Emotion Detection - Music Recommendation System using Machine Learning

A Project Report

Submitted to the APJ Abdul Kalam Technological University

in partial fulfillment of requirements for the award of degree

Bachelor of Technology

in

Computer Science and Engineering

by

Althaf P Shanavas (CHN19CS014)

Alwin John (CHN19CS016)

Devika J S (CHN19CS046)

Hemanth S (CHN19CS060)

Robin Roy (CHN19CS101)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COLLEGE OF ENGINEERING CHENGANNUR

KERALA

July 2022

DEPT. OF COMPUTER SCIENCE & ENGINEERING

COLLEGE OF ENGINEERING CHENGANNUR

CERTIFICATE

This is to certify that the report entitled **Emotion Detection - Music Recommendation System using Machine Learning** submitted by **Althaf P Shanavas** (CHN19CS014), **Alwin John** (CHN19CS016), **Devika J S** (CHN19CS046), **Hemanth S** (CHN19CS060), & **Robin Roy** (CHN19CS101) to the APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Computer Science and Engineering is a bonafide record of the project work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Mrs. Naseena N
(Project Guide)
Dept.of CS
College of Engineering
Chengannur

Mrs. Angel Thomas
(Project Coordinator)
Dept.of CS
College of Engineering
Chengannur

Dr. Manju S Nair
Associate Professor and Head
Dept.of CS
College of Engineering
Chengannur

DECLARATION

We hereby declare that the project report **Emotion Detection - Music Recommendation System using Machine Learning**, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under supervision of Mrs. Naseena N and Mrs. Angel Thomas

This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources.

We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Althaf P Shanavas

Alwin John

Devika J S

Hemanth S

Robin Roy

Chengannur

27-07-2022

Abstract

Problem Statement: It is often confusing for a person to decide which music they have to listen to from a massive collection of existing options. Several suggestion frameworks have been available for issues like music, dining, and shopping depending on the user's mood

Solution: The proposed system detects the emotions, if the subject has a negative emotion then a specific playlist will be presented that contains the most suitable types of music that will improve his mood. On the other hand, if the detected emotion is positive, a suitable playlist will be provided, including different music types that will enhance the positive emotions

Implementation: The project works by getting a live video feed from a webcam, and passing it through the model to get a prediction of emotion. Then according to the emotion predicted, the app will fetch a playlist of songs from Spotify through the Spotipy wrapper and recommend the songs by displaying them on the screen

Technology used:

TensorFlow - For image processing and training the model

OpenCV - For face detection

Flask - To develop a web application

Social Impact: Developing a recommendation system could assist a user to make a decision regarding which music one should listen to helping the user to reduce his/her stress levels. The user would not have to waste any time searching or looking up for songs and the best track matching the user's mood is detected, and songs would be shown to the user according to their mood

Acknowledgement

We take this opportunity to express our deepest sense of gratitude and sincere thanks to everyone who helped us to complete this work successfully. We express our sincere thanks to **Dr. Manju S Nair**, Head of Department, Department of Computer Engineering, College of Engineering Chengannur for providing us with all the necessary facilities and support.

We would like to express our sincere gratitude to the Department of Computer Engineering, College of Engineering Chengannur for the support and co-operation.

We would like to place on record our sincere gratitude to our project co-ordinators **Mrs. Naseena N** and **Mrs. Angel Thomas** for the guidance and mentorship throughout this work

Althaf P Shanavas

Alwin John

Devika J S

Hemanth S

Robin Roy

Contents

Abstract **i**

Acknowledgement **ii**

List of Figures **v**

1 Introduction **1**

1.1 Introduction 1

1.2 Motivation 2

1.3 Problem Statement 2

1.4 Proposed Solution 2

2 Software Requirement Specification **3**

2.1 Introduction 3

2.1.1 Purpose 3

2.1.2 Intended Audience and Reading Suggestions 3

2.1.3 Overview 3

2.2 Functional requirements 4

2.3 Interface requirements 4

2.4 Non functional atributes 5

3 Project Design **6**

3.1 Introduction 6

3.2 Model Architecture 6

3.3 Application Architecture Design 7

3.4 GUI Design 8

3.5 Technology Stack 8

4 Implementation **10**

4.1 Dataset 10

4.2 Proposed System 11

4.3 Image Processing and Training 12

4.4 User Interface 13

5 Results **14**

5.1 Current condition 14

5.2 Project Components 14

6 Conclusion & Future Work **15**

References **17**

List of Figures

1.1 Face detection 1

3.1 Model Architecture - Facial-expressions-recognition 6
3.2 Structural Design of the Application 7
3.3 Graphical User Interface for the Application 8

4.1 Dataset of one of the expressions 10
4.2 Normalization: Scaling data to the range of 0-1 12

Chapter 1

Introduction

1.1 Introduction

Communicating between individuals is a major aspect of everyday life. It conveys accurate details and millions of information among humans, whether in the form of words, tone or expression. The expression provided by the face and body is the best means to understand people in communication. In particular, the facial expression is one of the basic elements of human communication and can be considered a form of nonverbal communication [1]. Facial expressions refer to a movement of one or more facial muscles or skin. The aim of these movements is to express the facial emotions of an individual such as happy, sad, natural and surprised.

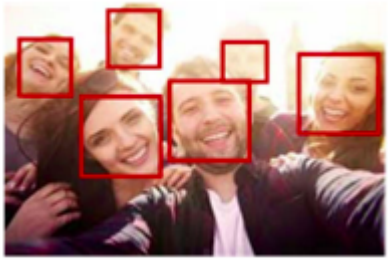

Figure 1.1: Face detection

Face detection and emotion detection are two separate concepts that are considered are given high significance in recent research fields like digital image processing, pattern recognition communities, and computer vision. Face detection is defined as the process

that is composed of two steps: first, finding faces in a picture or video and does not care if the face refers to a person or not. Then, it draws a surround box of every face in the image in any expression, orientation, facial pose, illumination, and occlusions, shown in figure 1. Emotion detection is considered as one of the facial expressions and is the main process for identifying emotions of human's face.

1.2 Motivation

Music is the form of art known to have a greater connection with a person's emotion. If a user receives a recommendation based on his preference, it will also improve his listing experience. Developing a recommendation system could assist a user to make a decision regarding which music one should listen to helping the user to reduce their stress levels.

1.3 Problem Statement

Nowadays, emotion detection is considered as one of the most important techniques that are used in many applications such as smart card application, surveillance, image database investigation, criminal, video indexing, civilian applications, security and adaptive human-computer interface with multimedia environments. It is often confusing for a person to decide which music they have to listen to from a massive collection of existing options. Several suggestion frameworks have been available for issues like music, dining, and shopping depending on the user's mood

1.4 Proposed Solution

The proposed system detects the emotions, if the subject has a negative emotion then a specific playlist will be presented that contains the most suitable types of music that will improve his mood. On the other hand, if the detected emotion is positive, a suitable playlist will be provided, including different music types that will enhance the positive emotions

Chapter 2

Software Requirement Specification

2.1 Introduction

2.1.1 Purpose

The purpose of this document is to define and describe the requirements of the project and to spell out the system's functionality and its constraints.

2.1.2 Intended Audience and Reading Suggestions

This Software Requirement Specification is for developers, project managers, users and testers. Further, the discussion will provide all the internal, external, functional and also non-functional information about Emotion Detection - Music Recommendation System.

2.1.3 Overview

The emotion recognition model is trained on FER 2013 dataset. It can detect 5 emotions. The project works by getting live video feed from web cam, pass it through the model to get a prediction of emotion. Then according to the emotion predicted, the app will fetch playlist of songs and recommend the songs by displaying them on the screen.

2.2 Functional requirements

- **Songs** - List of Malayalam songs are collected. So users can get recommendations based on their emotions.
- **Song Recommendations** - Based on the model trained for emotion detection, It can detect 5 emotions, then according to the emotion predicted, the app will fetch playlist of songs
- **Real time expression detection** - The dataset used for this project is the famous FER2013 dataset. Models trained on this dataset can classify 5 emotions. By getting live video feed from web cam, we pass it through the model to get a prediction of emotion.
- **Simple UI** - The web-app has a simple User Interface.

2.3 Interface requirements

User Interface for the project was built using Flask. It is a micro web framework written in Python. Flask abstracts away lower-level tasks, such as setting up a development web server, managing information flow from the browser to the Python interpreter, and more. Using Flask thus allows the developer to focus on the application logic rather than worrying about infrastructural things. Flask is a framework for attaching your code to user actions in a browser, via simple callback functions.

The hardware requirements include the requirements specification of the physical computer resources for a system to work efficiently. The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. The Hardware Requirements for the proposed system include Intel/AMD - 1 GHz or faster processor, 4 GB RAM and 512 GB Hard disk.

The software requirements are description of features and functionalities of the target system. Requirements convey the expectations of users from the software product. The

requirements can be obvious or hidden, known or unknown, expected or unexpected from client‘s point of view. The Hardware Requirements for the proposed system is the Operating System: Windows 7, Windows 10/11.

2.4 Non functional atributes

A non-functional requirement (NFR) is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviours. Non-functional attributes are:

- **Reliability** - The model should be able to detect emotion of each person, since the characteristics of the human expressions differs.
- **Easy to Use** - Objective is to provide the service first, so that the user can get going on deciding the music without waiting much longer.
- **Compatibility** - As of now there is only web version of this, so any laptop, desktop applications is compatible with this.
- **Availability** - It only requires the face of the user and a stable internet connection to get smooth experience.
- **Usability** - It is simple and easy to use as there is no need of sign-up or anything.
- **Maintainability** - As number of feed backs are got from the user, It is important for us to improve the efficiency of the model. As of now it is only 63% accurate

Chapter 3

Project Design

3.1 Introduction

The Emotion Detection-Music Recommendation System project follows a very minimalist approach in terms of development in order to be accessible with ease to its target audience and meet their needs. The working of the software is similarly user friendly.

3.2 Model Architecture

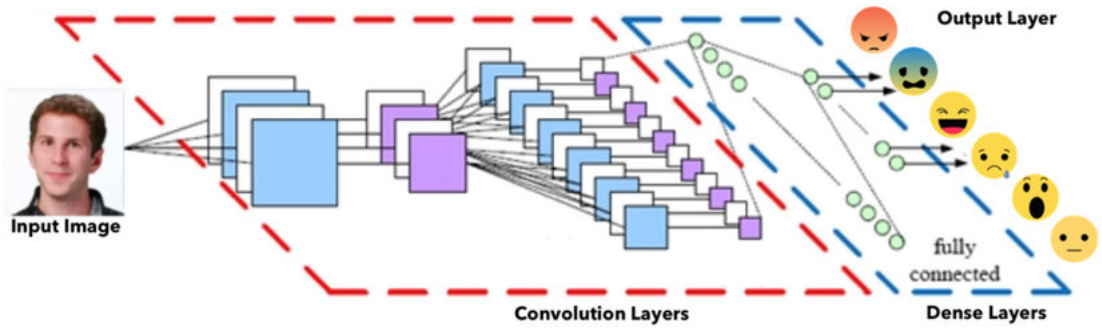

Figure 3.1: Model Architecture - Facial-expressions-recognition

The model architecture is a sequential model consisting of Conv2d, Maxpool2d, Dropout and Dense layers:

- Conv2D layers throughout the model have different filter size from 32 to 128, all with activation 'relu'
- Pooling layers have pool size (2,2)
- Dropout is set to 0.25 as anything above results in poor performance
- Final Dense layer has 'softmax' activation for classifying 5 emotions
- Used 'categorical crossentropy' for loss with 'Adam' optimizer with 'accuracy' metric

3.3 Application Architecture Design

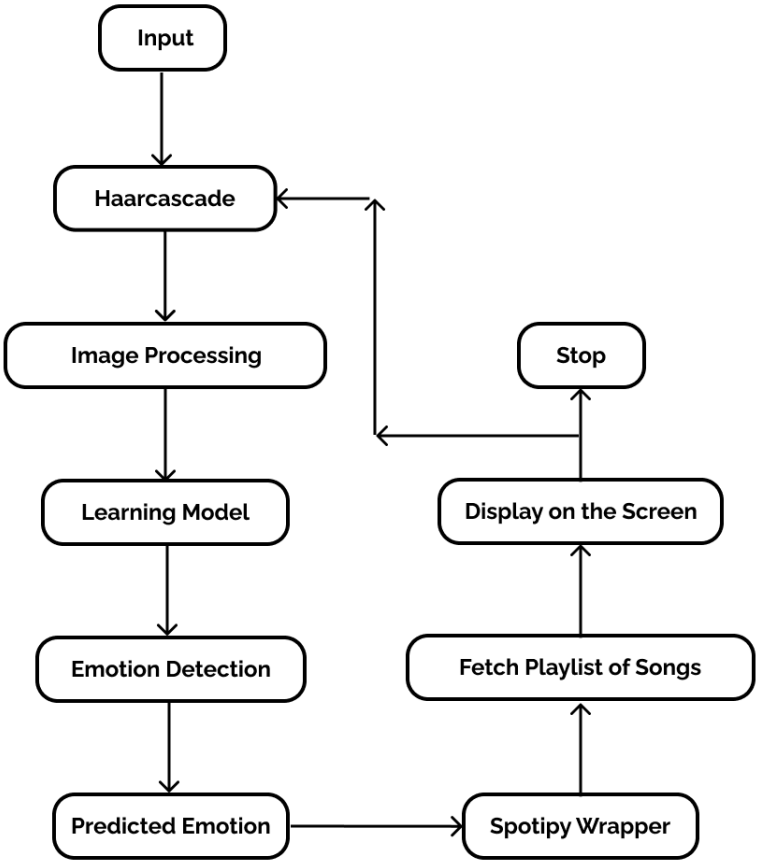

Figure 3.2: Structural Design of the Application

The application is structured in such a manner where the input is taken from the webcam and are sent to the model, then the prediction of the emotion is done. By means of the model predictions are sent determine the emotion then it list the songs then shows the web page where the songs and emotion is displayed in real-time.

3.4 GUI Design

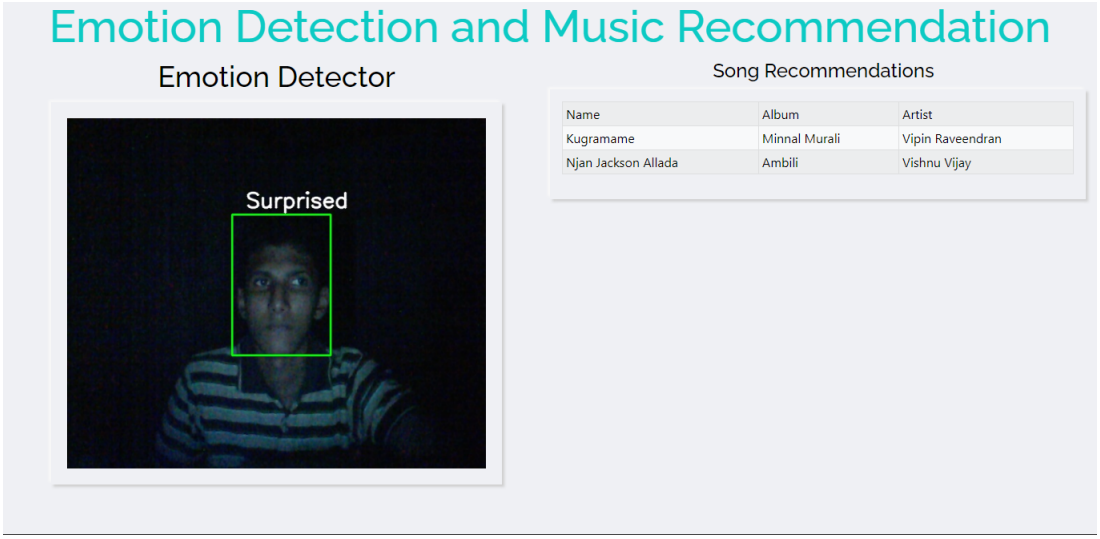

Figure 3.3: Graphical User Interface for the Application

3.5 Technology Stack

Python

It is an object-oriented, high-level programming language with integrated dynamic semantics primarily for web and app development. It is extremely attractive in the field of Rapid Application Development because it offers dynamic typing and dynamic binding options. The libraries of python used in the development of this application includes numpy, pandas, sklearn and matplotlib. NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices. pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool. Scikit-learn (Sklearn)

is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

Flask

Flask is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around Werkzeug and Jinja and has become one of the most popular Python web application frameworks.

HTML CSS

Hypertext Markup Language and Cascading Style Sheets are two of the core technologies for building Web pages. HTML provides the structure of the page while CSS provides the (visual and aural) layout for a variety of devices. Along with graphics and scripting, HTML and CSS are the basis of building Web pages and Web Applications.

Chapter 4

Implementation

4.1 Dataset

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.

The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

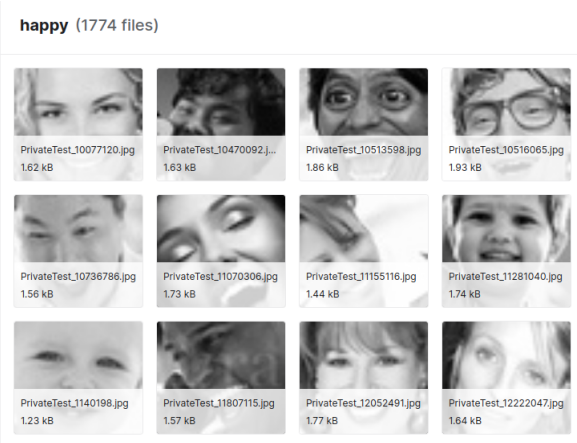

Figure 4.1: Dataset of one of the expressions

4.2 Proposed System

In any machine learning application, the ”data set” required is of at-most importance. It is the key feature that determines the extent of accuracy a method or system can achieve. Hence the data set required for this project is collected from- Kaggle, a subsidiary of Google LLC, an online community of data scientists and machine learning practitioners.

Output is displayed in an effective manner using a web application built using flask which will be later deployed using deployment services such as Heroku, or AWS. This model is very beneficial to the growing younger generation with a better understanding about music aspects. This system will also be helpful for encouraging a love of music in today's youth and helping them stay involved in the field.

4.3 Image Processing and Training

The images were normalised, resized to (48,48) and converted to grayscale in batches of 64 with help of 'ImageDataGenerator' in Keras API.

Training took around 13 hours locally for 75 epochs with an accuracy of 66

```
# Normalization: Scaling data to the range of 0-1 is traditionally referred to as normalization.
train_data_gen = ImageDataGenerator(
    rescale=1.0/255,
    horizontal_flip = True,
    zoom_range=0.2
)

test_data_gen = ImageDataGenerator(
    rescale=1.0/255,
    validation_split = 0.2
)


# Making train data from the set of images.
train_data = train_data_gen.flow_from_directory(
    directory=dir_train,
    target_size= (48,48), #Change later to find the values.
    batch_size=48,
    color_mode='grayscale',
    class_mode='categorical'
)
# Making test data from the set of images.
test_data = train_data_gen.flow_from_directory(
    directory=dir_test,
    target_size= (48,48), #Change later to find the values.
    batch_size=48,
    color_mode='grayscale',
```

Figure 4.2: Normalization: Scaling data to the range of 0-1

4.4 User Interface

The interface for the project was built using Flask. It is a micro web framework written in Python. It is classified as a micro-framework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. The HTML templates that flask with render. These allow the user to input their own data and will present the results. Some of the good things that come with using flask as your framework to develop the web application is little dependency to update and be on the lookout for security bugs. One of the main advantages of using the Flask framework is that it is light weight and extensible Python web framework. The web app receives input values from the webcam and this is passed through the model and the server get the predictions accordingly from the model. Functions for predicting the emotion for each face expressions, along with songs corresponding to the emotion is displayed in the app.

Chapter 5

Results

5.1 Current condition

The entire project works perfectly fine. Live detection gives good frame rates due to multithreading.

5.2 Project Components

- haarcascade is for face detection.
- camera.py is the module for video streaming, frame capturing, prediction and recommendation which are passed to app.py.
- app.py is the main flask application file.
- index.html in 'templates' directory is the web page for the application. Basic HTML and CSS.
- utils.py is an utility module for video streaming of web camera with threads to enable real time detection.
- train.py is the script for image processing and training the model.

Chapter 6

Conclusion & Future Work

The system has successfully been able to capture the emotion of a user and recommend music. It has been tested in a real-time environment for this predicate. However, it has to be tested in different lighting conditions to determine the robustness of the developed system. The system has also been able to grab the new images of the user and appropriately update accordingly.

Limitations

- Opencv tries to open the camera on whatever device the app is running on. Code in current state makes use of webcam if available on server side not client side. So when app is run locally on a laptop Video Streaming through webcam is possible. But if it's deployed to a cloud, the app is stored in a data center somewhere which obviously doesn't have web camera connected to it and hence it doesn't work.
- The system still is not able to record all the emotions correctly due to the less availability of the images in the image dataset being used.
- The quality of the image should be at least higher than 320p for the classifier to predict the emotion of the user accurately.
- Not Every song is listed

Further Work

- Instead of CSVs, create a databse and connect it to application. The DB will fetch songs for recommendations and new songs can be updated directly onto database
- Add a feature which will update specified playlists for better and more recent recommendations, a specific day over a fixed duration say every sunday and append it to database
- Deploy it in Different Platforms
- Directly play the song or redirect to the song on Spotify when user clicks on it.
- Rewrite code such that Video Streaming is done on client side instead of server side so as it make the app deployable

Note: Model accuracy is not that great. It is 66%. Further training and fine tuning is required.

References

[1] Manas Sambare , *Learn facial expressions from an image*, FER-2013, Jan. 2020 from @online Kaggle, `https://www.kaggle.com/datasets/msambare/fer2013` Online; accessed 10-July-2022

[2] Vijay Prakash Sharma; Azeem Saleem Gaded; Deevesh Chaudhary; Sunil Kumar; Shikha Sharma, *Emotion-Based Music Recommendation System*, 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 15 November 2021

[3] R. R. Londhe, P. Vrushshen and Pawar, *Analysis of Facial Expression and Recognition Based On Statistical Approach*, Jan. 2012.

[4] S. K. et al., *Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques*, TURCOMAT, vol. 12, no. 2, pp. 912-917, Apr. 2021.

[5] S. Deebika, K. A. Indira and Jesline, *A machine learning based music player by detecting emotions*, 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), pp. 196-200, 2019