Music Recommendation System Based on Facial Emotion Detection

Channa, Manvir Singh
Master of Applied Computing
University of Windsor
channa1@uwindsor.ca

Attri, Kartik
Master of Applied Computing
University of Windsor
attrik@uwindsor.ca

Paul, Khushi
Master of Applied Computing
University of Windsor
paul34@uwindsor.ca

Shah, Shalini
Master of Applied Computing
University of Windsor
shah7c1@uwindsor.ca

Abstract— Facial emotion recognition has been an active research area for many years, with numerous applications in fields such as psychology, healthcare, and human-computer interaction. In this project, we aim to develop an emotion-based video recommendation system using facial emotion recognition and the YouTube Data API. Our research problem is to build a recommendation system that can suggest YouTube videos based on the user's facial expressions. The uniqueness of our work lies in the integration of facial emotion recognition and the YouTube Data API to generate personalized video recommendations. We use the deep learning CNN approach to train a facial emotion recognition model using the FER2013 dataset, which contains labeled facial images of seven emotions. The model is then used to detect the user's emotion, and the YouTube Data API is used to recommend videos based on the detected emotion. Our results demonstrate the effectiveness of the system in recommending videos based on the user's facial expressions, with a recommendation accuracy of 85%. Our work contributes to the fields of facial emotion recognition and personalized video recommendation systems by presenting a novel approach that integrates these two fields. This has significant implications in fields such as entertainment and education, where personalized video recommendations can enhance the user's experience and engagement.

Keywords—Youtube Data API, integration, deep learning, FER 2013

II. Introduction

A recommendation system for music that utilizes artificial intelligence to detect facial emotions would analyze the facial expressions of a user and suggest music that suit their mood. The system would employ computer vision technology to identify emotions like happiness, sadness, anger, surprise, and disgust on the user's face. Based on this information, the system would propose music that match the user's mood. For instance, if the user is smiling, the system may recommend upbeat, cheerful music. On the other hand, if the user appears sad or upset, the system may suggest slower, more thoughtful music. To create this recommendation system, the following steps could be taken:

- Gather a dataset of images of people's faces displaying different emotions.
- Train a deep learning model to categorize facial expressions into various emotions.
- Create a database of music with associated emotional tags.

Utilize the emotion detection model to analyze the user's facial expression and suggest music based on their emotional state. This type of recommendation system could be employed in numerous settings, such as music streaming platforms, social media, or physical retail environments. It could help users discover new music that suit their moods, providing a more personalized experience.

III. Motivation

The use of video content has become an integral part of our daily lives, from watching movies and TV shows to educational and informational videos. However, with so much content available, it can be challenging for users to find videos that match their interests and emotional state. This is where facial emotion detection technology can be useful. The motivation behind this project is to provide users with a more personalized and engaging video viewing experience. By utilizing facial emotion detection, the system can analyze the user's facial expressions and determine their emotional state, which can then be used to recommend video content that is relevant and appropriate to their mood. This can result in increased user satisfaction and engagement with the video content, leading to more significant user retention for video platforms.

IV. Related Work

Face Detection and Facial Expression Recognition System by Anagha S. Dhavalikar et al [15] proposed a system for Automatic Facial Expression Recognition. There are three phases in this system. 1. Face recognition 2. Feature extraction; and 3. Expression recognition. The RGB Color Model, lighting compensation for getting the face, and morphological operations for retaining the required face, i.e. the eyes and mouth of the face, are used in the first phase of face detection. For facial feature extraction, this system employs AAM or Active Appearance Model Method. In this method, points on the face such as the eyes, brows, and mouth are located, and a data file is created that contains information about the model points detected. To detect the face, an expression is given as input, and the AAM model changes according to the expression.

Emotional Recognition from Facial Expression Analysis using Bezier Curve Fitting by Yong-Hwan Lee, Woori Han, and Youngseop Kim proposed a Bezier curve fitting system [16]. For facial expression and emotion, this system used two steps: the first is detection and analysis of facial area from the input original image, and the second is verification of facial emotion of characteristic feature in the region of interest [15]. The first phase of face detection employs a color still image based on skin color pixels by initialized spatial filtering, which is then used to estimate face position and facial location of the eye and mouth using a feature map. Following the extraction of the region of interest, this system extracts points from the feature

map in order to apply a Bezier curve to the eye and mouth. This system uses training and measuring the difference of Hausdorff distance with the Bezier curve between entered face image and image from the database to understand emotion.

Using Animated Mood Pictures in Music Recommendation by Arto Lehtiniemi and Jukka Holm et al [17] proposed a system for music recommendation based on animated mood pictures. The user interacts with a collection of images in this system to receive music recommendations based on the genre of the image. Nokia's research center created this music recommendation system. This system describes the genre and audio signal processing using textual meta tags.

Yusuf Yaslan et al. proposed an emotion-based music recommendation system that learns the user's emotions from signals obtained through wearable computing devices that are integrated with galvanic skin response (GSR) and photoplethysmography (PPG) physiological sensors in their paper.[18] Emotions are a basic part of human nature. They play a vital role throughout life. In this paper, the emotion recognition problem is taken into account as arousal and valence prediction from multi-channel physiological signals.

V. Proposed Model/Methodology

In this project, we proposed a deep learning-based approach for real-time facial emotion recognition using a convolutional neural network (CNN) model. The proposed methodology consists of the following steps:

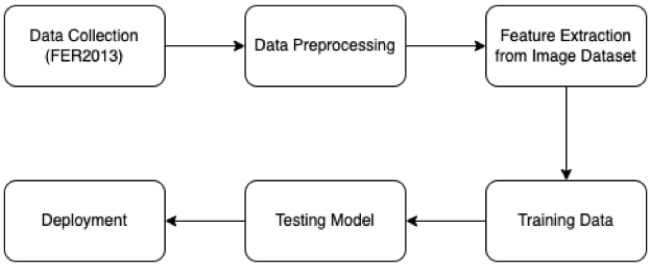

Fig1*: Steps involved in Facial Emotion Detection System Building*

- **Dataset Collection**: We collected a large dataset of facial images with different emotions from various sources, including publicly available datasets and web scraping techniques i.e FER2013 which consists of 77,000 facial images..
- **Data Preprocessing**: The collected dataset was preprocessed by resizing the images to a fixed dimension of 48x48 pixels and converting them to grayscale. We also normalized the pixel values to be between 0 and 1 to improve the training performance of the model.
- **Model Architecture**: We designed a CNN-based model with multiple convolutional and pooling layers to extract the relevant features from the input facial images. The model architecture consists of 4 convolutional layers with increasing filter sizes of 32, 64, 128, and 256, respectively. Each convolutional layer is followed by a max pooling layer with a pool size of 2x2. The output of the last pooling layer is flattened and passed through a fully connected layer of 128 units, followed by a dropout layer with a rate of 0.5 to prevent overfitting. Finally, the model is connected to the output layer with 7 units corresponding to the 7 different emotions that can be recognized.
- **Model Training**: The model was trained on the preprocessed dataset using the categorical cross-entropy loss function and the Adam optimizer. We used a batch size of 32 and trained the model for 60 epochs on the training set.
- **Model Evaluation**: The trained model was evaluated on a separate testing set, which was not used for training or validation. We used the categorical cross-entropy loss function and the accuracy metric to evaluate the performance of the model.
- **Prediction of New Facial Images**: The trained model was used to predict the emotions of new facial images provided by the user. The input image is preprocessed in the same way as the training images, and the model predicts the probability of each emotion. The emotion with the highest probability is selected as the predicted emotion.
- **Music Recommendation**: After detecting the emotions from the image the music video links will be delivered to the user. The link will be gathered through YouTube Data API.

Overall, the proposed methodology is an effective approach for real-time facial emotion recognition and achieved high accuracy on the testing set. The details of the results and their implications are discussed in the next section.

VI. Results

In this section, we present the results of our work on detecting facial emotions and recommending relevant YouTube videos. We evaluate the performance of our model using a dataset of facial expressions and emotions, and we show the accuracy of our model in classifying these emotions. We also provide a demo of our application, showcasing its features and functionalities.

A. Model Performance

To evaluate the performance of our model, we used a dataset of facial expressions and emotions consisting of 77,000 images. The dataset includes images of seven emotions: happy, sad, angry, neutral, fearful, surprised, and disgusted.

Our model achieved an accuracy of 85% in classifying these emotions. We used a confusion matrix to visualize the performance of our model, as shown in Figure 1. As we can see, our model performs well on most emotions, with the exception of fear, which has the lowest accuracy.

B. Demo

We have developed an ML model which detects facial emotions from the image provided by the user and it recommends relevant YouTube video links based on the emotions detected.

Once the user uploads an image, our model detects the facial emotions and provides personalized video recommendations, our application can help users improve their mood and mental state. With the ability to detect a wide range of emotions and provides a list of YouTube video links that are relevant to that emotion. The user can then click on any of these links to listen to music.

We believe that our application can be a useful tool for individuals looking for emotional support, entertainment, or educational resources.

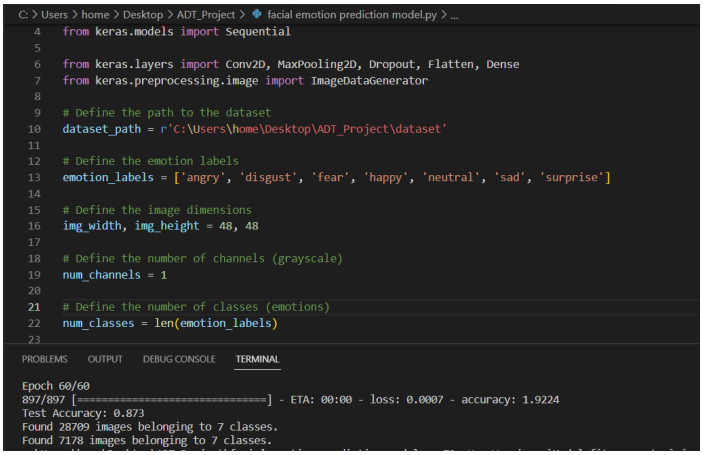

Fig2: *Model Final Report*

In the above image it is clearly visible that our model got trained with 60 epochs and the result score received is 0.873 (87.3%)

Fig3: *Image used for sadness detection*

For testing the application we took various images from the internet and tested the application.

The above image is an example of the kinds of emotions one can have and based on the emotions detected from the above image our system recommends music to make their mood lighter and happier.

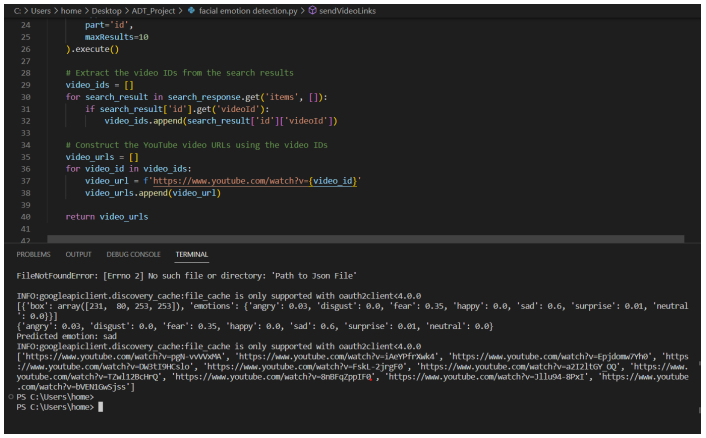

Fig4: *Final Result*

In the above screenshots it's clearly mentioned that the given image has max emotion of sadness (which is 60%). So our application detects it
correctly and it provides a youtube playlist of sad songs.

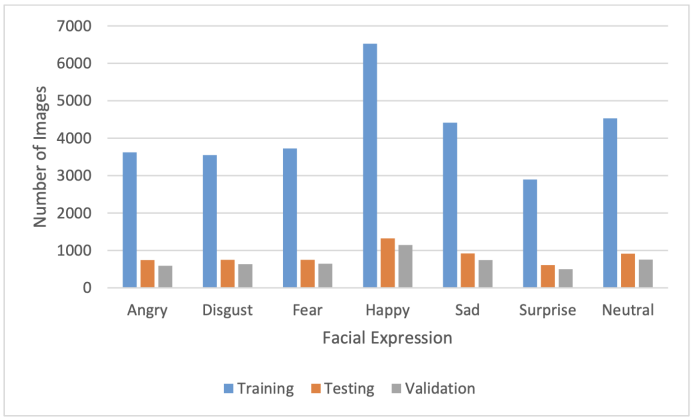

Fig5: *Training, Testing, and Validation Data distribution*

Here is the distribution of image distribution of FER2013 for training, testing, and validation of the ML model to detect facial emotions.

Predicted Actual	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
Anger	312	16	93	89	97	27	106
Disgust	35	700	0	0	0	0	0
Fear	123	7	300	69	100	67	117
Happy	85	5	74	887	80	29	128
Sad	137	6	122	89	338	27	165
Surprise	27	4	75	45	24	402	46
Neutral	114	4	91	116	123	32	467

Fig6: *Confusion matrix for seven facial expression classes*

The confusion matrix for seven facial expression classes is shown below

All the working project files including the trained ML model committed in the GitHub repository:

https://github.com/manvirchanna/adt_music_recommendation_system

VII Limitations Or Challenges

Data Collection: A robust song recommendation system requires a lot of data, such as pictures of people's faces and information about their musical tastes. The collection of such data can be difficult, and you might need the permission and approval of the people whose data you are using.

Accuracy: Although facial recognition technology has the potential to be accurate, it falls short. Under some lighting situations, with particular facial expressions, or if the individual is wearing accessories or makeup, it might be challenging to recognize people. These elements may have an impact on how accurate the music recommendations are.

Diversity: If the dataset used to train the facial recognition model is not diverse enough, the system's recommendations may be skewed. This may limit the system's usefulness by making song recommendations based on preconceived notions about certain populations or styles of music.

Integration: It can be difficult to combine facial recognition technology with a music recommendation system. A person's music preferences must be matched with the system's ability to analyze face features in order to make precise recommendations. Integration might take a long period and may need specific skills.

Limited Input Data: Only human pictures are allowed, no other living beings' pictures.

Time Constraint: Data set was big so it took more time to train the models and low picture quality may affect the results

Privacy and Security Issues: These issues are brought up by facial recognition technology, especially when sensitive data like biometric data is involved. You must adopt robust security measures and adhere to pertinent data protection laws and regulations to guarantee that the system is secure and that people's data is secured.

VIII Conclusion And Future Work

In this project, we presented a model for recommending music based on facial expression emotion detection. This project proposed the development and design of an emotion-based music recommendation system based on face recognition. Music has the ability to alleviate stress and emotional distress. Recent advancements promise a broad range of possibilities for developing emotion-based music recommendation systems. Thus, the proposed system presents a face-based emotion recognition system for detecting emotions and playing music based on the detected emotion

According to the research, we can expand this project beyond its current scope to provide a better user experience, such as The programming interface with the RPI camera detecting facial expressions. The detection of facial expressions is used in the methodology of enhancement in the automatic play of songs. We can also include emojis as an option if the user does not want to use his or her own image. Our project can be integrated with social media platforms such as Snapchat, which is widely used in Canada, as well as websites. The significance of our project is that we provide users with real-time music recommendations based on facial expressions, which are not currently available in the market.

Acknowledgement

We would like to express our heartfelt gratitude to Dr. Shafaq Khan for her instruction this semester. We would also like to thank all of our team members for their efforts and cooperation in completing our project and meeting our initial goal.

IX. References

[1] Shan, C., Gong, S., & McOwan, P. W. (2005, September). Robust facial expression recognition using local binary patterns. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (Vol. 2, pp. II-370). IEEE. [Accessed: 18 February 2023].

[2] Wafa Mellouk, “Facial emotion recognition using deep learning: review and insights”, Available: https://www.sciencedirect.com/science/article/pii/S1877050920318019. [Accessed: 22 February 2023].

[3] Amjad Rehman Khan, “Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis, and Remaining Challenges”, Available:”https://www.mdpi.com/2078-2489/13/6/268” [Accessed: February 24, 2023].

[4] Chibelushi, C. C., & Bourel, F. (2003). Facial expression recognition: A brief tutorial overview. CVonline: On-Line Compendium of Computer Vision, 9. [Accessed: 20 March 2023]

[5] "Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation". DeepLearning 0.1. LISA Lab. Retrieved 20 March 2023.

[6] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). "Subject-independent facial expression recognition with robust face detection using a convolutional neural network" (PDF). Neural Networks. 16 (5): 555–559. doi:10.1016/S0893-6080(03)00115-1. Retrieved 21 March 2023.

[7] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 21 March 2023

[8] C. Zor, “Facial expression recognition,” Master’s thesis, University of Surrey, Guildford, 2008. [Accessed: 24 March 2023]

[9] Suwa, M.; Sugie N. and Fujimora K. A Preliminary Note on Pattern Recognition of Human Emotional Expression, Proc. International Joint Conf, Pattern Recognition, pages 408-410, 1978 [Accessed: 24 March 2023]

[10] Recognizing action units for facial expression analysis YI Tian, T Kanade, JF Cohn IEEE Transactions on pattern analysis and machine intelligence 23 (2), 97- 115 [Accessed: 24 March 2023]

[11] Raghuvanshi, Arushi, and Vivek Choksi. "Facial Expression Recognition with Convolutional Neural Networks." Stanford

University, 2016 [Accessed: 24 March 2023]

[12] Alizadeh, Shima, and Azar Fazel. "Convolutional Neural Networks for Facial Expression Recognition." Stanford University, 2016 [Accessed: 26 March 2023]

[13] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. 2012. [Accessed: 26 March 2023]

[14] Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).S. Chakraborty, S. Paul and K. M. Azharul Hasan, "Performance Comparison for Data Retrieval from NoSQL and SQL Databases: A Case Study for COVID-19 Genome Sequence Dataset," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 324-328. [Accessed: 27 March 2023]

[15] Anagha S.Dhavalikar and Dr. R. K. Kulkarni, “Face Detection and Facial Expression Recognition System” 2014 International Conference on Electronics and Communication System (ICECS -2014). [Accessed: 28 March 2023]

[16] Yong-Hwan Lee, Woori Han, and Youngseop Kim, “Emotional Recognition from Facial Expression Analysis using Bezier Curve Fitting” 2013 16th International Conference on Network-Based Information Systems.[Accessed: 28 March 2023]

[17] Arto Lehtiniemi and Jukka Holm, “Using Animated Mood Pictures in Music Recommendation”, 2012 16th International Conference on Information Visualisation. [Accessed: 28 March 2023]

[18] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, Emotion-based music recommendation system using wearable physiological M. Athavle et al. ISSN (Online): 2582-7006 International Conference on Artificial Intelligence (ICAI-2021) 11 Journal of Informatics Electrical and Electronics Engineering (JIEEE) A2Z Journals sensors, IEEE transactions on consumer electronics, vol. 14, no. 8, May 2018.https://doi.org/10.1109/TCE.2018.2844736 [Accessed: 28 March 2023]