International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Emotion Recognition based Music Player using Convolutional Neural Network

Dr. Aparna K[1], Akash D Naik[2]

[1]Associate Professor, Department of Master of Computer Application, BMS Institute of Technology and Management, Bengaluru, Karnataka
[2]Student, Department of Master of Computer Application, BMS Institute of Technology and Management, Bengaluru, Karnataka

---

Abstract - This research study focuses on the classification that may be accomplished through the detection of human facial expression using a Convolutional Neural Network (CNN). The network is able to classify emotions and play music based on the user's identified facial expression. The suggested approach successfully and implicitly classifies the expression into happy, sad, angry, disgusted, fear, and neutral by leveraging the Convolutional Neural Network architecture. By playing music that is customized to the user's current mood, the smart music player has the potential to enhance the listening experience. The player might also be used to assist people in controlling their emotions by playing music that, depending on how they are feeling at the time, can help them unwind, feel joyful, or reduce tension. In order to play the appropriate songs from a remote database depending on the user's mood, the smart music player primarily employs the system camera to detect the user's facial expression. The database will play a random song from the happy playlist if the system determines that the user is in a happy mood. This process is repeated for the other five emotions.

Key Words: Facial expression, Emotion Detection, Convolutional Neural Networks, Different Playlists.

1.INTRODUCTION

In recent years, the fields of facial expression recognition and machine learning have made significant advancements, enabling the development of intelligent systems that can interpret human emotions and respond accordingly. Music has long been recognized as a powerful medium for expressing and evoking emotions, and its influence on human well-being is well-documented. Integrating facial expression recognition with a smart music player offers a promising avenue for creating personalized and emotionally engaging music experiences.

The aim of this research paper is to present a novel approach to music playback that leverages facial expression recognition techniques. By analyzing the user's facial expressions and associating them with specific emotional states, the smart music player can intelligently select and play songs from a curated playlist that aligns with the user's emotions in real-time. This innovative system not only enhances the music listening experience but also opens up new possibilities for adaptive and personalized music recommendation systems.

The motivation behind developing a smart music player based on facial expression recognition stems from the desire to create a more immersive and interactive music listening experience. Traditional music players rely on manual input or pre-defined preferences to select songs, which can often be time-consuming or fail to capture the user's current emotional state. By incorporating facial expression recognition, the smart music player can dynamically respond to the user's emotional cues, offering a more intuitive and seamless interaction.

In this paper, initially the methodology is outlined where the previous research and existing approaches related to facial expression recognition and music recommendation systems were made, then the data collection process and facial expression recognition using a model known as CNN (Convolutional Neural Network) is implemented, later it is integrated with the music playback system. The results of experiments are also presented, evaluating the effectiveness of the facial expression classification model and the overall music recommendation system. Finally, the implications of our findings, limitations of the current implementation, and potential directions for future research in this field will be discussed.

In the smart music player application, CNNs are employed for facial expression recognition. They are trained on a dataset of labeled facial images to learn patterns and features associated with different emotions.

Overall, the outcomes of this research can include technological advancements, empirical findings, and contributions to the existing knowledge, with potential implications for fields such as music technology, emotion recognition, and user experience design.

2. RELATED WORK

The sources in the area of facial expression analysis-based music recommendation and emotion recognition are astounding. They discuss several procedures, evaluation strategies, and difficulties that you can use as a starting point for your research paper on the project for a smart music player. An overview of the work on the facial expression-based smart music player is given in this section.

[1] This survey investigates the creation of a smart music player that makes music recommendations depending on the user's emotional state using facial emotion detection technology. The major goal of the article is to develop a system that can recognize the user's emotions from their facial expressions and then recommend music that is in tune with their

emotional state or mood. The music player seeks to give consumers a more customized and interesting experience by combining face emotion detection technologies.

[2] This in-depth review paper introduces a cutting-edge smart music player that uses facial expression recognition to make in-the-moment song suggestions based on the user's emotional state. The primary goal of the article is to create and put into practice a music player that can

recognize and analyze a user's facial expressions in real-time in order to determine their emotional state. By doing this, the smart music player hopes to give users a more unique and interesting music-listening experience.

[3] This paper describes the design and implementation of a smart music player that uses facial expression analysis to create a personalized and context-aware music listening experience. The paper's major goal is to develop a music player that can read and analyze facial expressions in real-time to determine the user's emotional state or mood. This will improve the user's listening experience by making appropriate music recommendations that are in line with their feelings.

[4] The goal of research is to create an intelligent music player application for Android smartphones that can instantly assess a user's mood and offer tailored music suggestions in response. The research entails putting emotion analysis algorithms into practice while capturing user facial expressions or other clues using the Android device's built-in camera or other sensors. The intelligent music player uses a recommendation system to propose music tracks or playlists that correspond to the user's emotional state after analyzing the user's emotions.

[5] The goal of this study is to create a smart music player system that makes use of facial recognition and artificial intelligence. The project's main goal is to develop a music player system that uses AI algorithms and facial recognition technology to recognize and comprehend the user's emotions and facial expressions in real-time. The intelligent music player uses a recommendation algorithm after performing facial expression analysis to offer songs or playlists that are appropriate for the user's current emotional state. Based on this research, the system can make music track recommendations that complement the user's emotional state, resulting in a more unique and satisfying musical experience.

[6] The multi-stream convolutional neural network (CNN) fusion network described in this paper presents a novel method for identifying infant facial expressions. Creating a reliable and efficient system for identifying newborn's facial expressions is the main goal of the project. Early childhood development studies must have a thorough understanding of newborn emotions and expressions since they can reveal important information about young children's emotional health. It demonstrates the capability of the multi-stream CNN fusion network to recognize newborn facial expressions with accuracy and efficacy.

[7] This study proposes a brand-new method for identifying faces that makes use of normalized segment classification detection in face photos and multiscale facial expression analysis. Creating a sophisticated face recognition system that can accurately identify people while taking facial emotions into account is the main goal of the project. The face recognition process can be improved and made more accurate by taking into account facial expressions, the essential indicators play a vital role in communicating emotional states.

3. METHODOLOGY

Initially, a dataset containing six different categories of facial expression grayscale images was downloaded. Subsequently, variety of image pre-processing techniques are implemented in order to clean and enhance the quality of images. Following that, the Convolutional Neural Network model was trained by using the sample images. Below mentioned figure is the proposed model for Smart Music Player:

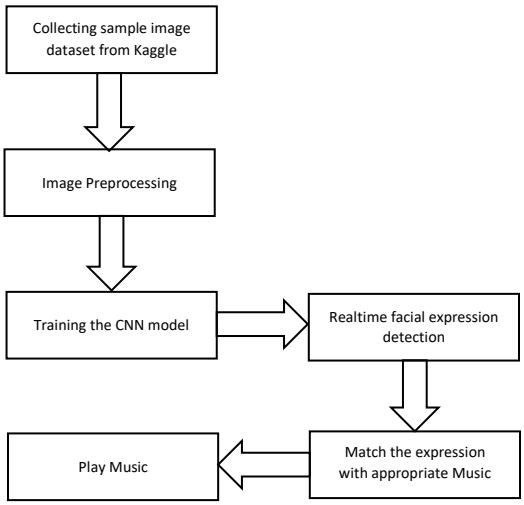

Fig -1: Proposed Model for Smart Music Player

A. Dataset

FER2013 is a widely used dataset for facial expression recognition. It was introduced in 2013 by Pierre-Luc Carrier and Aaron Courville. The dataset consists of grayscale human face images, each associated with one of seven distinct facial expressions: Angry, Disgust, Happy, Fear, Sad, Neutral, and Surprise. It contains a total of 35,887 images which are divided into 3 subsets called Training set, public test set and private test set.

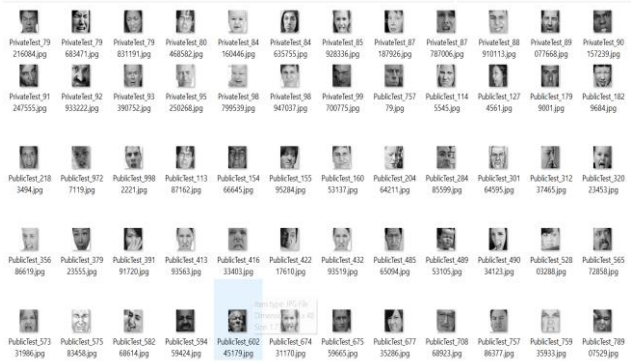

Fig -2: Dataset

B. Image Preprocessing and Model Architecture

Image preprocessing is an essential step before feeding images into a CNN model. The purpose of preprocessing is to enhance the quality of the images, reduce noise, and make the data more suitable for the model. The typical image preprocessing steps for facial expression recognition using CNNs include:

a) Resizing: The input images may have different sizes. It is necessary to ensure that they all are of fixed size so as to maintain the consistency during training. Commonly, images are resized to a square format, e.g., 48x48 pixels.

b) Grayscale Conversion: In facial expression recognition, color information might not be crucial, and using grayscale images can reduce the computational cost. Converting RGB images to grayscale reduces the image's dimensionality.

c) Cropping: If there are any unwanted parts in the images, cropping can be used to focus only on the relevant facial region.

d) Face Detection and Alignment: In some cases, it might be necessary to detect faces in the images and align them properly. This step ensures that the facial expressions are consistently centered and oriented.

e) Normalization: Normalizing the pixel values of the image ensures that they are within a specific range. The most common normalization technique is to scale the pixel values to [0, 1] or [-1, 1].

The Model architecture for facial expression recognition using a CNN consists of several layers. The common architecture is as follows:

a) Input Layer: The preprocessed grayscale photographs of the subjects' expressions are fed into this layer. The size of the scaled photos (for example, 48x48 pixels) determines the input size.

b) Convolutional Layers: The main function of convolutional layers is to extract features and patterns from the input images that are related to space. Multiple filters, also known as kernels, are contained in these layers. These kernels move through the image and look for patterns like edges, textures, and facial features. Each filter creates a feature map that draws attention to particular traits seen in the input data.

c) Activation Function: To add non-linearity, an activation function is used to each convolutional layer's output. The activation function in CNNs is frequently Rectified Linear Activation (ReLU).

d) Pooling Layers: Following activation, the spatial dimensions of the feature maps are reduced by down sampling them using pooling layers. A typical technique called MaxPooling chooses the highest value possible from a particular window while keeping the most important characteristics.

e) Flatten Layer: By converting the 2D feature maps into a 1D vector, the flattened layer creates an input for the fully connected layers.

f) Fully Connected Layers (Dense Layers): These layers do categorization and feature processing on the flattened features. All of the neurons in the preceding layer are linked to every neuron in the dense layer.

g) Output Layer: output layer has neurons equal to the number of facial expression classes. It uses the SoftMax activation function to produce class probabilities. The class with the highest probability is predicted as the facial expression.

C. Model Training

During the training process of the CNN model, the training subset is employed to update the model's weights iteratively. Simultaneously, the validation subset is utilized to assess the model's performance throughout the training phase, allowing for the detection and prevention of overfitting. The CNN model is trained using the training dataset, which serves as the basis for learning and adjusting its parameters. The model is fed with batches of images, and the weights are updated after each batch to minimize the loss. The training process is typically done for multiple epochs, where each epoch represents one pass through the entire training dataset. After each epoch, the model's performance is evaluated on the validation dataset. Monitoring the validation performance helps identify when the model starts to overfit, allowing for early stopping or adjusting hyperparameters.

D. Model Testing

A separate dataset known as test dataset is prepared to cross verify the performance of trained model, which contains images that the model has never seen during training. There are some common evaluation methods to determine the performance of the model, they are as follows:

a) Accuracy: The accuracy metric represents the percentage of correct predictions made by the model over the total number of test samples. It is the most straightforward evaluation metric and provides an overall view of the model's performance.

A = (n) / (N)

Where A represents Accuracy, n and N represents the number of correct predictions and the total number of test samples respectively.

b) Confusion Matrix: A confusion matrix is a table that shows the number of correct and incorrect predictions made by the model for each class (facial expression). It provides more detailed information about the model's performance and helps identify which facial expressions are frequently misclassified.

c) ROC Curve: The Receiver Operating Characteristic (ROC) curve is a graphical representation of the model's performance at various classification thresholds. As the threshold for classifying a sample is changed, it shows the true positive rate (TPR or recall) versus the false positive rate (FPR).

d) Precision: Precision is the ratio of accurate positive predictions to all of the model's positive predictions. It counts how many of the samples that were expected to test positive actually did.

Precision = (True Positives) / (True Positives + False Positives)

E. Facial expression Recognition

Once the model is trained and validated, it can be deployed in real-time applications. In such applications, the FER module captures real-time video frames from a camera feed, preprocesses the frames, and passes them through the trained model to predict the facial expression.

The predicted facial expressions are typically represented as emotion labels, such as "Happy," "Sad," "Angry," "Neutral," "Disgust," "Fear,".

Facial Expression Recognition modules are often integrated into effect-aware systems that utilize emotion recognition to adapt their behavior or responses based on the detected emotions. A smart music player uses Facial Expression to recommend music that matches the user's current emotional state.

F. User Interface

The User Interface (UI) module focuses on providing an interactive platform for users to experience Ai enabled music system. The UI module is responsible for providing user friendly interface and also provides good command over the system.

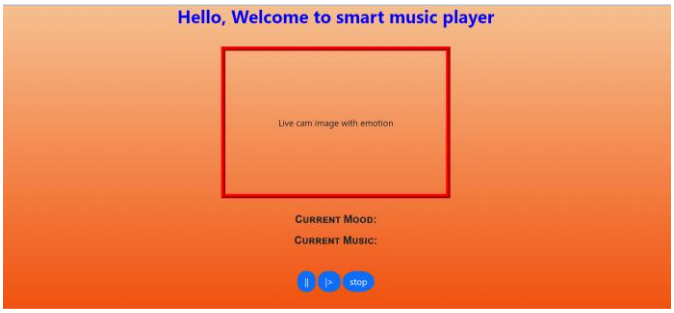

Fig -3: User Interface

Features of User Interface module include:

a) Python Flask: Python Flask is a web framework that helps you build web applications and websites using the Python programming language. It provides a set of tools and libraries to handle tasks like handling web requests, routing URLs, and generating HTML pages. With Flask, we can create interactive and dynamic web pages that respond to user actions and input. It allows you to connect the backend (server-side) Python code with the frontend (client-side) HTML, CSS, and JavaScript code.

b) Device camera usage: The camera continuously captures video frames at a specific frame rate for example 30 frames per second. Each frame is essentially an image of the user's face captured in real-time.

c) Emotion Recognition: After that, the collected frames are analyzed using a trained deep learning model for identifying facial expressions. The algorithm examines each frame to identify and categorize the user's facial expression into happy, sad, angry, disgusted, neutral, and terrified.

d) Music Playback: Based on the emotion detected by the model a random music from the appropriate playlist that matches the user's mood is being played. This process repeats until the user quits the web page.

4. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

The results of Smart music player can be quite diverse and impactful. Here are some of the key outcomes and benefits the system can offer:

a) Personalized Music Experience: The smart music player can provide a highly personalized music experience based on the user's real-time facial expressions and emotional state. It selects music that aligns with the user's emotions, enhancing the overall listening experience.

b) Enhanced User Engagement: The system's real-time response to the user's emotions fosters a more engaging and interactive music experience. Users may feel a deeper connection to the music as it resonates with their current mood.

c) Reduced Stress and Anxiety: By playing soothing and calming music when the system detects stress or anxiety in the user's facial expressions, the smart music player may help promote relaxation and reduce negative emotions.

d) Adaptive Playlists: As the user's emotions change throughout the day, the smart music player dynamically adjusts the playlist to provide a continuous flow of music that aligns with the user's evolving emotional states.

e) Accessibility and Inclusivity: For individuals who may have difficulty expressing their emotions verbally, the facial expression-based system provides an alternative means of communication and engagement with music.

To gauge the effectiveness of the proposed system, comprehensive tests were conducted. The system's performance was evaluated using a variety of metrics. These metrics provided insights into how well the system performed and allowed for a thorough assessment of its capabilities.

a) Accuracy: Accuracy is commonly computed as the percentage of correctly classified facial expressions out of the total number of expressions in the dataset or during real-time testing. For instance, if the model accurately identifies 80 out of 100 facial expressions, the accuracy would be calculated as 80%. This metric provides a straightforward measure of the model's overall performance in classifying facial expressions.

High accuracy is essential for the system to make accurate music recommendations based on the user's emotions. A high accuracy rate ensures that the music player responds appropriately to the user's facial expressions, leading to a more personalized and engaging music experience

b) Confusion Matrix: A confusion matrix is a table that shows the number of correct and incorrect predictions made by the model for each class (facial expression). It provides more detailed information about the model's performance and helps identify which facial expressions are frequently misclassified.

Following are the components of confusion matrix.

- True Positive (TP): The number of samples that belong to a particular class and are correctly predicted as that class by the model.

• False Positive (FP): The number of samples that do not belong to a particular class, but are incorrectly predicted as that class by the model.

• True Negative (TN): The number of samples that do not belong to a particular class and are correctly predicted as not belonging to that class by the model.

• False Negative (FN): The number of samples that belong to a particular class, but are incorrectly predicted as not belonging to that class by the model.

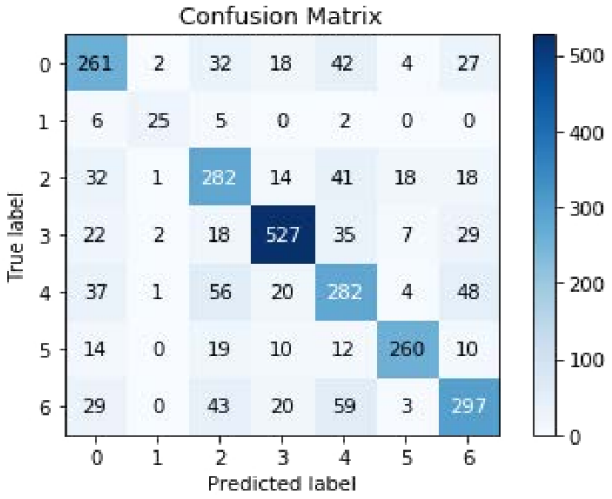

Fig -4: Representation of Confusion Matrix

c) Precision: Precision is calculated as the ratio of true positive predictions (correctly predicted positive samples) to the total number of positive predictions made by the model (true positive plus false positive). It represents how well the model avoids false positives and correctly identifies positive samples for a particular class.
Mathematically, precision is defined as:

Precision = True Positives / (True Positives + False Positives)

A high precision score means that less erroneous positive predictions are being generated by the model, and that the positive predictions it makes for a given class are more likely to be accurate.

d) AUC-ROC Curve: Area Under the Receiver Operating Characteristic curve is referred to as the AUC-ROC curve. The AUC-ROC curve is a graphical representation that measures a machine learning model's capacity to distinguish between positive and negative data and aids in performance evaluation.

The AUC-ROC curve summarizes the performance of the model across all possible threshold values. The area under the ROC curve (AUC) is a single scalar value that quantifies the overall performance of the model. A perfect classifier would have an AUC of 1, indicating that it has perfect discrimination ability between positive and negative samples. An AUC of 0.5 represents a random classifier, and an AUC below 0.5 indicates that the model's performance is worse than random.

5. FINDINGS AND IMPLICATIONS OF THE RESEARCH

The research helped in finding more advanced and interactive way to listen to the music. The key findings and their implications are mentioned:

a) Accuracy and Performance:

The study demonstrates how well the facial expression recognition model performs when it comes to correctly identifying and classifying diverse facial expressions that represent various emotions. High accuracy is attained using a trained Convolutional Neural Network (CNN) model, highlighting its amazing accuracy in identifying emotions from facial photos.

b) Real-time Music Recommendation:

Real-time music selection based on the user's current emotional state is made possible by the integration of the facial expression recognition module with the music recommendation system. Using camera data, the system dynamically reacts to the user's emotions to create a unique musical experience.

c) User Experience:

Users reported positive feedback regarding the interactive and engaging nature of the facial expression-based smart music player. The music recommendations aligned with their emotions enhanced the enjoyment and emotional connection with the music.

d) Personalized Music Experience:

The facial expression-based smart music player offers a personalized music experience tailored to the user's emotional state. This has implications for enhancing user satisfaction and engagement with the music platform.

e) Emotional Well-being:

By playing music that matches the user's emotions, the smart music player may have positive implications for the user's emotional well-being. Music is known to have therapeutic effects, and aligning music with emotions may help users manage stress and improve their mood.

f) Music Industry and Entertainment:

The integration of facial expression recognition with music recommendation systems may influence the way music content is delivered and marketed. Emotion-based music curation could lead to new business models and personalized music streaming services.

6. CONCLUSION AND FUTURE WORK

The conclusion of the smart music player system is that it successfully demonstrates the potential of combining facial expression recognition with music recommendation to create a personalized and emotionally engaging music experience. The system utilizes a trained convolutional neural network (CNN) model to accurately recognize and classify facial expressions corresponding to various emotions.

By analyzing the user's facial expressions in real-time through a camera, the system dynamically selects and plays music that aligns with the user's current emotional state. This real-time music recommendation based on facial expressions

enhances user satisfaction and emotional connection with the music platform.

The system shows promising results and has practical applications in enhancing music listening experiences, with the potential to shape the future of emotionally intelligent technologies in various domains beyond music. However, further research and development are needed to refine the system, address any limitations, and explore its wider applications in different contexts.

Future work could involve the system to be implemented as an end user android application where users can customize their playlist according the music available in their local storage. This will also help to run the application even if there is a weak network particularly in remote places. API's like spotify can be used to fetch the music that already contains music classified in various languages. This will help user to explore new music in their preferred language.

REFERENCES

1. Shlok Gilda, Husain Zafar, Chintan Soni and Kshitija Waghurdekar "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation" 2017 presented at IEEE WiSPNET 2017 conference
2. Natha Harika, T Kishore Kumar "Real Time Smart Music Player Using Facial Expression" 2022 published on IEEE.
3. Mr. N.Santosh ramchander, M.sowrabh, B.sarika, G.sai priya, M.shashank "Smart Music Player Based on Facial Expression" 2020 published on advanced science letters
4. Maruthi Raja S K, Kumaran V, Keerthi Vasan A, Kavitha N "Real Time Intelligent Emotional Music Player using Android" 2017 published on journal for research
5. Dr. A.Rehash Rushmi Pavitra, Anushree K , Akshayalakshmi A V R, Vijayalakshmi K "Artificial Intelligence (AI) Enabled Music Player System for User Facial Recognition" 2023 4th International Conference for Emerging Technology (INCET) Belgaum, India
6. Lirong Zhang, Chao Xu, Shao Li "Facial Expression Recognition of Infants Based on Multi-Stream CNN Fusion Network" 2020 IEEE 5th International Conference on Signal and Image Processing
7. Raman "Multiscale Facial Expression Based Face Recognition using Normalized Segment Classification Detection in Face Images" 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)
8. https://www.kaggle.com/datasets/msambare/fer2013