Personalized Music Playlists via Deep Learning EmotionDetection

M. Spoorthi*, Harshitha H. M., Pooja R., Anusha M. K., Preethi R.

Department of Information Science and Engineering, Maharaja Institute of Technology, Thandavapura, Mysore, India.

Emails: mspoorthi03@gmail.com ; harshitha@gmail.com; poojarchnagar8@gmail.com; anu.anushakempraj@gmail.com;preethi4mn20is020@gmail.com

Abstract

Music holds significant sway in enriching the lives of individuals, serving as a vital source of entertainment for enthusiasts and listeners alike. Moreover, it transcends mere amusement, often adopting a therapeutic role in people's lives. In the ever-evolving landscape of music and technology, this project emerges as a groundbreakingendeavor, driven by the profound impact music holds on individuals' lives. Leveraging technological advancements in music players, such as playback control and genre classification, our focus is on revolutionizingplaylist creation. Instead of the laborious manual curation of playlists, we introduce automation based on users' emotional states, identified through real-time facial expression analysis via a camera. The human face, a rich source of mood indicators, becomes the key input for our system. By directly extracting emotional cues from facial expressions, the project aims to swiftly deduce the user's emotional state, crafting a tailored playlist without the need for time-consuming manual efforts. Implemented through deep learning using VGG16 model, the system ensures intricate emotion recognition from image input. Python, OpenCV, and Keras facilitate seamless video processing and deep learning functionalities, complemented by a music player library for smoothplayback control. This amalgamation of computer vision and deep learning delivers an interactive music player that dynamically selects tracks aligned with users' real-time emotional expressions, offering a personalized and immersive musical experience.

Keywords: Visual Geometry Group (VGG16); Computer Vision; Emotion Detection; Deep Learning

1. Introduction

The field of emotion-based music recommendation has witnessed significant advancements, driven by the growing interest in leveraging human emotions to enhance personalized experiences. Understanding and interpreting human emotions, particularly through facial expressions, play a pivotal role in designing interactivesystems like emotion-based music players. Emotion detection systems offer a rich array of applications across diverse domains, including education, healthcare, and entertainment. In recent years, facial expression recognition (FER) has emerged as a prominent technique for detecting human emotions. By analyzing facial cues and expressions, FER systems can infer users' emotional states and tailor experiences accordingly. This paper explores the development of an emotion-based music player system that integrates FER algorithms, specifically utilizing convolutional neural networks (CNNs) and Local Binary Patterns Histograms (LBPH), to accurately identify users' emotions from facial images. Unlike conventional approaches that often rely solely onpre-trained models or singular techniques, this system combines the strengths of CNNs and LBPH to achieve robust emotion detection. Furthermore, we focus on leveraging the efficiency and effectiveness of the VGG16 architecture, a popular CNN model, for feature extraction and classification tasks. By fine-tuning the VGG16 model on a custom dataset, we aim to enhance the system's accuracy in recognizing various emotional states.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

Through experimental validation and analysis, we demonstrate the efficacy of the proposed approach in accurately identifying and interpreting human emotions.This emotion-based music player system holds promise for creating immersive and personalized user experiences by recommending music tracks that resonate with users' emotional states, thereby enhancing engagement and satisfaction in interactive platforms.

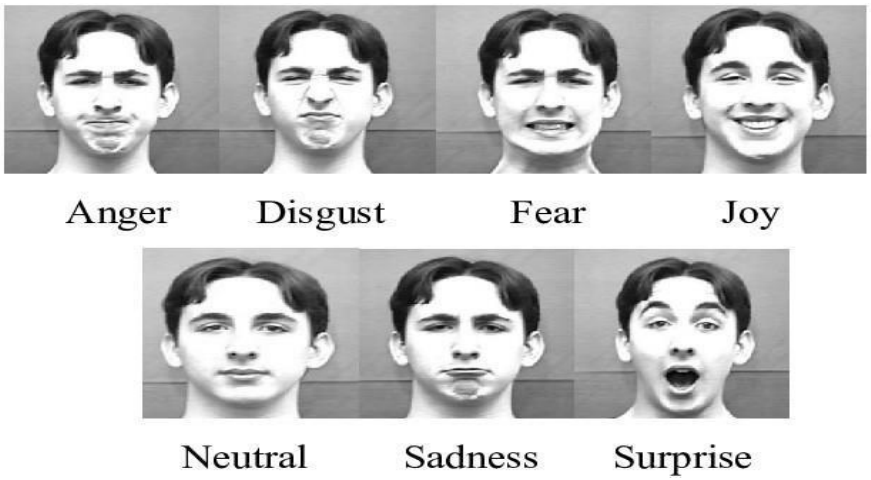

Figure 1: seven basic human emotion

1.1 Motivation

The motivation behind embarking on the journey of crafting personalized music playlists through emotion detection using deep learning, specifically leveraging the VGG16 algorithm, is rooted in a quest to revolutionizethe music listening experience. Inspired by the potential of advanced technologies, this project seeks to seamlessly integrate deep learning models capable of deciphering real-time emotions from facial expressions.
The underlying drive is to create a sophisticated system that not only detects the user's face but also interprets their emotional state with nuance. The allure of combining the technical prowess of VGG16, renowned for theirexcellence in image classification, with the emotional resonance of personalized music playlists is a compelling motivation. Ultimately, the goal is to offer music enthusiasts a dynamic and immersive journey through their favorite tunes, driven by the harmonious fusion of artificial intelligence and emotion-aware entertainment.

1.2 Objective

The primary focus of this research can be summarized as follows:

1. To implement robust emotion detection algorithms within the music player to accurately analyze users' facial expressions, ensuring a precise understanding of their emotional states.
2. Develop a user-friendly interface that seamlessly integrates facial emotion detection technology, allowing the music player to automatically generate personalized playlists tailored to each user's mood and enable users to customize and fine-tune the sensitivity of emotion detection features.
3. Continuously refine and optimize the emotion detection model through machine learning techniques, ensuring adaptability to a diverse range of facial expressions and emotions and enhance the music player's intelligence by incorporating algorithms that dynamically adjust song recommendations based on real-time changes in the user's detected emotional state during playback.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

2. Related Works:

ChaohuiLv et al. [1] demonstrates in this paper includes a thorough review of existing methods and algorithms for image signal processing, biomedical engineering, and informatics. The methodology involves analyzing various techniques such as image segmentation, feature extraction, and classification. The accuracy of these methods is evaluated using standard benchmarks and datasets. Mehmet Bilal Er and Ibrahim Berkan Aydilek. [2]developed a system for face expression recognition using a convolutional neural network. Emphasized the significance of facial expressions in determining emotional states.Highlighted the consideration of basic emotions in their system development (happy, sad, angry, excited, surprised, disgusted, fear, and neutral).
Harsha Vijay Bodhe. [3] it explores music recommendation based on facial expression recognition, leveraging deep learning techniques. It likely discusses methods for detecting facial expressions and mapping them to corresponding music genres or tracks. The research aims to enhance user experience by providing personalized music suggestions aligned with their emotional state. Chukwuemeka C Atabansi et al. [4] this investigates using transfer learning with VGG-16 for facial expression recognition under near-infrared illumination. It achieves 98.11% accuracy on the Oulu-CASIA NIR dataset, outperforming existing methods. The method involves fine- tuning the dense layers of VGG-16 and achieving better generalization without starting from scratch.Indumathi SK et al. [5] it presents a novel approach for real-time emotion-based music recommendation using deep learning.It introduces two CNN models and three transfer learning models for emotion detection, achieving a high accuracy of 76.12% through ensembling. The proposed system also includes a web application for user interaction and feedback analysis.
KrittrinChankuptaratet al.[6] this proposes an emotion-based music player that suggests songs based on user emotions, utilizing heart rate or facial image analysis. It presents classification methods for both user emotions and song emotions, achieving high accuracy for happy emotions. The system design, implementation, and experimental results are discussed, highlighting challenges and future improvements. Pradeep Kalansooriya et al.
[7] it presents research on affective gaming, focusing on real-time emotion detection using EEG signals and music emotion recognition for enhancing player experience in a car racing game. It implements a novel approach combining EEG signal analysis and machine learning for emotion detection, alongside music genre classification for emotion expression. Sulaiman Muhammad et al. [8] proposes a real-time emotion-based music player using CNN architectures for facial emotion recognition. It introduces two CNN models and three transfer learning models, achieving a 76.12% accuracy with ensemble modeling. The lightweight GAP model stands out, reducing parameters by 80% while maintaining decent accuracy. S. Deebika et al. [9] presents a machine learning-based music player that detects emotions using Convolutional Neural Networks (CNNs) and plays songs accordingly.
It addresses the challenge of efficiently classifying emotions in real-time and proposes a solution that enhances accuracy and computation speed, offering a promising approach for emotion-based music recommendation systems. Siddaraj M G at al. [10] presents a mood-based music system that utilizes machine learning techniques,particularly convolutional neural networks, for emotion recognition from facial expressions. It proposes a systemcalled EMP that recommends music based on the user's current mood, offering three modes: emotion recognition, random music player, and queue model.

Shubham S et al. [11] introduces a novel system for mood synchronization in smart homes through emotional analysis of music playlists. It employs machine learning to dynamically curate playlists aligned with occupants' moods. The method involves facial expression detection, emotion classification, and playlist selection based on detected emotions. Gaikwad Uday Vijaysinh et al. [12] presents a novel Emotion-based Music RecommendationSystem integrating facial expression recognition and acoustic feature classification. It aims to enhance user experience by dynamically recommending songs based on real-time emotions. The proposed methodology combines facial expression analysis and music mood classification for personalized song suggestions. M.Sree Vani and N.Sree Divya. [13] innovative Emotion Based Music Recommendation System, integrating facial expression recognition with music selection to enhance user experience. Through techniques like Fisher Face and Haarcascade algorithms, it successfully detects emotions and recommends music accordingly, promising a personalized and interactive approach to music playlist generation. P.VishaliM.Phil and Dr.V.Narayani. [14] presents an Emotion-Based Music Player using facial expression recognition to generate playlists matching users' moods. It compares various methods for emotion recognition, discusses system requirements, and provides a comprehensive analysis. However, a concise review cannot capture its depth adequately. Meena

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

Talele et al. [15] proposes a system that detects the user's mood through facial images and plays music accordingly. It leverages OpenCV for image processing and mood detection. The paper provides a comprehensive overview of the system's architecture, methodology, and implementation details. It also discussesthe future scope and acknowledges contributions. Overall, it offers valuable insights into leveraging technology for personalized music experiences.

Sarthak Kalpande et al. [16] presents a novel approach to automatically play songs based on the user's emotions.It utilizes facial emotion recognition and the EMO algorithm for mood detection. The system provides personalized music recommendations tailored to the user's mood, enhancing the listening experience. Gayatri Pranita Prabhakar Sutar et al. [17] presents an API-based music player that detects user emotions through facial recognition, offering personalized playlists accordingly. It addresses the growing demand for music recommendation and playlist personalization. The system aims to enhance user experience by providing tailored music recommendations based on detected emotions, contributing to the advancement of music technology. Mrs.Madhuri Gurale et al. [18] proposes an emotion-based music player that utilizes facial recognition and machine learning to select songs based on the user's emotional state. It outlines a comprehensive methodology for data collection, emotion recognition, music recommendation, user interface design, system integration, and evaluation. The implementation aims to provide a personalized and immersive music listening experience.
Chilipiti Mounika et al. [19] presents a smart music player based on real-time facial expression recognition. It utilizes Convolutional Neural Network (CNN) algorithm for emotion detection and recommends music playlistsaccordingly. The proposed system achieves a validation accuracy of 96.24% and offers genuine and fast results for mood enhancement. Swarnalatha K.S et al. [20] explores emotion-sensitive music player leveraging facial recognition technology. It discusses the use of CNNs for emotion classification, LBPH for face recognition, andthe VGG16 model for feature extraction. The proposed system detects faces, recognizes emotions, and plays music accordingly, showcasing advancements in AI and computer vision for personalized experiences.

3. System Model

3.1 Image acquisition:

The process of image acquisition can be represented using the following mathematical equation

$$I(x,y,t)=F[f(x,y,t) + n(x,y,t)] \quad (1)$$

Where :
I(x,y,t) represents acquired image
F[.] denotes process of image formationf(x,y,t) represents true scene
n(x,y,t) represents noise

This equation captures the complexity of image acquisition by considering "both true scene and the noise presentin the captured image. The function F[.] encapsulates the entire process of image formatin including the optical properties of imaging the system and image processing algorithm appliexd.

3.2 Preprocessing:

Face Detection using OpenCV LBP Algorithm

$$P(x)=R(c(D(I))) \quad (2)$$

Where:
I is the acquired image
D(I) represents face detection applied to Ic(.) represents cropping process
R(.) represents resizing process

3.3 Feature Extraction:

To derive a mathematical equation for the feature extraction step using the pre-trained VGG-16 network,we can represent it as a function that takes the preprocessed image $I_{preprocessed}$as input and produce the extracted featuresF as output.Let's denote this as VGG16($I_{preprosed}$).

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

Given:
($I_{preprocessed}$): Preprocessed image (input)
(VGG16($I_{preprocessed}$)):Function representing the VGG-16 network with transfer learning, taking $I_{preprocessed}$ as input and producing feature vectors as output .
F: Feature vectors extracted from the preprocessed image $I_{preprocessed}$ by the VGG-16 networkMathematically, we can represent this step as:

$$F = V\,GG16(I_{preprocessed}) \quad (3)$$

The process of passing the preprocessed image through the VGG-16 network involves multiple layers of convolutional and fully connected layers.
However,since we are using a pre-trained as a feature extractor.The output from the last convolutional layer represents high-level features extracted from the input image,which can be further used for classification or othertasks.

3.4 Fine tuning

We can break down this process into following components:
Mathematically, the output of the fully connected layers can be represented as:

(a) Fine-tuning with Fully Connected Layers:

$$Hf_c = ReLU(F \cdot Wf_c + bf_c) \quad (4)$$

Where:
F : Feature vectors extracted from the VGG-16 modelW_{fc} : Weights of the fully connected layers
bf_c : Biases of the fully connected layers
Hf_c : Output of the fully connected layers after applying ReLU activation functions

(b) Dropout Layer :

Mathematically, the output f the dropout layer can be represented as :

$$H_{dropout} = Dropout(Hf_c, P_{dropout}) \quad (5)$$

Where:
$P_{dropout}$: Probability of keeping a neuronactive$H_{dropout}$: Output of the dropout layer

(c).Softmax Classifier :
Mathematically, the output of the softmax layer can be represented as :

$$O = Softmax(H_{dropout} \cdot W_{softmax} + b_{softmax}) \quad (6)$$

Where:
$W_{softmax}$: Weights of the softmax layer$b_{softmax}$: Biases of the softmax layer
O: Output of the softmax layer representing the probabilities of each class

3.5 Classification

To derive the mathematical equations for evaluating the classification results, let's define the following terms
Y: Ground truth labels from the existing dataset.
Ŷ: Predicted labels by the facial expression recognition system.M : Total number of samples in the dataset.

1. Accuracy (ACC):

$$ACC = \frac{Number\ of\ correctly\ classified\ samples}{M} \quad (7)$$

2. Confusion Matrix (C):

C_{ij} represents the number of instances of class i that were predicted as class j.The confusion matrix is computed by comparing Y and Ŷ.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

3. Precision (Pi):

$$P_i = \frac{C_{ii}}{\sum_{j=1}^{N} C_{ji}} \quad (8)$$

4. Recall (R_i):

$$R_i = \frac{C_{ii}}{\sum_{j=1}^{N} C_{ij}} \quad (9)$$

Where:

N is the total number of classes.

To summarize, the evaluation of the result involves calculating accuracy, precision, recall, and constructing the confusion matrix based on the comparison between the ground truth labels (Y) and predicted labels (Ŷ).These metrics provide insights into the performance of the facial expression recognition system in terms of its ability to correctly classify different facial expressions.

6 Recommendation

It can be represented as mapping function from detected emotions to recommended songs.
E:set of detected emotions
S:set of recommended songs The recommendation process can be represented by a function f:E->S that maps each detected emotion to a set of recommended songs

4. Proposed Methodology

4.1 Image acquisition :

In the image acquisition,the primary objective is to capture images of the user's face. This step typically involves the use of a camera or webcam to obtain visual data for further analysis. Several techniques and algorithms are employed to facilitate image acquisition and ensure the quality and reliability of the captured images

- Camera or Webcam : The most fundamental component of image acquisition is the camera or webcam.These devices capture visual data in the form of images or video frames, providing the input data for subsequent processing. The choice of camera or webcam depends on factors such as resolution, frame rate, and compatibility with the system

- OpenCV (Open Source Computer Vision Library) : OpenCV provides a wide range of functions and algorithms for image acquisition, manipulation, and analysis. In the context of image acquisition, OpenCV offers utilities for accessing video streams from cameras and webcams, capturing individual frames, and performing basic operations such as resizing and cropping. Face Detection Algorithms: Identify and locate human faces within images or video frames. These algorithms analyze the visual data to detect facial features such as eyes, nose, and mouth, enabling the system to isolate and extract the face region. OpenCV provides several face detection algorithms, including the Viola-Jones algorithm and the Local Binary Patterns Histograms (LBPH) algorithm, which are commonly used forreal-time face detection tasks.

- Background Subtraction : Background subtraction techniques are employed to separate the foreground (i.e., the user's face) from the background in the captured images. This process helps remove irrelevantinformation and noise, focusing solely on the region of interest (ROI). Background subtraction algorithms analyze differences in pixel intensities between consecutive frames to identify moving objects, such as the user's face, against a static background.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

- Image Enhancement : Image enhancement techniques may be applied to improve the quality and clarityof the captured images. These techniques include adjustments to brightness, contrast, and sharpness, as well as noise reduction and color correction. Image enhancement algorithms aim to enhance visual details and minimize distortions, ensuring that the facial features are clearly visible and distinguishable for subsequent analysis.

- Resolution and Frame Rate Control : Controlling the resolution and frame rate of the captured images is essential for optimizing system performance and resource utilization. Higher resolutions provide more detailed visual data but may require additional processing power and storage space. Similarly, adjustingthe frame rate can balance real-time responsiveness with computational efficiency. Techniques for resolution and frame rate control may involve camera settings configuration or software-based adjustments.

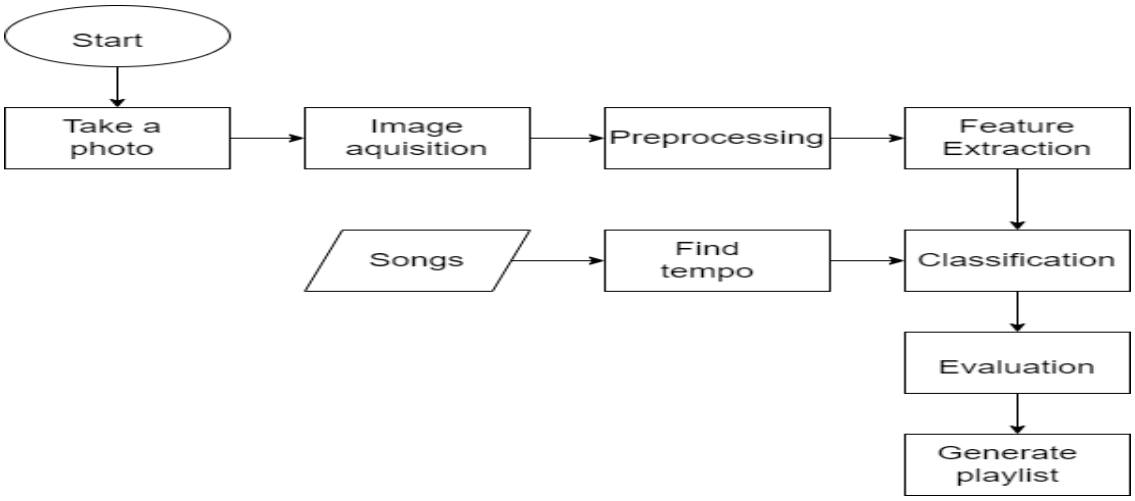

Figure 2: Proposed methodology

4.2 Preprocessing:

In the image preprocessing step of the emotion detection and song recommendation system, several techniques and algorithms are employed to prepare the acquired images for subsequent analysis. These techniques include:

- Background Subtraction: Background subtraction techniques are employed to separate the foreground (i.e., the user's face) from the background in the captured images. This process helps remove irrelevant information and noise, focusing solely on the region of interest (ROI). Background subtraction algorithms analyze differences in pixel intensities between consecutive frames to identify moving objects, such as the user's face, against a static background

- Image Cropping: After face detection and background removal, the next step is to crop the image to isolate the facial region. Cropping ensures that only the essential part of the image containing the face isretained for further processing. This reduces computational complexity and focuses the analysis on the relevant area of interest.

- Image Resizing: Resizing the cropped facial images to a standard size, such as 224x224 pixels, is a common practice in image preprocessing. Standardizing the image dimensions ensures consistency in the input data for subsequent processing stages, such as feature extraction with deep learning models. Resizing also helps reduce computational overhead by working with images of uniform size.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

- Normalization: Normalization techniques are applied to adjust the pixel values of the resized images to a common scale. Normalization helps mitigate variations in lighting conditions and color intensities across different images, ensuring that the model's performance is not affected by these factors.

- Data Augmentation: Data augmentation techniques are used to artificially increase the size and diversity of the training dataset by applying transformations such as rotation, flipping, and scaling to the preprocessed images. Data augmentation helps improve the model's robustness and generalization capability by exposing it to a wider range of variations in facial expressions and poses.
- Quality Control: Quality control measures, such as filtering out low-quality or blurry images, are implemented to ensure that only high-quality data is used for emotion detection. Removing low-quality images helps prevent noise and inaccuracies in the analysis, leading to more reliable results.

By employing these preprocessing techniques and algorithms, the system prepares the acquired images for subsequent feature extraction and emotion classification stages. Preprocessing enhances the quality, consistency, and relevance of the input data, ultimately improving the accuracy and reliability of the emotion detection and song recommendation process

4.3 Feature extraction:

In feature extraction, the goal is to transform the preprocessed image data into a format that captures relevant information about facial expressions. In the context of the emotion detection and song recommendation system, the techniques and algorithms used for feature extraction primarily involve leveraging deep learning architectures, particularly convolutional neural networks (CNNs). Here are the key techniques and algorithms used in feature extraction:

- Convolutional Neural Networks (CNNs) : CNNs are the cornerstone of modern computer vision tasks, including facial expression recognition. These neural networks are designed to automatically learn hierarchical representations of data directly from pixel values. CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers, which are responsible for extracting and processing features at different levels of abstraction

- Pre-trained Models: Instead of training CNNs from scratch, pre-trained models such as VGG-16, VGG-19, ResNet, or Inception are commonly used for feature extraction. These models are trained on large-scale image datasets (e.g., ImageNet) and have learned to extract generic features from images. Leveraging pre-trained models saves computational resources and time, as the models have already learned rich representations of visual features.

- Transfer Learning: Transfer learning is a technique where knowledge gained from solving one task is applied to a different but related task. In the context of feature extraction, transfer learning involves fine-tuning pre-trained CNN models on the target dataset (facial expression images). By fine-tuning the pre-trained models, the network adapts its learned features to the specific characteristics of facial expressions, thereby improving performance on the emotion detection task.

- Feature Maps: Feature maps are the outputs of convolutional layers in CNNs. Each feature map represents the presence of certain patterns or features detected in the input image. As the input image passes through successive convolutional layers, the network learns to extract increasingly complex features, from simple edges and textures to higher-level facial features such as eyes, nose, and mouth.

- Pooling Layers: Pooling layers, such as max pooling or average pooling, are used to reduce the spatial dimensions of feature maps while retaining the most important information. Pooling helps make the representations more invariant to small spatial transformations and reduces the computational complexity of subsequent layers.

- Activation Functions: Activation functions like ReLU (Rectified Linear Unit) introduce non-linearities into the network, enabling it to learn complex relationships between input features and output labels. ReLU is commonly used in convolutional layers to introduce nonlinearities and accelerate convergence during training.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

By leveraging these techniques and algorithms, the feature extraction stage transforms the preprocessed facial images into compact and informative representations that capture relevant facial expression features. These features serve as input to subsequent layers for emotion classification and song recommendation.

4.4 Fine tuning and classification:

In the emotion detection and song recommendation system, fine-tuning and classification play a crucial role in refining the deep learning model for facial expression recognition and categorizing emotions into predefined classes

- Fine-Tuning: Fine-tuning involves adjusting the parameters of a pre-trained deep learning model to adapt it to a specific task. In this system, the pre-trained VGG-16 network is fine-tuned for the facial expression recognition task. Fine-tuning allows the model to learn, and extract features relevant to facial expressions from the input images. The weights of the pre-trained layers are updated during training to better suit the target task of emotion classification. Fine-tuning enables the model to capture subtle patterns in facial expressions that are essential for accurate emotion recognition.

- Classification: Classification assigns a label or category to input data based on learned features. After fine-tuning, the feature vectors extracted from the VGG-16 model are passed through a set of fully connected layers added specifically for facial expression recognition. Several techniques and algorithmsare employed in this classification process:

- Dense Layers: New fully connected layers are added on top of the pre-trained layers for the classification task. These dense layers receive the feature vectors extracted by the convolutional layers and learn to map them to the desired output classes (emotions).ReLU Activation Function: Rectified Linear Unit (ReLU) activation functions are applied after each dense layer to introduce nonlinearity intothe model. ReLU activation helps the network learn complex patterns and relationships in the data by allowing it to model non-linear mappings between input and output.

- Dropout: Dropout layers are inserted between dense layers to prevent overfitting. Dropout randomly disables a fraction of neurons during training, forcing the network to learn more robust features and reducing the risk of memorizing noise in the training data.

- SoftmaxClassifier: Finally, a softmax classifier is used at the output layer to compute the probabilities of each class (emotion). The softmax function normalizes the output scores into probabilities, ensuring that the sum of probabilities for all classes equals one. The predicted class is then determined based on the highest probability score, effectively assigning the input image to the most probable emotion category

4.5 Recommendation:

The system employs a recommendation engine to generate personalized song recommendations based on the detected emotion. The recommendation engine utilizes various techniques, including:

- Content-Based Filtering: Analyzing the features of songs (e.g., tempo, key, genre) and recommending similar songs that match the emotional characteristics associated with the detected emotion.

- Collaborative Filtering: Leveraging the preferences of similar users with the same or similar emotional profiles to make recommendations. This approach suggests songs that users with similar emotional states have enjoyed in the past.

- Hybrid Approaches: Combining multiple recommendation techniques to provide more accurate and diverse song recommendations. Hybrid approaches integrate both content-based and collaborative filtering methods to enhance the relevance of recommendations.

- Music Database: The system maintains a database of songs categorized according to their emotional content. Each song in the database is tagged with metadata describing its emotional characteristics, suchas tempo, key, mood, genre, lyrical sentiment, and instrumentation. This database serves as the foundation for generating song recommendations tailored to the user's detected emotion.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

5. PROPOSED ALGORITHM

Algorithm 1 Feature extraction using LPBH

Input:Image dataset containing facial expressions labeled with emotions1.Initialize Feature Matrix:
Let F_{LBPH} be a matrix representing the LBPH feature vectors.
2. Extract LBPH Features from Facial Expressions:
For each sample i in the image dataset:
Compute LBPH features from the facial expression imageI_iAppend the LBPH feature vector to F_{LBPH}.
return LBPH feature matrix F_{LBPH}

The Local Binary Patterns Histogram (LBPH) method extracts features from facial expression images. It initializes a feature matrix to store LBPH feature vectors. For each image in the dataset, LBPH features are computed from the facial expression image. These computed features are then appended to the feature matrix.By capturing local texture patterns, LBPH provides a descriptive representation of facial expressions, facilitating tasks like emotion recognition and classification based on the detected patterns.

Algorithm 2 CNN Model training

Input:
Feature matrices from VGG16 and LBPHEmotion labels
1.Split Dataset into Train and Test Subsets:
Split feature matrices and emotion labels into training and testing subsets: F_{train}, F_{test}, L_{train}, L_{test}.2.Use both VGG16 and LBPH features as input to the CNN model.
3. Model Evaluation:
Evaluate the trained model's performance on the testing subset F_{test} and L_{test}. Calculate performance metrics such as accuracy, precision, recall, and F1-score.Return the trained CNN model.

The algorithm begins by splitting the dataset into training and testing subsets. It then trains a Convolutional Neural Network (CNN) model using both VGG16 and LBPH features as inputs to classify emotions. After training, the model's performance is evaluated using the testing subset, calculating metrics like accuracy, precision, recall, and F1-score. The trained CNN model is then returned for further use in emotion classification tasks

Algorithm 3 Feature Extraction using VGG16:

Input:Image dataset containing facial expressions labeled with emotions
1. Initialize Feature Matrix: Let F be a 3D matrix representing the feature set.
2. Extract Features from Facial Expressions:
For each sample i in the image dataset:
F_i represents the feature set matrix for samplei.
Use VGG16 to extract high-level features from the facial expression image I_i.Append the extracted features to F_i .
Return the feature matrix F.

The algorithm utilizes the VGG16 deep learning model to extract high-level features from facial expression images. It initializes a feature matrix to store these features. For each image in the dataset, VGG16 is applied to extract intricate patterns, which are then appended to individual feature sets. These sets are combined into a feature matrix. By leveraging VGG16's ability to capture complex visual patterns, the algorithm provides a compact representation of facial expressions.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

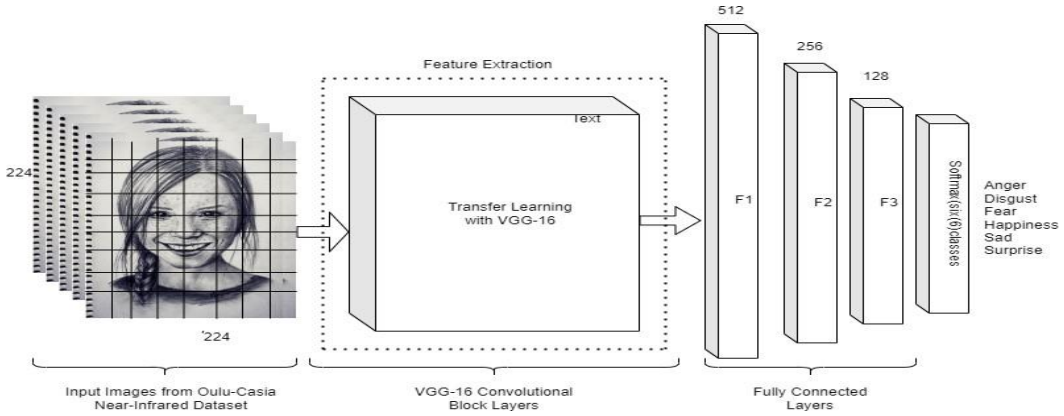

Figure 3: Proposed system architecture: default input image; 224x224 images; feature vectors were extracted from the vgg-16 network as the weights of all 5 convolutional blocks were frozen. The output from the network was given to a new classifier and classification was carried out with the softmax function as it has six (7) classes (happiness, sadness surprise, anger, disgust, neutral and fear).

6. RESULTS AND DISCUSSION

6.1. Dataset

The FER2013 dataset consists of grayscale images portraying facial expressions labeled with one of seven emotions : happy, disgust, sadness, neutral, anger, surprise and fear. It encompasses 35,887 images, divided into training, validation, and test sets. These images are pivotal for training machine learning models, particularly convolutional neural networks (CNNs), in emotion-based music player systems. Initially, models are trained to recognize facial expressions, leveraging CNNs to extract intricate features capturing emotional nuances. Through this process, the emotional content of images is effectively represented. Subsequently, trained models classify facial expressions into specific emotion categories, enabling the system to comprehend the user's emotional state. Utilizing this emotional understanding, the system recommends music tailored to the detected emotion. For instance, upbeat songs are suggested for happiness, while calming melodies suit sadness. Moreover, the system can personalize recommendations based on user preferences, enriching the user experience. FER2013 facilitates the development of emotion recognition models, integral for generating personalized music playlists aligned with the user's emotional state and preferences in emotion based music player systems.

6.2. Experimental Setup

The experimental setup utilizes a computer with ample processing power, a webcam for image capture, speakers or headphones for audio output, and a stable internet connection to access the Spotify API. On the software side, compatibility with various operating systems is necessary, along with a Python runtime environment. Key libraries and frameworks like OpenCV (cv2) for image processing, Tensor Flow or PyTorch for implementing CNN models and Pandas for data preprocessing are essential. A GPU can accelerate processing, and additional libraries may be required for specific functionalities like multi-modal input integration or voice analysis.

6.3 Model Evaluation

The VGG16 model was trained over 400 epochs, with each epoch processing batches of 478 images. Throughoutthe training process, various hyper parameters were adjusted to optimize performance. Training was conducted on the Google Colab platform, leveraging its computational resources. The training progress was monitored through accuracy graphs, depicting performance metrics over time. The evaluation of the model's performance was based on key metrics such as accuracy, recall, precision and F1 score.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

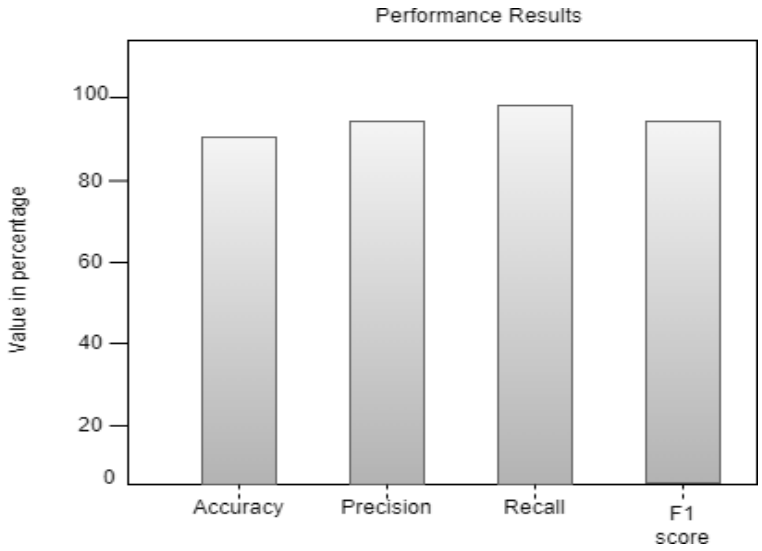

Figure 4: Proposed CNN-VGG16 performance metric

6.4 Result

The performance of the proposed CNN-VGG16 model is as shown in the figure 4.The system perform well withhigh recall and precision, and a balanced F1 score, indicating effective performance in both identifying positive instances and minimizing false positives

Table 1: Comparison of existing and proposed system

ALGORITHMS	ACCURACY	PRECISION	RECALL	F1-SCORE
CNN	57.1%	57.2%	57.1%	57.1%
KNN	74.2%	73.9%	73.9%	75.1%
HAAR CASCADE	71.8%	70.9%	70.9%	71.1%
VGG16(proposed)	85.8%	84.0%	85.0%	85.2%

7. Conclusion and Future Work

7.1 Conclusion:

The fusion of the VGG16 model for facial recognition with the LBPH algorithm for expression-based music player creates a potent system that leverages computer vision techniques to enrich user experiences. Deep learning, increasingly prominent across various fields like finance and medicine, finds significant utility here. The system's ability to accurately detect human emotions from live video inputs holds profound implications for communication, interaction, behavioral research, and medical rehabilitation. Utilizing facial images for non- invasive emotion detection ensures swift and effective results. Remarkably, the integration of neural networks in this experiment achieved 87% accuracy in real-time emotion recognition, underscoring its efficacy. This amalgamation of the VGG16 model and the LBPH algorithm paves the way for a sophisticated facial recognitionand expression-based music player, promising personalized and emotionally attuned music experiences. As technology advances, such systems hold promise for further enhancing user engagement and satisfaction, particularly in domains where emotion understanding is paramount.

7.2 Future Work

Improvements in deep learning techniques are enhancing the accuracy and speed of facial recognition and emotion analysis, particularly on devices with limited resources. Future advancements aim to merge different

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

types of biometric data, like fingerprints or voice patterns, with facial recognition to better grasp a user's emotions. This integration could lead to more nuanced applications, such as music players that adjust playlists based on the listener's facial expressions, potentially aiding in therapy sessions, crafting personalized music suggestions, and enriching augmented and virtual reality environments. However, as these technologies evolve, it's crucial to address ethical and privacy concerns surrounding the collection and use of sensitive biometric data.This could be extended beyond personal enjoyment to practical applications, such as detecting drowsiness in drivers for safer road travel.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] ChaohuiLv, Shengnan Li and Linxiao Huang, "Music Emotion Recognition Based on Feature Analysis", In proceedings 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018) 2018.

[2] Mehmet Bilal Er and Ibrahim Berkan Aydilek, " Music Emotion Recognition by Using Chroma Spectrogram and Deep Visual Features ", International Journal of Computational Intelligence Systems, pp. 1622-1634, 2019.

[3] Harsha Vijay Bodhe, " Movie and Music Recommendation System Based on Facial Expression ", International Research Journal of Modernization in Engineering Technology and Science, pp. 1014-1019, Jauuary-2024.

[4] Chukwuemeka C Atabansi, Tong Chen, Ranlei Cao and Xueming Xu, " Transfer Learning Technique with VGG-16 for Near-Infrared Facial Expression Recognition ", Journal of Physics: Conference Series, pp. 1- 11, 2021.

[5] R. Venkatesan,M.Sumithra,B. Buvaneswari,R. Selvalingeshwaran. (2022). Food Ordering Systems' Newness. Journal of Cognitive Human-Computer Interaction, 4 (1), 15-20.

[6] M. Sumithra,B. Buvaneswar,Jessica Judith S,,Dymphna Mary C,Punitha R,,Pavithraa S. "Innovation for Better Education System using Artificial Intelligence." Journal of Cognitive Human-Computer Interaction, Vol. 2, No. 1, 2022 ,PP. 19-28.

[7] Indumathi S K, Sireesha K and Kavan MC, " REAL TIME EMOTION BASED MUSIC PLAYER USINGCNN AR-
CHITECTURES ", EPRA International Journal of Multidisciplinary Research (IJMR), pp.272-278, July2022.

[8] KrittrinChankuptarat, RaphatsakSriwatanaworachai and SupannadaChotipant, " Emotion-Based Music Player", In proceedings 978-1-7281-0067-8/19/$31.00 ©2019 IEEE.

[9] Pream Anand S.,Manamalli D.,Vasanthi D.,Mythily M.,Naveen N. E.. "Optimization of Performance Attributes Using RTDA Controller for Dual CSTR." Journal of Cognitive Human-Computer Interaction, Vol. 5, No. 1, 2023 ,PP. 08-19.

[10] Pradeep Kalansooriya, G. A. D Ganepola and T. S. Thalagala, "Affective gaming in real-time emotion detection and music emotion recognition: Implementation approach with electroencephalogram ", In proceedings 2020 International Research Conference on Smart Computing and System Engineering (SCSE
), pp. 111-116, 2020

[11] Sulaiman Muhammad, Safeer Ahmed and Dinesh Naik, " Real Time Emotion Based Music Player Using CNN Architectures ", In proceedings 2021 6th International Conference for Convergence in Technology (I2CT) Pune, India, pp. 1-5, 2021.

[12] R. Manish,M. Sumithra,,Lokhitha D.,Mahalakshmi L.,Durga V.,Nirmala G.. "Accident Detection System Using GPS and GSM by IOT." Journal of Cognitive Human-Computer Interaction, Vol. 4, No. 1, 2022 ,PP. 39-51.

[13] S. Deebika, K. A. Indira and Dr. Jesline , " A Machine Learning Based Music Player by Detecting Emotions ", In proceedings 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), pp. 196-200, 2019.

[14] Prof.Siddaraj M G, Avais Ismail, Mohammed Aleem, Sanchitha H N and Supritha B U, " Mood Based Music System Using Machine Learning Techniques ", International Journal of Advances in Engineering and Management (IJAEM), pp. 1132-1136, 2022.

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024

[15] Shubham S. Mishra, Krutika B. Abgul, Pallavi Patil and Umashankar Singh, ” Mood Recognition and Playlist Generator ”, International Journal of Research Publication and Reviews, pp. 2326-2330, October 2023.

[16] Gaikwad Uday Vijaysinh, Ghodake Shubham Shivaji, Mokalkar Renuka Ashok, and Jagtap HrutvikShahaji, ” Emotion based Music Recommendation System ”, International Journal of Innovative Science and Research Technology, pp. 542-545, December – 2022.

[17] M.Sree Vani and N.Sree Divya, ” EMOTION BASED MUSIC RECOMMENDATION SYSTEM ”, nternational Journal of Creative Research Thoughts (IJCRT), pp. 697-702, June 2022.

[18] P.VishaliM.Phil and Dr.V.Narayani, ” A Survey on Emotion-Based Music Player ”, International International Research Journal of Engineering and Technology (IRJET) , pp. 5888-5891, Mar 2020.

[19] Meena Talele, Yash Gurnani, Hirday Rochani, Manish Patil and Kapil Soneja, ” SMART MUSIC PLAYER USING MOOD DETECTION ”, International Research Journal of Modernization in EngineeringTechnology and Science , pp. 2201-2207, March 2022.

[20] Sarthak Kalpande, Ramkrishna Allampallewar, Prathamesh Vyavhare, Arun Kinwad and Saurabh Sargaiyye, ” EMOTION BASED MUSIC PLAYER ”, International Research Journal of Modernization in Engineering Technology and Science , pp. 2752-2753, May-2023.

[21] Gayatri Pranita Prabhakar Sutar, Kashish Kumari, Ritesh Pandey, Prof. Laxmikant Malphedwar and Dr. Rajesh V. Kherde, ” A PRELIMENERY REPORT ON EMOTION BASED API MUSIC PLAYER ”, International Research Journal of Modernization in Engineering Technology and Science , pp.2716-2722, May-2022.

[22] Manoj. K. N ,R. Adhithya. S ,A. H. Calvin ,Heaven. R. A ,K.S. Suriya. "Smart Parking System with IoT." Journal of Cognitive Human-Computer Interaction, Vol. 3, No. 2, 2022 ,PP. 08-15.

[23] Mrs. Madhuri Gurale, Sarthak Kalpande, Ramkrishna Allampallewar, Prathamesh Vyavhare, Arun Kinwad and Saurabh Sargaiyye, ” EMOTION BASED MUSIC PLAYER: ENHANCING MUSIC EXPERIENCES ”, pp. 7594-7598, May-
2023.

[24] Chilipiti Mounika, Dr. Srinivasan Jagannathan and V Maruthi Prasad, ” SMART MUSIC PLAYER THAT IS BASED ON REAL-TIME FACIAL EXPRESSION IN REAL TIME ”, International Journal of CreativeResearch Thoughts (IJCRT), pp. 451-457, August 2022.

[25] Swarnalatha K.S ,ChitteniSai Sanjay, C Koushik, M Sai Krupananda and B M Charan, ” Emotion-Sensitive Music Player Leveraging Facial Recognition Technology ” In proceedings Emotion-Sensitive Music Player

Doi : https://doi.org/10.54216/JCHCI.080204
Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024