SUGGESTING PLAYLIST AND PLAYING PREFERRED MUSIC BASED ON EMOTION FROM FACIAL EXPRESSION

BY

MST. LIMA AKTER ASHA

ID: 201-15-13842

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised by

Md. Sazzadur Ahamed

Assistant Professor

Department of CSE

Daffodil International University

Co-Supervised by

Ms. Sharun Akter Khushbu

Senior Lecturer

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

January 2024

APPROVAL

This Project titled **"Suggesting Playlist and Playing Preferred Music Based on Emotion from Facial Expression"**, submitted by **"Mst. Lima Akter Asha"**, ID No: **201-15-13842** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on January 24, 2024.

BOARD OF EXAMINERS

Dr. Md. Ismail Jabiullah(MIJ) — **Chairman**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Taslima Ferdous Shuva(TFS) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Shayla Sharmin

Sharun Akter Khushbu (SAK) — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Risala Tasin Khan (RTK) — **External Examiner**
Professor
Institute of Information Technology
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Sazzadur Ahamed, Assistant Professor, and Department of CSE** at Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

____

Mr. Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Co- Supervised by:

____

Ms. Sharun Akter Khushbu
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

____

Mst. Lima Akter Asha
Id Number: 201-15-13842
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

Firstly, I would want to thank Almighty God from the bottom of our hearts for His glorious grace, which has enabled us to successfully finish the final year project.

I would like to express my sincere gratitude and debt of gratitude to **Md. Sazzadur Ahamed**, Assistant Professor, CSE Department, Daffodil International University, Dhaka. Our supervisor has extensive knowledge and a strong interest in research. The completion of this project has been made possible by his unending patience, academic guidance, ongoing encouragement, persistent and vigorous supervision, constructive criticism, insightful counsel, and his reading of several subpar drafts and correction of them at every level.

I would like to extend my sincere appreciation to the CSE department head, Professor **Dr. Sheak Rashed Haider Noori**, as well as to the other academic members and personnel of Daffodil International University, for their kind assistance in seeing our project through to completion.

I express my gratitude to every student at Daffodil International University who participated in this discussion throughout the course of their studies.

Lastly, I must respectfully thank my parents for their unwavering patience and support.

ABSTRACT

Facial expression is a powerful indicator of human emotion and plays a crucial role in interpersonal communication. The mood of a person or his intention can be analyzed by detecting his expression. Automatic machine-based analysis of facial emotions is an essential aspect of artificial intelligence and has significant applications in various areas, including music recommendation. By analyzing facial features and expressions, a music recommendation system can predict the user's mood and recommend songs that align with their emotional state. Many researchers have worked on this. Our proposed system works on 8 moods of humans which are angry, contempt, disgust, fear, happy, neutral, sad, and surprise. This study utilizes a machine learning concept to achieve this goal. The methodology involves data collection, model training using a combination of Convolutional Neural Network (CNN) and VGG16 CNN, and recommending songs from the Spotify music dataset. The results show that both CNN and VGG16 CNN performed well in detecting facial expressions, with CNN achieving 89% accuracy and VGG16 achieving 97% accuracy. This system effectively recommends songs from the Spotify dataset based on the user's mood.

TABLE OF CONTENTS

CONTENTS **PAGE**

CHAPTER

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Introduction

In this vast domain of entertainment, music is one of the best forms of expression that touches the soul with its universal language, evoking emotions, memories, and connections, unlike any other art form. There are many types of music and songs and not every music is enjoyable in every situation. Most of the time it depends on the listener's mood and music can also change the mood of a person. Everyone loves to enjoy music based on their emotion and mood. Most of the time, they can't find the preferred song from the huge song list that they have on their device. A lot of songs are using device storage and are never played. Users have to play songs manually all the time and it's difficult to create a playlist for that. So, our project can identify the mood of the user and suggest the music based on that mood by using the machine learning concept.

Fig-1.1.1 Basic emotions of humans.

In this project, we will work with these basic emotions. These emotions will be collected from a person's facial expression from an image. Then the system will detect the expression and suggest the playlist or song.

1.2 Motivation

I was motivated to do this project to save my time while I wanted to listen to music. The music I usually enjoy depends on my mood and to find and play that music is a hassle and time-consuming. So, my proposed system will detect my mood and find the preferred music and playlist for me. So, it's a very time-saving system. With the help of this system, I can also separate and get rid of the music that I never listen to but consumes my storage.

1.3 Objective

My main objective is to make a system that can detect a person's mood by his facial expression and can suggest the music and playlist that is appropriate for his mood in that situation.

The main objectives of this research are:

- To detect the user's face.
- To identify the facial expression of the user.
- To classify the expression.
- To find out the playlist for that expression.
- To play the music based on that expression from the playlist.
- To make different playlists based on different expressions.
- To enhance and speed up the process of enjoying music.

These all are the main objectives I wish to fulfill.

More specifically there are some technical objectives which are as follows:

1. **Facial Expression Recognition:** Develop a system capable of accurately recognizing and categorizing human facial expressions into eight different emotions (angry, contempt, disgust, fear, happy, neutral, sad, surprise) using machine learning algorithms.
2. **Data Collection and Preprocessing:** Gather a comprehensive dataset of facial images representing diverse expressions and emotions to train the proposed model effectively. Ensure the data is properly labeled and preprocessed to enhance model performance.
3. **Model Development:** Implement and train a combined Convolutional Neural Network (CNN) and VGG16 CNN model to effectively detect and classify facial expressions. Fine-tune the model's architecture to optimize accuracy in emotion recognition.

4. **Music Recommendation System Integration:** Integrate the facial expression recognition model with a music recommendation system. Develop algorithms that correlate detected emotions with specific music genres or songs from the Spotify dataset, aligning them with the user's emotional state.
5. **Evaluation and Validation:** Evaluate the performance of the developed model using appropriate metrics such as accuracy, precision, recall, and F1-score. Validate the model's effectiveness in detecting emotions and recommending music aligned with the user's mood.

1.4 Research Questions

Some potential research questions for my research project are:

1. How effective are Convolutional Neural Network (CNN) and VGG16 CNN models in accurately detecting the eight different human emotions from facial expressions?
2. What are the key facial features and patterns utilized by the machine learning models (CNN and VGG16 CNN) to distinguish between different emotions such as anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise?
3. How does the integration of machine learning techniques with facial expression analysis contribute to the accuracy and reliability of predicting human emotions?
4. What is the comparative performance of CNN and VGG16 CNN models in recommending songs from the Spotify dataset based on the detected mood of the user?
5. How do users perceive the accuracy and relevance of song recommendations made by the proposed system according to their mood?
6. To what extent can the proposed system generalize across different demographics or cultural backgrounds in accurately predicting and recommending music based on detected emotions?
7. What are the potential ethical considerations or biases in using facial expression analysis and machine learning algorithms for predicting human emotions and recommending content such as music?

These questions serve as a starting point for my research, allowing me to delve deeper into various aspects of my proposed system and its implications.

1.5 Expected Outcome

Those are the outcomes we will get at the end of this project.

1. **Accurate Facial Expression Detection:** Expect the developed model to accurately detect and classify facial expressions, achieving high accuracy in recognizing the eight defined emotions. Anticipate achieving high accuracy in emotion recognition.

2. **Model Performance Comparison:** Compare the performance of the individual CNN and VGG16 CNN models in detecting facial expressions. Expect the VGG16 CNN to show superior performance due to its deeper architecture and learned feature representations.

3. **Effective Music Recommendation:** Anticipate the system recommending music that aligns with the user's detected emotional state. Users should experience songs that reflect their moods accurately, enhancing their music-listening experience.

4. **System Robustness and User Satisfaction:** Assess the robustness of the integrated system under various conditions and its ability to consistently recommend suitable music based on detected emotions. Expect positive user feedback regarding the accuracy and relevance of the music recommendations.

5. **Contribution to the Field:** Contribute to the existing research by providing a comprehensive methodology that effectively combines facial expression recognition and music recommendation systems based on user emotions, potentially opening avenues for emotion-aware personalized services.

I expect these results at the end of this project.

1.6 Layout of the Report

Chapter 1: The general structure of the study is presented in Chapter 1, which also contains the research's rationale, goal, questions, and anticipated results.

Chapter 2: This part discusses the difficulties of the research as well as related publications, a summary of the study, and a comparative analysis.

Chapter 3: The research's theoretical topics are illustrated in Chapter 3. It first outlines the whole research project's workflow before talking about the process of obtaining and preparing the data. The necessary algorithms for the suggested model are then displayed, along with an example of how to correctly execute them.

Chapter 4: Chapter 4 presents the results of implementation and assesses the suggested model's effectiveness. The project's correct explanation and clarification are provided in this part.

Chapter 5: The study project is concluded in Chapter 5 with a summary of future directions. This section summarizes the project's overall efforts and results and offers some suggestions for further research in this area.

CHAPTER 2
BACKGROUND STUDY

2.1 Introduction

When it comes to entertainment, music is a powerful medium that can speak to people of all languages and resonate deeply within them, evoking feelings, memories, and relationships in a manner that other art forms just cannot match. Not every song, despite the wide variety of musical styles and compositions, is appropriate for every situation or emotion. The power of music to mold and adjust the listener's emotional environment is what makes it so beautiful.

Understanding the close connection between feelings and taste in music, our initiative aims to improve and streamline the music-listening experience. Users frequently find it difficult to browse through the enormous song collections that are saved on their devices, which results in wasted storage space and lost opportunities to enjoy music.

These days, there's a growing curiosity with using facial expressions as a key component when suggesting music. This growing interest stems from the realization that facial expressions are powerful markers that provide deep insights into a person's emotional domain—a personal dimension closely linked to their musical preferences. Many research works have attempted to interpret the complex relationship between facial expressions and the emotional content expressed by music.

An interesting story is revealed by investigating this connection between musical feelings and facial expressions. Scholars have enthusiastically delved into deciphering the complex relationships between the subtleties of face expressions and the multitude of emotions captured in musical compositions. This investigation aims to uncover the subliminal messages concealed in facial expressions that reflect and resonate with the emotional fabric of different musical genres and melodies.
The deep insights gained from these experiments highlight the potential power of facial expressions as a

means of predicting and comprehending a person's musical tastes. Through deciphering the complex language of emotions imprinted on a person's face, these projects aim to open a door to a more intuitive, emotionally relevant, and customized music recommendation system. This intriguing trend for improving how we interact with and find music is presented by the convergence of facial expression analysis and musical emotions, which promises a more personalized and rich aural experience that is in perfect harmony with our emotional states.

2.2 Related Works

In recent years, there has been a growing interest in the use of facial expressions as a tool for music recommendation. This interest stems from the recognition that facial expressions can provide valuable insights into a person's emotional state, which is closely linked to their music preferences. Several studies have explored the relation-ship between facial expressions and musical emotions.

By extracting data from the facial expressions using a camera, this input can be utilized to generate playlists that align with the detected mood. This eliminates man-ual sorting of songs and allows for the creation of appropriate playlists based on emotional features. H. Immanuel James et al. [1] proposed a system that focuses on detecting emotions in order to develop an emotion-based music player. It aims to improve upon existing approaches by utilizing facial recognition algorithms for emo-tion detection.

Ziyang Yu et al. [2] worked with microexpression for this detection. They detected the microexpression, extracted the features, and classified the category. They used Convolutional Neural Network for the detection of their suggested automated sys-tem.

MetFlorence et al. [3] explored the potential benefits of incorporating facial ex-pression recognition in music recommendation systems. Through their study, they found that incorporating facial expression recognition improved the accuracy of music recommendations compared to traditional methods. The system uses webcam technology to capture the user's image and determine their current emotion. Based on this information, it suggests an appropriate song from the user's playlist that aligns with their emotional state, eliminating the need for manual searching or browsing.

Jingye Zhang et al. [4] wrote a research paper that propose a system that recom-mends songs and movies using a deep learning method. They used ResNet-38 for the detection and their proposed system was convenient and applicable.

serves to introduce a novel mode of Recommendation system to counter these drawbacks. In this paper, a convenient, accurate, and widely applicable recommen-dation system for films and pop songs, using deep learning as an efficacious tool, is propounded. They devised a refined Deep Residual Network (ResNet-38) for the emotion detection of the users, which outperforms other traditional models.

ShanthaShalini K et al. [5] aims to develop a dynamic music recommendation system based on human emotions. It uses feature extraction and machine learning techniques to detect emotions from real faces. The system then plays songs for spe-cific moods, providing a personal touch to users. The goal is to identify human feel-ings and create an emotion-based music player using computer vision and machine learning techniques.

Sarika Vidyasagar et al. [6] suggested a technique, based on the extraction of fa-cial expression, that creates a playlist, which reduces the time and labor necessary to manually render the process. The system will also suggest songs based on their lyrics. This makes it possible for the recommended system to reduce the amount of time that must be spent calculating in order to produce the desired results and the overall cost of the planned system, which raises the system's overall accuracy. Facial expres-sions are captured using an integrated camera. computes the precision of systems for identifying emotions in real-time photos. For the overflowing concerns to be handled and user-specific data to be suggested, including a cooperative filter, content-based suggestion, and utilizing Sentiment-based music and distributing the burden between mobile, low-end devices, and high-end servers

Ankita Mahadik et al. [7] proposed a system, where a mood-based music player is created that performs real-time mood detection and suggests songs as per detected mood. This turns into an extra feature for the conventional music player apps that are already pre-installed on our phones. One key advantage of using mood detection is customer contentment. This apparatus aims to Analyze the user's appearance fore-cast the user and recommend music that fits the identified mood.

Madhuri Athavle et al. [8] proposed a new approach for playing music automati-cally using facial emotion. The majority of current methods include employing wear-able computers, manually playing music, or categorizing audio aspects. They suggest altering the manual sorting and playing instead. To identify emotions, they have employed a convolutional neural network. Pygame and Tkinter are used to suggest music. The system they've suggested tends to shorten the amount of time needed to compute the outcomes and the total cost of the system's design, raising the system's overall precision. The FER2013 dataset is used to test the system. Face expressions are recorded using an integrated camera. On the input, feature extraction is done. Face photos are used to identify emotions including happiness, anger, sadness, sur-prise, and neutrality. A playlist of songs is automatically created by determining the current sentiment of the consumer.

G. Chidambaram et al. [9] proposed a system that mainly aims to detect human emotions for developing an emotion-based music player. Their method involves em-ploying Deep Neural Networks (DNN) to directly extract the best suitable feature abstractions from the data collected in an uncontrolled setting and manage the re-strictions of features that are handmade. The suggested system uses the VGG16 CNN model to identify a user's facial expressions and make decisions based on those ex-pressions. Following the classification of the emotion, the tune that matched the us-er's feelings would play.

A. Phaneendra et al. [10] suggested a system that combines computer vision and machine learning methods to identify facial emotions and suggest music depending on those emotions. The method involves learning the best feature abstraction via Deep Neural Networks. With the usage of a convolutional neural network model, the suggested system can recognize the user's facial expressions. The music that best expresses the user's feelings will be shown once the emotion has been identified. One primary webpage for this project is created with the Streamlit framework, which rec-ords a picture of the user. After the picture is taken, it is sent to the model to forecast the user's sentiment. After identifying the mood, the Spotify API is contacted by the Spotify Python program to submit requests for music tracks, which are then shown on the user interface.

C. Rashmi et al. [11] proposed a project which's goal is to use a person's facial ex-pressions to convey their emotions. The web camera interface for computer systems is used by a music player to record human emotion. The program records the picture of the user, followed by the use of image processing and segmentation methods takes facial characteristics from a particular individual, and attempts to identify the mood of the person attempting to communicate. The project's goal is to uplift the user's spirits by playing music by taking a picture of the user and recording their needs.

Divyansh Shukla et al. [12] suggest a unique method for creating individualized music suggestions utilizing facial expression analysis in music recommendation sys-tems. The system analyzes the user's facial expressions using machine learning algo-rithms to ascertain their emotional state. Based on this information, it then suggests music that fits the user's mood. The suggested method makes use of a collaborative filtering algorithm to provide tailored music suggestions and a deep learning-based model to identify and categorize facial emotions. A collection of participant evalua-tions for music preferences and facial expression photos was used to assess the sug-gested approach.

Shakthi Sri et al. [13] suggested a method for anticipating facial expressions. More picture samples are gathered and compared with various classifications. Shape- and texture-oriented elements are the main components of the picture for preparing the dataset beforehand and using more than one CNN architecture. Compare each archi-tecture since they all provide varying degrees of accuracy. They constructed a hierar-chical model to store the trained model in the end.

Rakhitha et al. [14] proposed an approach that aims to improve the robustness and accuracy of facial emotion identification by offering personalized activation keys that reduce the influence of face masks and highlight the visible facial parts. A novel approach that combines state-of-the-art techniques, specific convolutional neural network (CNN) layers, and altered activation keys has been devised to correct-ly identify facial expressions and provide music. Face masks provide a problem, thus the approach makes use of state-of-the-art techniques to assess the facial character-istics that remain visible, such as the brows and eyes, to effectively collect emotional information. The researcher proposes adding specialized layers that focus exclusively on the visible face areas to current CNN architectures to achieve this.

These studies demonstrate the diverse approaches taken to understand the com-plex interaction between music and facial expressions. The understanding of the relationship between facial expressions and music can contribute to the development of AI-based music recommendation systems.

2.3 Comparative Analysis and Summary

The literature review presents an array of studies investigating the fusion of facial expressions and music recommendation systems. Various methodologies utilize facial recognition algorithms, convolutional neural networks (CNNs), and deep learning models to detect emotions from facial expressions for personalized music suggestions. These studies focus on leveraging computer vision and machine learning techniques to interpret human emotions and recommend music that aligns with the detected emotional states. Approaches include real-time emotion detection through webcam technology, microexpression analysis, feature extraction, and employing CNN architectures like VGG16 and ResNet-38. The goal across these studies is to create emotion-based music players or recommendation systems that cater to individual moods, aiming to enhance user experience by automating playlist creation based on facial expressions. Moreover, some studies address challenges related to facial masks impacting emotion recognition accuracy. Overall, these research endeavors underscore the potential of understanding the intricate connection between facial expressions and musical preferences to advance AI-driven music recommendation systems.

2.4 Challenges

Here are some potential challenges associated with the research described in the abstract:

1. **Diversity and Bias in the Dataset:** It might be difficult to compile a varied dataset that fairly depicts a range of ages, ethnicities, and facial expressions. Predictions or suggestions based on a biased dataset may be incorrect, particularly if the dataset is not inclusive or diverse.
2. **Generalization Across Emotions:** Since emotions are multifaceted and intricate, it might be difficult to precisely group them into discrete groups such as surprise, happiness, or wrath. Mixed emotions can be conveyed by some facial expressions, making it challenging to generalize across different people.
3. **Accuracy and Reliability:** While CNN and VGG16 models may achieve high accuracies, it might be difficult to guarantee robustness in a variety of lighting situations, facial angles, and emotions. It may be challenging for real-world applications to reliably detect emotions with this level of accuracy.
4. **Real-Time Processing:** The computing burden and delay associated with real-time facial expression analysis for instantaneous music suggestions might be problematic. For a flawless user experience, accuracy and real-time performance must be balanced.
5. **Ethical Concerns:** The use of face recognition technology gives rise to ethical questions about permission from users, privacy, and the exploitation of facial data. It is important to ensure ethical

concerns in the gathering, storing, and use of data.

6. **User Acceptance and Adaptability:** Users may see the system differently when it comes to how well it predicts emotions and makes music recommendations. It might be difficult to guarantee user acceptance and adjust to personal preferences or changing emotional states.
7. **Integration and Scalability:** It might be difficult to integrate a produced system with already-existing platforms or apps, such as Spotify, and to make sure it can scale to accommodate a big user base. During deployment, compatibility problems or system scalability might provide challenges.
8. **Interdisciplinary Challenges:** Multidisciplinary cooperation is necessary to bridge the gap between computer vision, machine learning, and music recommendation systems. It is vital to guarantee efficient communication and cooperation among specialists from various domains.

To address these issues and develop more precise, dependable, and morally sound emotion-based music recommendation systems, specialists in domains such as computer vision, psychology, data ethics, and music recommendation systems must constantly explore, innovate, and work together.

CHAPTER 3
RESARCH METHODOLOGY

3.1 Introduction

After the user's image is captured and input is received, our system starts a laborious face detection procedure. By utilizing the well-established effectiveness of Convolutional Neural Networks (CNNs), the system is able to recognize and extract complex face traits with ease. This makes it possible to accurately identify and categorize the user's emotional state.

Our system's primary purpose is to use these face feature analyses to identify emotions. Our method determines the user's primary mood properly by assigning certain emotional categories to the retrieved face data. The core feature of our technology, personalized playlist selection, is based on this deep awareness of the user's emotional state.

Using a large music data collection, mostly from Spotify, our algorithm creates playlist recommendations that naturally correspond with the user's identified emotional state. With the help of facial expression analysis and music suggestion, users should be able to enjoy a highly customized and emotionally impactful music listening experience.

Our approach aims to merge powerful machine learning algorithms with sophisticated face identification techniques to bridge the gap between music tastes and facial emotions. This smooth integration not only increases user interaction but also demonstrates how emotion detection technology can be used to improve user-focused products such as music recommendation systems.

3.2 Workflow

The system works with an input by capturing an image of the user. Then the face detection process starts. Convolutional Neural Network is proven to be very effective for this task. By extracting the facial feature, the system detects the emotion of the user and recommends the playlist based on that mood.

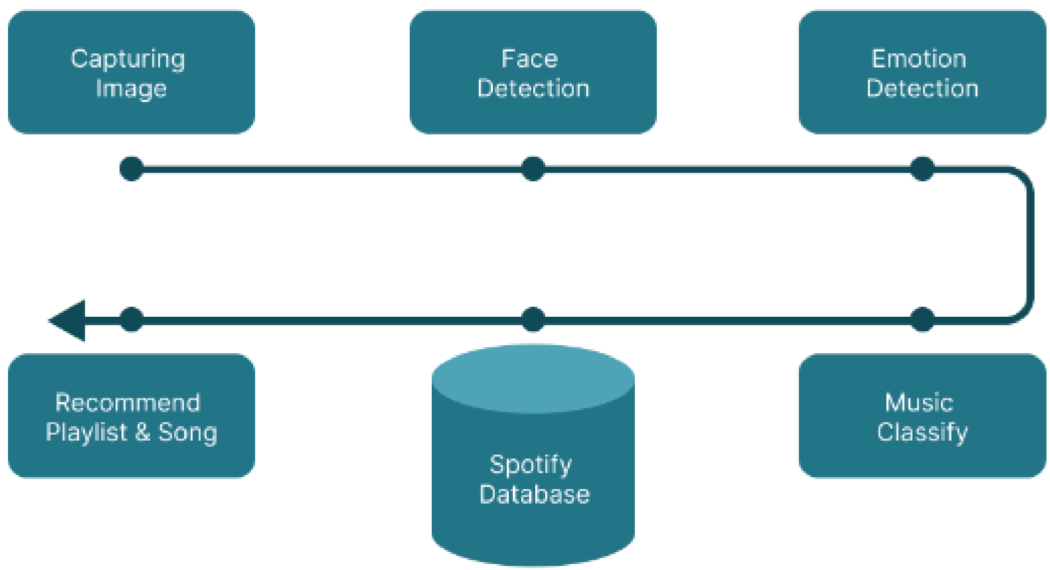

Fig 3.2.1: Workflow of the proposed system.

The details of the workflow:

1. Capturing Image:

- Using a camera or webcam to capture an image of the user's face.
- We need to ensure optimal lighting conditions and image quality for accurate processing.

2. Face Detection:

- Employing face detection algorithms to locate and isolate faces within the captured image.
- Handling potential challenges like multiple faces, partial occlusions, or varying angles.

3. Emotion Detection:

- Analyzing facial expressions to detect emotions such as happiness, sadness, anger, surprise, fear, or disgust.

- Utilizing specialized emotion recognition algorithms trained on the given datasets.

4. Music Classify:

- Categorizing or classifying music tracks based on their genre, mood, tempo, or other relevant attributes.

- Using music information retrieval techniques or rely on existing music databases.

5. Integrate with Spotify Database:

- Access and retrieve music tracks from the Spotify database based on recommendations.

6. Recommend Playlist & Song:

- Combining emotion detection results with music classification to suggest suitable music playlists or songs.

- Matching detected emotions with music genres or moods that align with those emotions.

- Considering personal preferences or listening history for more tailored recommendations.

3.3 Experiment Dataset

We trained the Model using a total of 37,675 images. We used 30,254 images for the train and 7,421 images for the test. The data is a mix of our own collected images and the Kaggle FER2013 dataset. All of the images used in grayscale and pixel value 48 * 48.

	contempt	fear	surprise	sad	happy	neutral	disgust	angry
train	215	4264	3386	5020	7424	5179	635	4131

	contempt	fear	surprise	sad	happy	neutral	disgust	angry
test	36	1046	861	1264	1810	1268	142	994

Fig 3.3.1: Data on every class

Our Dataset: We've collected around 1,800 images of 8 different expressions. About 225 participants voluntarily participated in our project for the image dataset. We've preprocessed those images to grayscale images with a size of 48 * 48 pixels.

Kaggle Dataset: The FER2013 data set from Kaggle was used for model training. FER2013 has 35,886 images in total where 28,708 images for training and 7,178 images for testing. All are grayscale images with a fixed size of 48 * 48. For recommending songs, we used Spotify music data by moods from Kaggle.

3.4 Proposed Methodology

The proposed system operates by initiating a face detection process upon capturing an image of the user. Leveraging Convolutional Neural Network (CNN), renowned for its exceptional effectiveness in image analysis, the system undergoes facial feature extraction to discern the user's emotional state. This intricate process enables the system to accurately detect and categorize the user's emotions from the facial expressions captured in the image. Based on this detected emotional state, the system proceeds to recommend a personalized playlist tailored to match the user's mood.

This streamlined approach harnesses the power of advanced computer vision techniques facilitated by CNN, allowing for precise recognition and interpretation of facial features indicative of various emotional expressions. By seamlessly integrating this technology, the system can discern nuanced emotions, encompassing a spectrum from happiness and sadness to surprise and neutrality, enhancing the accuracy of playlist recommendations. Ultimately, the system endeavors to intuitively connect the user's emotional state with a curated selection of music, aiming to deliver a highly personalized and enriching music listening experience.

3.5 Model

Optimizers: The optimizers are functions that change the parameters of our neural network, such as the model's learning rate and weight hyperparameters, to reduce losses and speed up results. We employ the two optimizers in our study.

1. Function of ReLU Activation.
2. Adam Optimizer.

Convolutional Neural Network: To detect facial expressions, the neural network is most fruitful. Neural network gains knowledge by ingesting examples. The network modifies the weights connected to the neurons. Subsequent modifications often impart additional knowledge to the model and improve its performance. In this instance, convolutional neural networks, or CNNs, are employed. Artificial Neural

Networks (ANN), which need 1-dimensional vectors, cannot be used since the picture is a multi-dimensional array. Consequently, a large increase in training parameter numbers is achieved at the expense of a subpar outcome. In contrast, CNN is a feed-forward network that analyzes and recognizes image input. Eight types of emotions are detected by the project: fear, surprise, anger, contempt, happiness, sadness, disgust, and neutral.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 48, 48, 32)        320

 conv2d_1 (Conv2D)           (None, 48, 48, 64)        18496

 batch_normalization (Batch  (None, 48, 48, 64)        256
 Normalization)

 max_pooling2d (MaxPooling2  (None, 24, 24, 64)        0
 D)

 dropout (Dropout)           (None, 24, 24, 64)        0

 conv2d_2 (Conv2D)           (None, 24, 24, 128)       73856

 conv2d_3 (Conv2D)           (None, 22, 22, 256)       295168

 batch_normalization_1 (Bat  (None, 22, 22, 256)       1024
 chNormalization)

 max_pooling2d_1 (MaxPoolin  (None, 11, 11, 256)       0
 g2D)

 dropout_1 (Dropout)         (None, 11, 11, 256)       0

 flatten (Flatten)           (None, 30976)             0

 dense (Dense)               (None, 1024)              31720448

 dropout_2 (Dropout)         (None, 1024)              0

 dense_1 (Dense)             (None, 8)                 8200

=================================================================
Total params: 32117768 (122.52 MB)
Trainable params: 32117128 (122.52 MB)
Non-trainable params: 640 (2.50 KB)
```

Fig 3.5.1: Summary of the Convolutional Neural Network.

VGG16 CNN: VGG16 CNN was trained by our dataset images to improve the model. The model performed very well.

.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 vgg16 (Functional)          (None, 1, 1, 512)         14714688

 batch_normalization (Batch  (None, 1, 1, 512)         2048
 Normalization)

 gaussian_noise (GaussianNo  (None, 1, 1, 512)         0
 ise)

 global_average_pooling2d (  (None, 512)               0
 GlobalAveragePooling2D)

 flatten (Flatten)           (None, 512)               0

 dense (Dense)               (None, 256)               131328

 batch_normalization_1 (Bat  (None, 256)               1024
 chNormalization)

 dropout (Dropout)           (None, 256)               0

 dense_1 (Dense)             (None, 128)               32896

 batch_normalization_2 (Bat  (None, 128)               512
 chNormalization)

 dropout_1 (Dropout)         (None, 128)               0

 dense_2 (Dense)             (None, 8)                 1032

=================================================================
Total params: 14883528 (56.78 MB)
Trainable params: 13146248 (50.15 MB)
Non-trainable params: 1737280 (6.63 MB)
```

Fig 3.5.2: Summary of the VGG16 CNN

3.6 Training the model

Training CNN: During the training phase of the Convolutional Neural Network (CNN), a comprehensive training regimen was employed, spanning across 60 epochs. The optimization process was facilitated using the Adam optimizer, fine-tuned with a specific learning rate of 0.0001 coupled with a decay rate set at 1e-6. Through this meticulously crafted training protocol, the CNN underwent iterative learning and refinement, culminating in the attainment of an approximate accuracy of 89% within the train set.

This rigorous training methodology aimed to optimize the CNN's ability to discern and classify facial expressions accurately. The utilization of the Adam optimizer, renowned for its effectiveness in optimizing deep neural networks, coupled with a carefully selected learning rate and decay parameter, facilitated robust and consistent learning throughout the training epochs. The achieved 89% accuracy within the train set serves as a testament to the network's proficiency in learning to identify and categorize various facial expressions, thereby laying a solid foundation for accurate emotion detection in subsequent phases of the system.

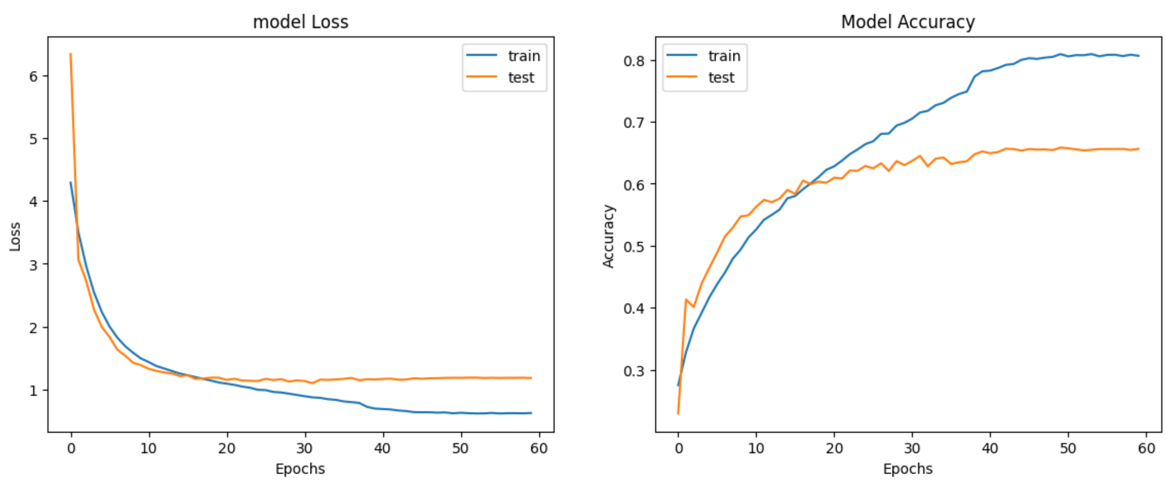

Fig 3.6.1: Accuracy graphs during training.

Training VGG16 CNN: During the training phase, the VGG16 CNN model exhibited exceptional performance metrics, achieving outstanding accuracy levels. Impressively, on the train set, the model demonstrated a remarkable accuracy of 97.43%, showcasing its proficiency in learning and accurately classifying the provided data. Moreover, during the validation phase, the VGG16 CNN model maintained a high level of accuracy, achieving a commendable rate of 92%.

These exemplary performance statistics underscore the robustness and effectiveness of the VGG16 CNN model in comprehensively learning from the training data while also demonstrating robust generalization capabilities during the validation process. The model's ability to achieve such high accuracy levels signifies its capacity to discern and classify intricate patterns within the data, showcasing its potential for reliable and precise facial expression recognition.

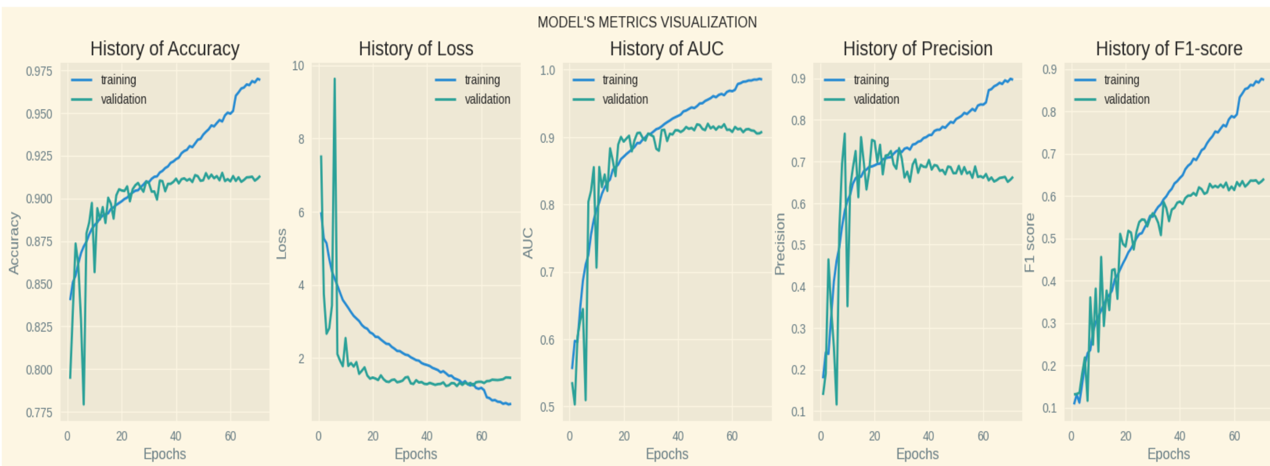

Fig 3.6.2: Accuracy graphs of VGG16 CNN

3.7 Redirecting to Songs

Following the successful detection of the user's emotional expression, the system will dynamically curate and suggest playlists along with individual songs sourced from the expansive Spotify music dataset. Leveraging the nuanced understanding of the user's detected mood, the system meticulously tailors its recommendations, ensuring alignment with the emotional state discerned from the facial expression analysis.

This intricate process aims to provide a rich and diverse selection of music, precisely matching the identified mood of the user. By tapping into the comprehensive Spotify music dataset, the system offers an extensive array of playlists and songs across various genres and themes, meticulously curated to resonate with the user's detected emotional context.

Furthermore, the recommendations are thoughtfully designed to encompass a spectrum of emotions—ranging from upbeat and cheerful melodies for moments of joy to soothing and contemplative tunes for instances of introspection or tranquility. This personalized approach endeavors to enhance the user experience by fostering a deep emotional connection through music that perfectly complements and mirrors the detected mood, ultimately enriching the listening journey.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Performance Analysis

We trained the CNN on epochs 60. With Adam optimizer and learning-rate=0.0001 with decay=1e-6. The accuracy we got was around 89% in the train set and around 66% in the validation set.

```
473/473 [==============================] - 25s 53ms/step - loss: 0.4392 - accuracy: 0.8889
116/116 [==============================] - 4s 31ms/step - loss: 1.1819 - accuracy: 0.6556
final train accuracy = 88.89 , validation accuracy = 65.56
```

Fig 4.1.1: Accuracy of CNN.

The VGG16 CNN model got an excellent performance. It got an accuracy of 97.43% in the train set and around 92% on validation set.

```
473/473 [==============================] - 49s 103ms/step - loss: 0.4013 - accuracy: 0.9743 - precision: 0.9134 - recall: 0.8772 - auc: 0.9920 - f1_score: 0.8950
116/116 [==============================] - 4s 31ms/step - loss: 1.4575 - accuracy: 0.9131 - precision: 0.6631 - recall: 0.6200 - auc: 0.9080 - f1_score: 0.6406
```

Fig 4.1.2: Accuracy of VGG16.

CNN Confusion Matrix:

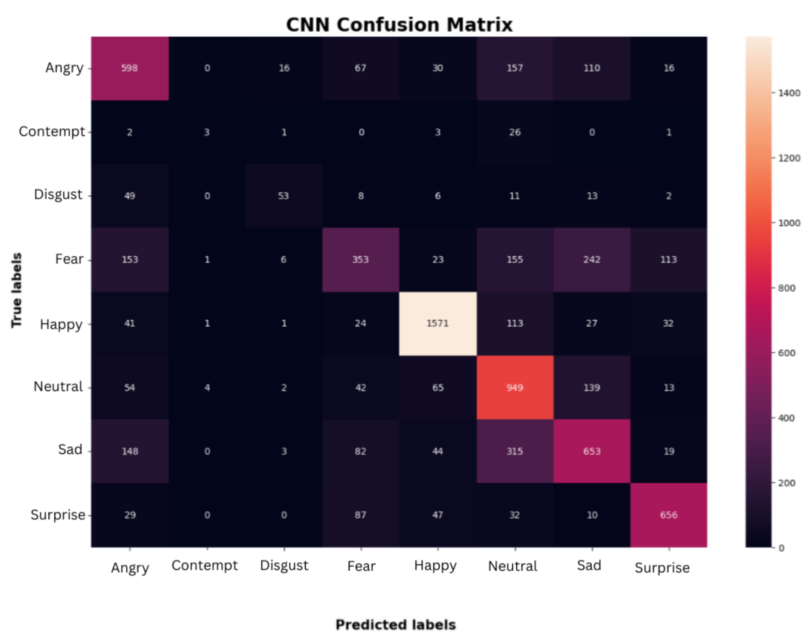

Fig 4.1.3: Confusion matrix of CNN.

VGG16 Confusion Matrix:

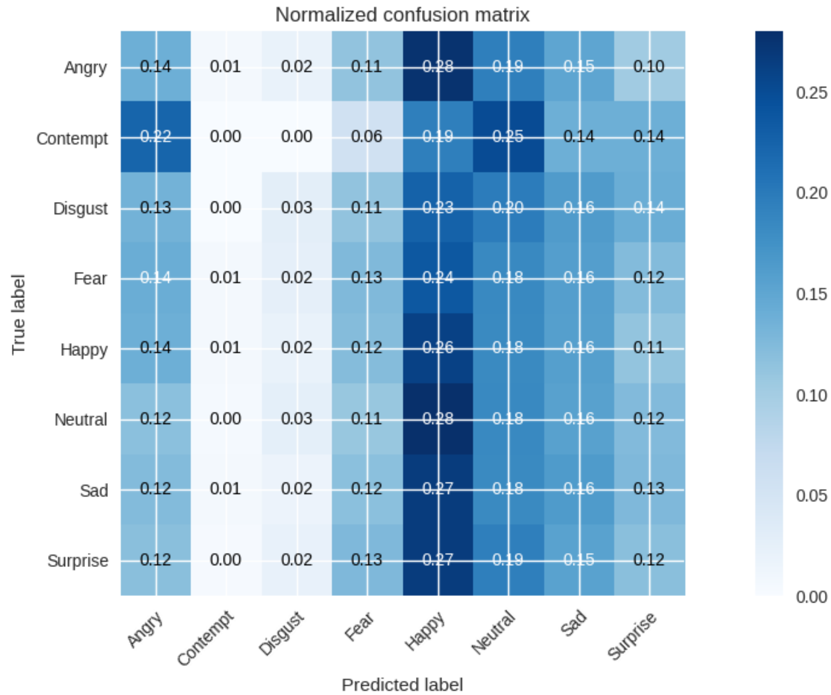

Fig 4.1.4: Confusion matrix of VGG16 CNN.

4.2 Result Discussion

The comprehensive evaluation of our project's models, namely the Convolutional Neural Network (CNN) and the VGG16 CNN, exhibited exceptional performance in detecting facial expressions from our dataset. Notably, our meticulously trained CNN model showcased a commendable accuracy level of approximately 89% on train and 66% on validation, demonstrating robust capabilities in discerning and categorizing various facial expressions with a high degree of precision. Equally impressive was the performance of our VGG16 CNN model, which remarkably achieved an outstanding accuracy rate of approximately 97% on train and 92% on validation, underscoring its prowess in accurately identifying and classifying facial expressions within our dataset.

Moreover, beyond the proficient facial expression detection, our system seamlessly integrates this capability into recommending songs sourced from the vast Spotify dataset, catering to the user's discerned emotional state. Leveraging the insights garnered from the accurate detection of facial expressions, our system adeptly aligns the user's detected mood with an array of song recommendations meticulously curated from Spotify's expansive music repository.

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

Our system stands as a testament to the successful fusion of facial expression detection and music recommendation, creating a seamless interface that aligns a user's emotional state with personalized song recommendations from the extensive Spotify dataset. This pioneering work not only demonstrates the effectiveness of detecting faces, emotions, and facial expressions but also signifies a significant leap forward in the realm of leveraging these insights for music-related applications.

While achieving commendable accuracy rates in emotion detection and song recommendation, the system's potential for further enhancement is evident. One avenue for improvement involves augmenting the existing datasets with a more extensive range of facial expressions, encompassing diverse demographics, cultural nuances, and emotional variations. By broadening the dataset, the system can strengthen its adaptability and accuracy in recognizing a wider spectrum of human emotions, thereby enhancing the precision of music recommendations tailored to individual moods.

Additionally, ongoing advancements in machine learning and computer vision techniques offer avenues for refining our methods. Exploring cutting-edge methodologies, such as leveraging state-of-the-art neural network architectures or integrating multimodal learning approaches that combine facial expressions with other behavioral cues, presents an exciting opportunity to further elevate the system's performance and accuracy.

Furthermore, user feedback and engagement are crucial for system optimization. Incorporating user preferences, ratings, and interactions with recommended playlists can facilitate a more personalized and intuitive recommendation system. Iterative improvements guided by user feedback would not only enhance the accuracy of song recommendations but also ensure an enhanced user experience, fostering increased user satisfaction and engagement with the system.

In conclusion, while our system showcases a commendable performance in its current iteration, the future trajectory involves continual refinement through the expansion of datasets, adoption of advanced methodologies, and active user involvement. These enhancements will collectively propel the system towards becoming a more sophisticated, adaptable, and user-centric platform at the intersection of emotion detection and music recommendation.

5.2 Future Works

The successful development of our proposed system, which effectively detects facial expressions and recommends music based on the user's mood, marks a significant milestone. Moving forward, several promising directions can be explored to further enhance and refine the system's capabilities.

Firstly, expanding and diversifying the datasets used for facial expression detection would be crucial. Incorporating larger and more diverse datasets encompassing various demographics, cultural backgrounds, and a wider range of emotional expressions could significantly improve the system's accuracy and applicability across different user groups.

Secondly, the exploration of advanced model architectures beyond the Convolutional Neural Network (CNN) and VGG16, such as ResNet, DenseNet, or Transformer-based models, holds promise. Experimenting with these sophisticated architectures, including ensemble models or multimodal approaches combining facial expressions with other behavioral cues, could potentially elevate the system's accuracy and efficiency.

Thirdly, efforts to fine-tune and optimize the models are essential. Rigorous experiments aimed at refining hyperparameters, optimizing training strategies, and addressing potential overfitting issues could lead to improved model performance and generalization across diverse user populations.

Additionally, focusing on user-centric improvements would be pivotal. This involves integrating mechanisms for user feedback, personalized recommendations based on user profiles, and analyzing sentiment in song lyrics to enhance the system's ability to tailor music recommendations according to individual preferences.

Ethical considerations and bias mitigation strategies should be continuously addressed. Ensuring fairness, transparency, and privacy in facial expression detection and music recommendation algorithms is critical for responsible and ethical deployment.

Furthermore, exploring real-time applications and optimizing the system's responsiveness for instant music recommendations based on live facial expressions or user interactions could significantly enhance the user experience.

Conducting comprehensive user studies and usability tests to gauge user satisfaction, acceptance, and the system's impact on user behavior and emotions would provide invaluable insights for refining the system based on real-world user feedback.

REFERENCES

[1] James, H. I., Arnold, J. J. A., Ruban, J. M. M., Tamilarasan, M., & Saranya, R. (2019). Emotion based music recommendation system. Emotion, 6(3).

[2] Yu, Z., Zhao, M., Wu, Y., Liu, P., & Chen, H. (2020, July). Research on automatic music recommendation algorithm based on facial micro-expression recognition. In 2020 39th Chinese control conference (CCC) (pp. 7257-7263). IEEE.

[3] Florence, S. M., & Uma, M. (2020, August). Emotional detection and music recommendation system based on user facial expression. In IOP conference series: Materials science and engineering (Vol. 912, No. 6, p. 062007). IOP Publishing.

[4] Zhang, J. (2020, October). Movies and pop songs recommendation system by emotion detection through facial recognition. In Journal of Physics: Conference Series (Vol. 1650, No. 3, p. 032076). IOP Publishing.

[5] Shalini, S. K., Jaichandran, R., Leelavathy, S., Raviraghul, R., Ranjitha, J., & Saravanakumar, N. (2021). Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques. Turkish journal of computer and mathematics education, 12(2), 912-917.

[6] VIDYASAGAR, B. S., & KARWANDE, V. (2021). EMOTION BASED MUSIC RECOMMENDATION SYSTEM BY USING DIFFERENT ML APPROACH. INTERNATIONAL JOURNAL, 6(6).

[7] Mahadik, A., Milgir, S., Patel, J., Jagan, V. B., & Kavathekar, V. (2021). Mood based music recommendation system. INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume, 10.

[8] Athavle, M., Mudale, D., Shrivastav, U., & Gupta, M. (2021). Music recommendation based on face emotion recognition. Journal of Informatics Electrical and Electronics Engineering (JIEEE), 2(2), 1-11.

[9] Chidambaram, G., Ram, A. D., Kiran, G., Karthic, P. S., & Kaiyum, A. (2021). Music recommendation system using emotion recognition. International Research Journal of Engineering and Technology (IRJET), 8(07), 2395-0056.

[10] Phaneendra, A., Muduli, M., Reddy, S. L., & Veenasree, R. (2022). EMUSE–An emotion based music recommendation system. International Research Journal of Modernization in Engineering Technology and Science, 4(5), 4159-4163.

[11] Rashmi, C., Akshaya, A., Sruthi, B., Lavanya, B., & Lohitha, A. (2023). EMOTION DETECTION MODEL-BASED MUSIC RECOMMENDATION SYSTEM. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 14(03), 26-33.

[12] Shukla, D., Singh, S. P., Saurabh, S., & Kumar, A. Music Recommendation using Facial Expression.

[13] Sathya, S., & Pandiarajan, T. (2023). An Improved Music Recommendation System for Facial Recognition and Mood Detection. In ITM Web of Conferences (Vol. 56, p. 01004). EDP Sciences.

[14] Wimalaweera, R. (2023). A Review on Music recommendation system based on facial expressions, with or without face mask.

Play list and playing preferences

ORIGINALITY REPORT

24%	19%	10%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	9%
2	webmail.thescipub.com Internet Source	1%
3	ijarsct.co.in Internet Source	1%
4	www.ijert.org Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	"Computer Vision and Image Processing", Springer Science and Business Media LLC, 2021 Publication	1%
7	Submitted to University of Westminster Student Paper	1%
8	Submitted to George Bush High School Student Paper	<1%

ijasret.com