing emotions into seven groups: happy, sad, neutral, surprise, fear, disgust, and angry. The system's architecture incorporates training Deep Neural Networks to recognize facial features and recommend music tracks accordingly. It uses the Stream lit framework for the user interface and connects to the Spotify API for song recommendations. The system achieves a 76% accuracy in emotion recognition. [8]This paper utilizes technologies such as React JS, Node JS, and Firebase for the frontend and backend. Leveraging algorithms such as Support Vector Machines (SVM) and OpenCV for facial recognition is used. Through algorithmic design, the system follows a step-by-step process from image upload to emotion detection to song

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0152
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1149 - 1156

selection. The user interface is intuitive, allowing users to easily upload images, detect emotions, and select songs. By providing a user manual, the system ensures seamless user interaction.

1.2 Existing System

While the concept of generating a playlist of songs in accordance with facial expressions using Convolutional [9] Neural Network (CNN) algorithms seems innovative, it comes with several drawbacks. Firstly, relying solely on facial gesture to determine emotions may not always accurately reflect the user's true mood. Additionally, CNN algorithms for emotion detection may not always be reliable or consistent. They can be prone to errors, especially in scenarios with fluctuating lighting conditions, facial angles, or cultural differences in facial expressions. This could lead to misinterpretations of the user's emotions, resulting in inappropriate song recommendations. [10] Furthermore, the automated generation of playlists based on detected emotions may lack the personal touch and customization that users desire. Music preferences are highly subjective and influenced by individual tastes, memories, and associations. Relying solely on facial expressions to curate playlists may overlook these nuances, resulting in a generic and potentially unsatisfying listening experience for the user. Additionally, there are privacy consideration associated with using facial recognition technology in this manner. Users may be uncomfortable with their emotions being continuously monitored and analyzed, provoking concerns about data security and consent. Overall, while the idea of leveraging facial expressions to tailor music playlists is intriguing, the drawbacks related to accuracy, personalization and privacy must be carefully considered and addressed for such a system to be truly effective and user-friendly.

1.3 Propsed System

Figure 1 displays the proposed application's system overview. The program will employ face detection to identify the user's emotion and assess the user's current mood before playing music from a music folder that was manually classified while the application was being created.

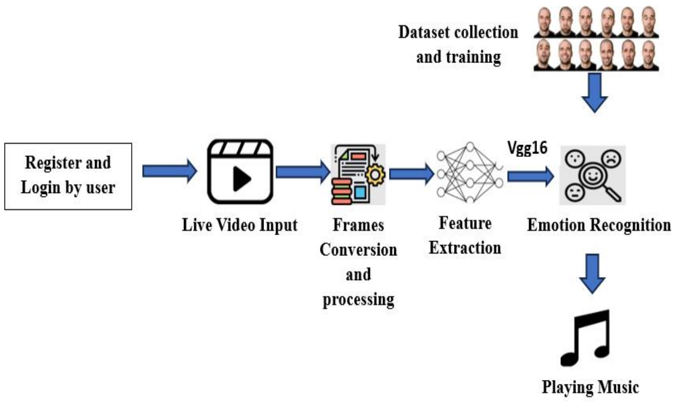

Figure 1 System Architecture

1.4 Dataset Collection

We collected an emotion dataset from reputable sources such as Kaggle, a popular platform for hosting datasets and machine learning competitions. The dataset comprises a diverse range of images depicting facial expressions representing various emotions, including happy, sad, anger, surprise, fear, disgust and neutrality (Table 1). Every image is tagged with the corresponding emotion category, providing valuable ground truth annotations for training and evaluating our emotion recognition model. [11] To ensure the dataset's quality and diversity, we conducted thorough screening and selection processes, prioritizing datasets with high-resolution images, balanced class distributions, and annotations provided by expert annotators or crowdsourcing platforms. Additionally, we verified the credibility and licensing of the datasets to comply with ethical and legal considerations regarding data usage. The collected dataset serves as a critical component in training and validating our emotion recognition model based on deep learning techniques. By leveraging this rich dataset, we aim to enhance the accuracy and robustness of our model in recognizing facial expressions across different individuals, demographics, and environmental conditions. This dataset acquisition process aligns with best practices in machine learning research, ensuring transparency, reproducibility, and ethical data handling throughout the project lifecycle.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0152
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1149 - 1156

Table 1 Collected Datasets

CLASS	DATASET COUNTS
Happy	7164
Sad	4938
Neutral	4982
Angry	3993
Fear	4103
Disgust	436
Surprise	3205

2. Method

2.1. Collect and Preprocess the Dataset

Collecting and pre-processing the set of data is an important step in developing an emotion-based music player using VGG16 for detecting emotions in video live stream input. Additionally, the model can be trained using live video stream data, which can be collected from various sources, such as webcams. Before using the collected data, it needs to be pre-processed to remove any noise or disturbances that may inhibit the emotion recognition process. For example, Video data can be pre-processed using techniques such as image resizing, normalization, and feature extraction from individual frames.

2.2. Build the VGG16 Model

The VGG16 model for the emotion-based music player will be built using the Keras deep learning library in Python. The model will consist of multiple convolutional layers (Figure 2) with ReLU activation, followed by max pooling layers to reduce dimensionality. The output will then be compressed and transmitted through fully connected layers with dropout regularization to prevent overfitting. [12]

Pseudo Code:

```
# Load pre-trained VGG16 model without the top (fully connected) layers

base_model=VGG16 (weights='imagenet', include_top=False, input shape= (224, 224, 3))

# Freeze the pre-trained layers

    For layer in base_model.layers:

        layer.trainable = False

# create a new model and add the VGG16 base

    Model = Sequential ()

    model.add (base_model)

# Add additional layers for recognizing emotional state

    model.add (Flatten ())

    model.add (Dense (256, activation='relu'))

    model.add (Dropout (0.5))

    model.add (Dense (7, activation='softmax')) // Assuming 7 classes for different emotions

# Compile the model

    model.compile (optimizer='adam', loss='categorical_crossentropy', metrics= ['accuracy'])

# Train the model with your dataset

# assuming you have data X_train, y_train and X_val, y_val for training and validation

model.fit (X_train, y_train, validation_data= (X_val, y_val), epochs=10, batch_size=32)

# Evaluate the model

    Loss, accuracy = model. Evaluate (X_test, y_test)
```

This pseudo code assumes you have preprocessed the data to fit the input shape of the VGG16 model (224x224x3). Replace X_train, y_train, X_val, y_val, X_test, and y_test with your actual training, validation, and test data.

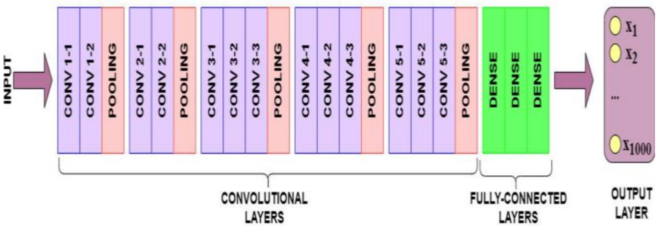

Figure 2 VGG16

2.3. Stream Video Input

Streaming video input is a fundamental aspect of the emotion-based music player that uses VGG16 for detecting emotions. The system requires a real-time video input to analyze the emotions of the person in the video stream and then selects music based on the detected emotion. [13] To achieve this, the system uses a video stream input from a webcam, which captures the live video feed of the user. The video stream is then processed using

OpenCV in Python to extract the features required for emotion detection. OpenCV offers a range of preprocessing functionalities including Standardizing, scaling and noise reduction, all of which contribute to enhancing the precision of the VGG16 model. The obtained features are then fed into the VGG16 model to estimate the user's emotional state accurately. [14] To ensure a smooth streaming experience, the system also utilizes a buffer to store the video input. The buffer allows for any latency or lag that might occur during the streaming process, thereby ensuring that the emotion detection and music selection process is not affected. Overall, the use of real-time video stream input is essential for the emotion based music player's functioning and ensures that the music selection accurately reflects the user's emotional state.

2.4. Play Music Based on Emotion

After detecting the emotions from the video input stream, the next step is to play music that matches the detected emotions. The emotion-based music player can be combined with the PyVLC media player to play music in real-time according to the detected emotions. [15] The PyVLC media player is a powerful media player library in Python that can play various types of media files and supports different video and audio codecs. By integrating the emotion-based music player with PyVLC, we can easily play the appropriate music file based on the emotions detected from the video input stream. For instance, if the model detects that the emotional state is happy, the music player can select upbeat and joyful music from a playlist, while sad emotions can trigger the player to select mellow and calming music.

2.5. User Emotion Classification

Face Detection: The primary objective of face detection is to locate human faces within images. This process typically involves identifying facial features such as the nose, mouth, and eyes, which serve as initial steps in face detection. Utilizing the sophisticated VGG16 Algorithm for facial detection ensures reliable results. This algorithm employs a machine learning-based object detection method, which requires a substantial number of positive photos for training the classifier. Additionally, negative images depicting objects without faces are also utilized.

Feature Extraction using VGG16 Method: Convolutional neural networks (CNNs) are a prevalent type of deep neural network fundamentally used for visual perception tasks in deep learning. CNNs operate based on a shared-weight architecture of convolution kernels or filters, which slide across input features to produce translation-equivariant responses known as feature maps. VGG16s are a type of CNN, with multilayer perceptron's adapted into their architecture. Multilayer perceptron's typically refer to fully connected networks, where each neuron in a layer is connected to every other neuron in the layer above. However, such networks are prone to overfitting due to their high connectivity. VGG16s employ a novel regularization strategy by leveraging the hierarchical structure of data to construct patterns of increasing complexity from smaller and simpler patterns imprinted in their filters.

User Emotion Recognition: Many platforms utilize facial expression recognition as a method for emotion analysis. Fisher Face is a technique rooted in principal component analysis and linear discriminant analysis principles. It involves categorizing and reducing photographic data before allocating it into appropriate groups, ultimately recording statistical values.

Emotion Mapping: Facial expressions can be categorized into basic emotions such as anger, happy, fear, neutral, sadness, disgust and surprise. The user's expression is compared to expressions in the dataset, thereby enabling emotion mapping.

2.6. VGG16 Working

Detecting faces is a popular topic with many practical applications. In today's smartphones and PCs, face detection software is already built in to help validate the user's identity. In addition to determining the user's age and gender and using some extremely amazing filters, several applications can record, recognize, and process faces in instantaneously. For feature extraction, VGG16 is utilized. For the emotion recognition module, we must train the system using datasets of seven emotions. VGG16 has the special ability to

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0152
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1149 - 1156

apply automatic learning to extract traits from dataset images for model building. VGG16 can provide an internal, two-dimensional visual representation. On this matrix, operations in three dimensions are carried out for teaching and testing reasons. Five-Layer Model: As its name indicates, this model has five layers. (Figure 3) A convolutional and a max-pooling layer, a fully connected layer with 1024 neurons, an output layer with 7 neurons, and a soft-max activation function are the layers that make up each of the first three phases. For the initial convolutional layers, 32, 32, and 64 5*5, 4*4, and 5*5 kernels, respectively, were used. Max-pooling layers come after convolutional layers, and they each employed kernels with 3*3 dimensions, a stride of 2, and the ReLu activation function.

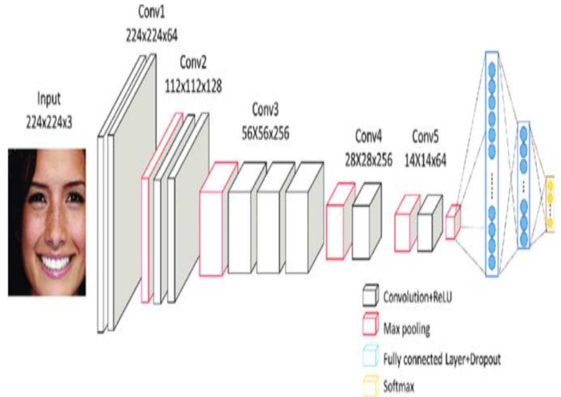

Figure 3 Emotion Recogntion using VGG16

The final layer will use softmax activation to output the predicted probabilities for each emotion class. The model will be trained using the dataset described previously, with a batch size of 32 and an Adam optimizer. The accuracy of the model will be evaluated using the validation set, and the best-performing model will be used to predict emotions in the live video stream input and play music accordingly.

3. Results and Discussion

3.1. Results

The Figure 4 indicate that the VGG16 model achieved a high level of accuracy in predicting the emotions reaching above 90%. The model demonstrated a strong ability to classify emotions such as happiness, sadness, anger, neutral, disgust, fear, and surprise with a significant level of precision. This high accuracy recommend that the model has effectively learned meaningful patterns and features associated with different emotions.

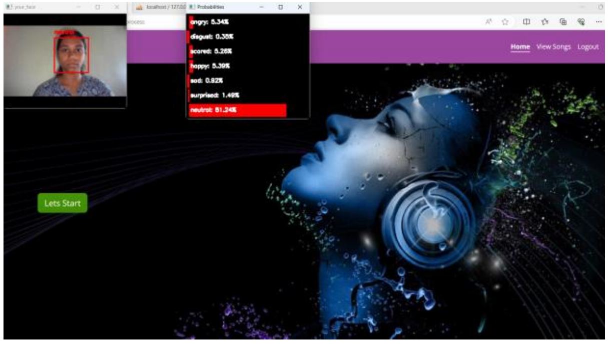

Figure 4 Capturing Image and Detect Emotion

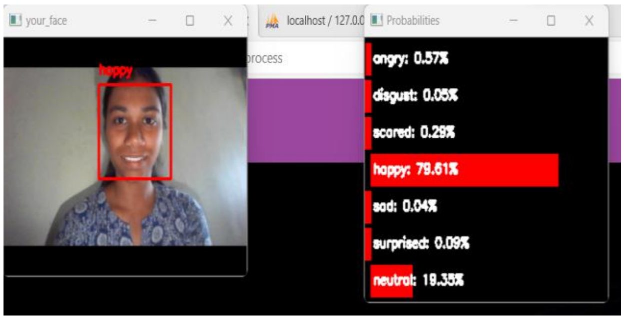

Figure 5 Happy Emotion Detection

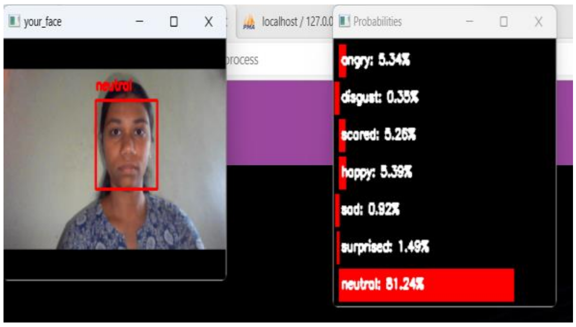

Figure 6 Neutral Emotion Detection

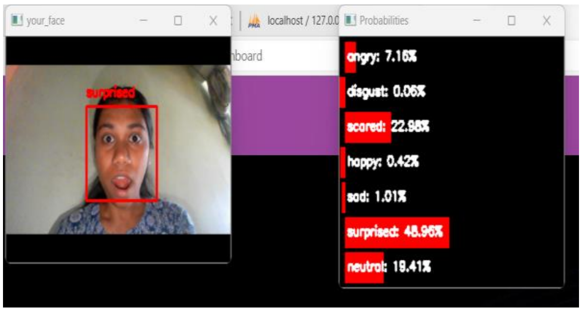

Figure 7 Surprised Emotion Detection

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0152
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1149 - 1156

3.2. Discussion

The discussion highlights the practical implications of such accurate emotion detection in the music player. As shown in Figure 5, Figure 6 and Figure 7 by reliably predicting the users emotional state (eg. Happy, Neutral, Surprised, Sad, Angry, Disgust, and Scared) the music player can provide a highly personalized and enjoyable experience. It can automatically select music tracks that precisely match the user's emotions, creating a seamless and immersive listening experience. This approach eliminates the need for the user to actively search for music that aligns with their mood, greatly enhancing convenience and user satisfaction. However, it is important to address the limitation of the current system regarding real-time input-based categorization. While achieving high accuracy in emotion detection, the system does not incorporate real-time indicators such as facial expression analysis to capture the user's changing emotional state. Integrating real-time emotion detection techniques could significantly enhance the system's ability to adapt and respond to the user's evolving emotions and preferences, leading to an even more refined and tailored music selection.

Conclusion

This study looked at an innovative method of classifying music based on the emotions and facial expressions of the listeners. It is advised to use neural networks and visual processing to categorize the seven fundamental universal emotions conveyed by music—happiness, sad, anger, disgust, surprised, scared, neutrality. First, the input image is run through a face detection algorithm. A feature extraction method based on image processing is then used to recover the feature points. Finally, instructions are supplied to a neural network to identify the emotion present in a collection of values obtained by analyzing the acquired feature points. Although the research is still in its early stages, success in the field of emotion identification and playing music from the supplied dataset is anticipated.

Acknowledgements

The authors express their sincerest gratitude to Dr. Rohini Nagapadma, the Principal, and Dr. Anitha R, Professor& HOD, from the Department of CSE at NIE, for their relentless support and encouragement. Additionally, we extend immense pleasure in thanking Ms. Priyanka R V, Assistant Professor in the Department of CSE at NIE, for her valuable suggestions and guidance during the process of this project study.

References

[1]. Jaladi Sam Joel, B. Ernest Thompson, Steve Renny Thomas, T. Revanth Kumar, Bini D. Emotion based Music Recommendation System using Deep Learning Model. International Conference on Inventive Computation Technologies (ICICT). Doi: 10.1109/ICICT57646.2023.10134389

[2]. Vijay Prakash Sharma et al., Emotion-Based Music Recommendation System (2021). 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). Doi: 10.1109/ICRIRT051393.2021.9596276

[3]. ShanthaShalini. K et al., "Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques" (2021). Turkish Journal of Computer and Mathematics Education vol. 12, no. 2, pp. 912–917, Apr.

[4]. Seshaayini K et al., Emotion Recognition Based Music Player (2023). Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT). Doi:10.1109/ICECCT56650.2023.10179716

[5]. Prachi Vijayeeta, Parthasarathi pattnayak. A Deep Learning approach for Emotion Based Music Player (2022). OITS International Conference on Information Technology (OCIT). Doi:10.1109/OCIT56763.2022.00060

[6]. Kundeti Naga Prasanthi et al. Machine Learning Techniques based Audio Player to Soothe Human Emotions (2022). International Conference on Sustainable

Computing and Data Communication Systems(ICSCDS).Doi:10.1109/ICSCDS53736.2022.9760912

[7]. Anushree K et al. Artificial Intelligence (AI) Enabled Music Player System for User Facial Recognition (2023). 4th International Conference for Emerging Technology (INCET) Belgaum, India, May26-28, 2023. Doi:10.1109/INCET57972.2023.10170476

[8]. Vinay p et al. Facial Expression Based Music Recommendation System (2021). International Journal of Advanced Research in Computer and Communication Engineering. Doi:10.17148/IJARCCE.2021.10682

[9]. Serhat Hizlisoy et al. Music emotion recognition using convolutional long short term memory deep neural networks (2021). Engineering Science and Technology, an International Journal Volume 24, Issue 3. Doi.org/10.1016/j.jestch.2020.10.009

[10]. Sulaiman Muhammad et al. Real Time Emotion Based Music Player Using CNN Architectures. 6th International Conference for Convergence in Technology (I2CT). Doi:10.1109/I2CT51068.2021.9417949

[11]. Sreenivas, V., Namdeo, V. & Kumar, E.V. Group based emotion recognition from video sequence with hybrid optimization based recurrent fuzzy neural network (2020). J Big Data 7, 56. Doi.org/10.1186/s40537-020-00326-5.

[12]. Soumya K, Suja Palaniswamy. Emotion Recognition from Partially Occluded Facial Images using Prototypical Networks (2020). Second International Conference on Innovative Mechanisms for Industry Applications. Doi:10.1109/ICIMIA48430.2020.9074962

[13]. Ishwar More et al. Melomaniac- Emotion Based Music Recommendation System (2021). IJARIIE, no. 3, pp.1323–1329.

[14]. Madhuri Athavle et al. Music Recommendation Based on Face Emotion Recognition (2021). Journal of Informatics Electrical and Electronics Engineering. Vol. 02, Iss. 02, S. No. 018, pp. 1-11

[15]. Rahul Arya, Chandradeep Bhatt, Mudit Mittal. Music Player Based on Emotion Detection Using CNN. IEEE North Karnataka Subsection Flagship International Conference (NKCon). Doi: 10.1109/NKCon56289.2022.10126761