CS 584: MACHINE LEARNING

Project Proposal

MUSIC RECOMMENDATION SYSTEM USING EMOTION RECOGNITION

Author:
Dikshitha Varman – A20521565
Jayasurya R – A20513480
Kiran Gopi – A20520561

Supervisor:
Professor Yan Yan

1 INTRODUCTION

1.1 ABSTRACT:

Managing a large music collection can be challenging for listeners who have to manually create and organize playlists with hundreds of songs. Keeping track of unused songs can also be difficult and time-consuming, leading to wasted device memory. Additionally, selecting songs based on mood and preference can be a hassle for users who have to do so manually each time they want to listen to music. Furthermore, changing preferences and play-styles can make it difficult to reorganize and play music. To address these issues, we have implemented a machine learning approach that utilizes facial scanning and feature tracking to detect the user's mood and generate a personalized playlist based on it.

1.2 PROBLEM DESCRIPTION:

The suggested system utilizes facial expression detection to extract facial landmarks from the user's facial expressions, which are then classified to determine the user's emotion. Once the user's emotion is determined, songs matching that emotion are displayed to the user, potentially reducing their stress levels and helping them choose music more efficiently. The proposed architecture consists of three modules: Emotion extraction, Audio extraction, and Emotion-Audio extraction. To ensure accurate results, the image fed into the classifier should be captured in a well-lit environment and have a minimum quality of 320p. The system eliminates the need for users to search or browse for songs, saving them time and effort.

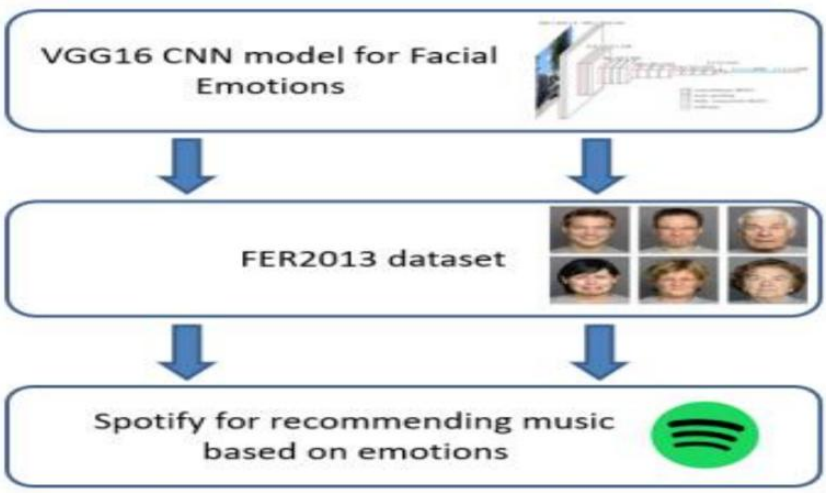

Fig 1.1: Proposed System

1.3 PROPOSED MODEL:

Our approach utilizes Deep Neural Networks (DNN) to overcome the limitations of handcrafted features by directly learning the most relevant feature abstractions from uncontrolled environment data. DNNs have recently emerged as a successful method for

recognizing visual objects, human poses, face verification, and many other tasks. With access to powerful computers and vast databases, DNNs can extract highly discriminative features from data samples. CNNs, in particular, have proven to be highly effective in tasks such as image recognition and classification by reducing the number of parameters without sacrificing model quality. Our proposed method employs a CNN model to recognize individual facial expressions and identify the user's emotion. The system then plays the song that best matches the user's emotions.

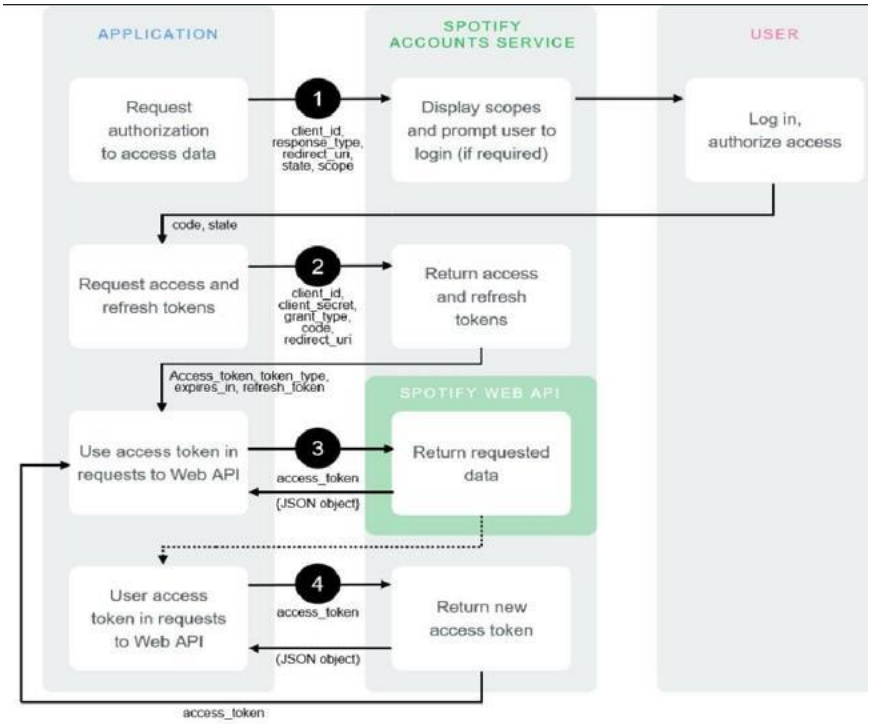

Fig 1.2: System Architecture

2 Dataset Description:

Facial Expression Recognition 2013 (FER2013) dataset is used by our model which is a publicly available dataset for training and evaluating facial expression recognition algorithms. The FER2013 dataset consists of 35,887 grayscale images of faces that are 48x48 pixels in size. The images are labeled with one of seven different facial expressions: angry, disgust, fear, happy, sad, surprise, and neutral. Each expression is represented by approximately 5,000 images in the dataset. The dataset is split into three subsets: a training set, a validation set, and a test set. The training set consists of 28,709 images, the validation set consists of 3,589 images, and the test set consists of 3,589 images.

In the FER2013 dataset, the numerical attribute is the pixel values of each image, which are represented as 48x48 grayscale values ranging from 0 to 255.

The categorical attribute in the FER2013 dataset is the facial expression label, which consists of seven categories: angry, disgust, fear, happy, sad, surprise, and neutral. Each image in the dataset is labeled with one of these categories.

3 Work done so far:

1. **Emotion detection:** The system has an "emotion function" that detects emotions using a pre-trained VGG16 model and a HaarCascade classifier from OpenCV. It takes an image as input, processes it, and applies the classifier. If a face is found, the model makes a prediction and returns an emotion label.

2. **Spotify API:** The system uses the Spotify API to access information about tracks, devices, and audio features. It uses two functions, "requestAuthorization()" and "callAuthorizationApi()", to obtain an access token for authentication purposes. The "callApi()" function is used to send requests to the Spotify accounts service for various types of information.

3. **Flask framework:** The system's server is built on the Flask framework. It has two routes, "home" and "emotion". The "home" route returns the application webpage, and the "emotion" route receives an image and returns an emotion label. The "emotion" route uses the "emotion function" for emotion detection.

4. **VGG16 model:** The VGG16 model used for emotion detection was trained for over 400 epochs with 478 images in each batch. Various hyperparameters were tuned throughout the course of training. The model achieved an accuracy of around 98% on the train set and around 70% on the test set.

5. **Image processing:** The input image is pre-processed by the "emotion function" before applying the HaarCascade classifier. If no face is found, a "nil" label is returned. If a face is found, the "emotion function" sends the last face found to the VGG16 model for emotion detection. The predictions are analyzed and an emotion label is returned.

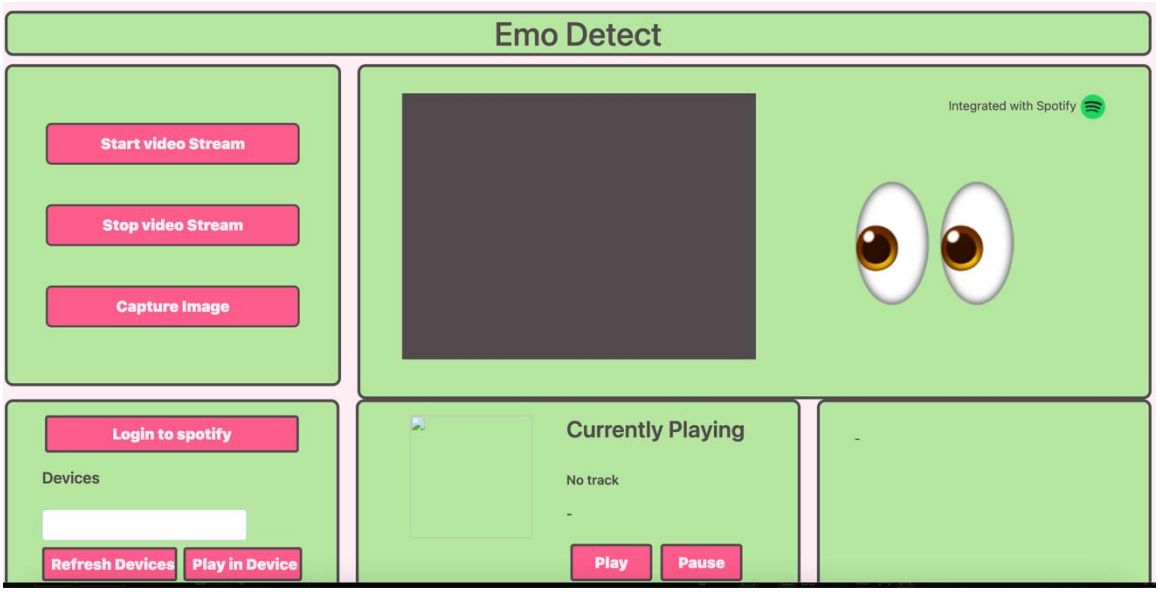

FIG 3.1 Homepage of our application

Below are some screenshots taken from the application where the user must start the video stream and capture the image. The emotion is detected, and songs will be played from the Spotify playlist.

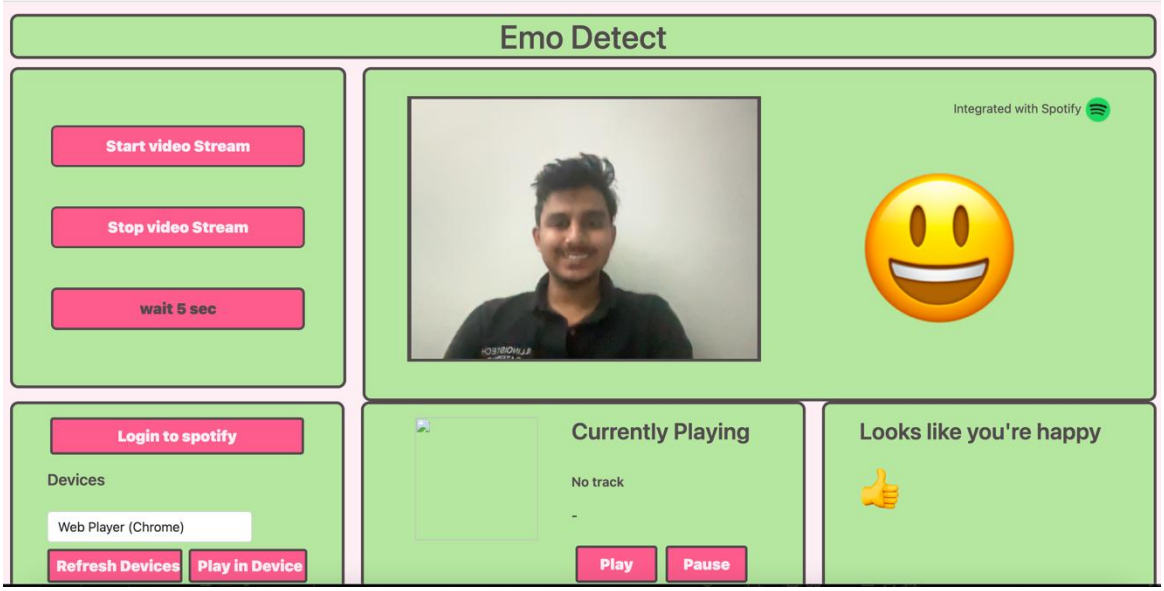

FIG 3.2 Emotion Detected: Happy

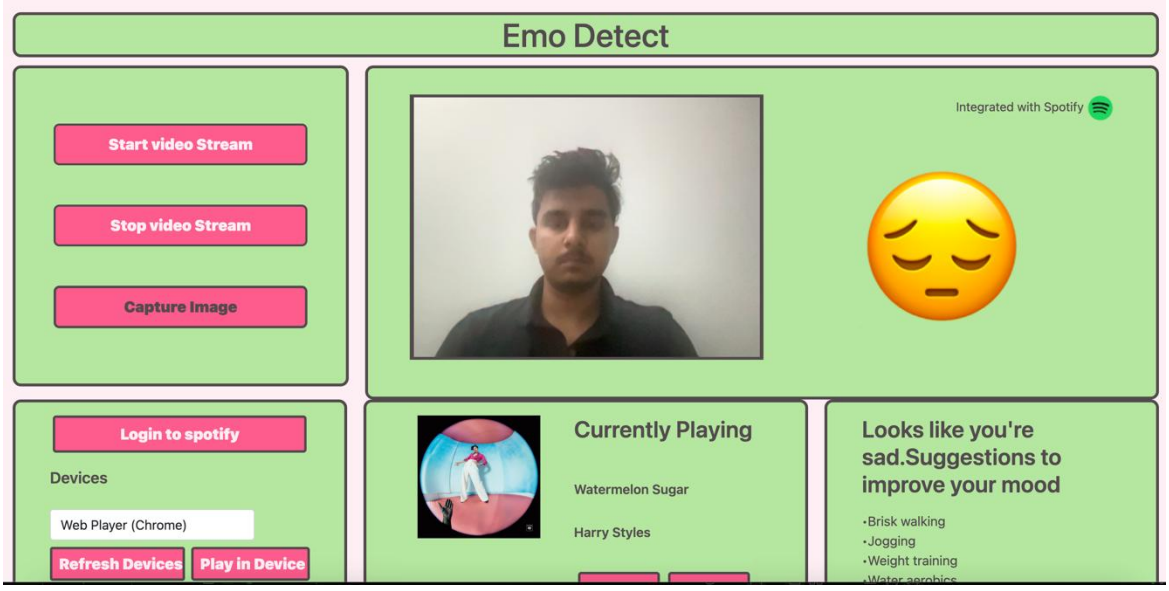

FIG 3.3 Emotion Detected: Sad

4 FUTURE SYSTEM IMPLEMENTATION:

The application aims to provide a simpler, additional hardware-free and reliable emotion-based music system to the operating system users. It could help to reduce the search time for music and thus reduce the unnecessary computational time and thus increase the overall accuracy and efficiency of the system.

- More work needs to be done to detect emotions like surprise, disgust, anger.
- Multiple face detection must be handled.
- More songs need to be analyzed and added to different emotion playlist as we have kept 10 songs for each playlist
- Currently our application plays music in the Spotify web player and we are working on to play the music in the primary application itself where we have provided with the "Play in Device" button in the UI.

5 References:

- *Metilda Florence S and Uma M, 2020, "Emotional Detection and Music Recommendation System based on User Facial Expression", IOP Conf. Ser.: Mater. Sci. Eng. 912,06/2007.*
- *EMOTION BASED MUSIC RECOMMENDATION SYSTEM H. Immanuel James, J. James Anto Arnold, J. Maria Masilla Ruban, M. Tamilarasan, R. Saranya IRJET (2019)*
- *Emophony – Face Emotion Based Music Player Banpreet Singh Chhabra – IRJET (JUNE 2020)*
- *Seungjae Lee, Jung Hyun Kim, Sung Min Kim, & Won Young Yoo. (2011). Smoodi: Mood-based music recommendation player, 2011 IEEE International Conference on Multimedia and Expo.*