RES MILITARIS Social Science Journal

EMOTION DETECTION MODEL BASED MUSIC RECOMMENDATION SYSTEM

Dr. S. Kavitha[1], Katkuri Arun[2], Kalakota Praveen[2], Chelimala Laxma Reddy[2]

[2]UG Scholar, [1,2]Department of Computer Science and Engineering

[1,2]Kommuri Pratap Reddy Institute of Technology, Ghatkesar, Hyderabad, Telangana.

ABSTRACT

Over the recent years much advancement is made in terms of artificial intelligence, machine learning, human-machine interaction etc. Voice interaction with the machine or giving command to it to perform a specific task is increasingly popular. Many consumer electronics are integrated with SIRI, Alexa, Cortana, Google assist etc. But machines have limitation that they cannot interact with a person like a human conversational partner. It cannot recognize Human Emotion and react to them. Emotion Recognition from speech is a cutting-edge research topic in the Human machines Interaction field. There is a demand to design a more rugged man-machine communication system, as machines are indispensable to our lives. Many researchers are working currently on speech emotion recognition (SER) to improve the man machines interaction. To achieve this goal, a computer should be able to recognize emotional states and react to them in the same way as we humans do. The effectiveness of the speech emotion recognition (SER) system depends on quality of extracted features and the type of classifiers used. In this work we tried to identify four basic emotions: anger, sadness, neutral, happiness from speech. Here this work used audio file of short speech taken from movies as training and testing dataset. This work uses CNN to identify different emotions using MFCC (Mel Frequency Cepstral Coefficient) as features extraction technique from speech.

Keywords: CNN MFCC, Speech Recognition, Audio files, Communication, Machine Learning, Audio Processing

1. INTRODUCTION

Computing is the study or practice of inventing, designing, building or using body-worn computational and sensory devices that leverages a new type of human-computer interaction with a body-attached component that is always up and running. As the number of wearable computing device users are growing every year, their areas of utilization are also rapidly increasing. They have influenced medical care, fitness, aging, disabilities, education, transportation, finance, gaming, and music industries [1], [2]. Recommendation engines are algorithms which aim to provide the most relevant items to the user by filtering useful information from a huge pool of data. Recommendation engines may discover data patterns in the data set by learning user's choices and produce the outcomes that co-relates to their needs and interests [3]. Most of the recommender systems do not consider human emotions or expressions. However, emotions have noticeable influence on daily life of people. For a rich set of applications including human-robot interaction, computer aided tutoring, emotion aware interactive games, neuro marketing, socially intelligent software apps, computers should consider the emotions of their human conversation partners. Speech analytics and facial expressions have been used for emotion detection. However, in case of human beings prefer to camouflage their expressions, using only speech signals or facial expression signals may not be enough to detect emotions reliably. Compared with facial expressions, using physiological signals is a more reliable method to track and recognize emotions and internal cognitive processes of people. The quest to imbue machines with the ability to understand human emotions has been a long and intricate journey, marked by significant strides in artificial intelligence (AI) and machine learning (ML). Over the past few years, the landscape of human-machine interaction has been reshaped dramatically, witnessing the integration of virtual assistants like Siri,

Alexa, Cortana, and Google Assistant into various consumer electronics. While these developments have undoubtedly enhanced our interactions with machines, there remains a critical limitation: the inability of machines to engage with users as empathetic conversational partners. Machines, as they stand, lack the capacity to discern human emotions and respond to them accordingly. This deficiency has spurred intense interest and research into emotion recognition, particularly from speech, within the realm of human-machine interaction. In the realm of entertainment and content recommendation, SER holds immense potential for enhancing user engagement and satisfaction. By leveraging emotion-aware algorithms, streaming platforms can dynamically adjust content recommendations based on the user's emotional state, delivering personalized experiences tailored to individual preferences and mood.

2. LITERATURE SURVEY

Yang et. al [4] offers new insights on music emotion recognition methods based on different combinations of data features that they use during the modelling phase from three aspects, music features only, ground-truth data only, and their combination, and provides a comprehensive review of them. Then, focusing on the relatively popular methods in which the two types of data, music features and ground-truth data, are combined, they further subdivide the methods in the literature according to the label- and numerical-type ground-truth data, and analyse the development of music emotion recognition with the cue of modelling methods and time sequence. Three current important research directions are then summarized. Although much has been achieved in the area of music emotion recognition, many issues remain.

Domínguez-Jiménez et. al [5] proposes a model for recognition of three emotions: amusement, sadness, and neutral from physiological signals with the purpose of developing a reliable methodology for emotion recognition using wearable devices. Target emotions were elicited in 37 volunteers using video clips while two biosignals were recorded: photoplethysmography, which provides information about heart rate, and galvanic skin response. These signals were analyzed in frequency and time domains to obtain a set of features. Several feature selection techniques and classifiers were evaluated. The best model was obtained with random forest recursive feature elimination, for feature selection, and a support vector machine for classification. The results show that it is possible to detect amusement, sadness, and neutral emotions using only galvanic skin response features. The system was able to recognize the three target emotions with accuracy up to 100% when evaluated on the test data set.

Ivana Andjelkovic et. al [6] presented and evaluated MoodPlay –a hybrid recommender system for musical artists which introduces a novel interactive visualization of moods and artists. The system supports explanation and control of a recommender system via manipulation of an avatar within the visualization. Design and implementation of an online experiment (N=279) was presented to evaluate the system through four conditions with varying degrees of visualization, interaction and control.

Jianhua Zhang et. al [7] the emotion recognition methods based on multi-channel EEG signals as well as multi-modal physiological signals are reviewed. According to the standard pipeline for emotion recognition, they review different feature extraction (e.g., wavelet transform and nonlinear dynamics), feature reduction, and ML classifier design methods (e.g., k-nearest neighbor (KNN), naive Bayesian (NB), support vector machine (SVM) and random forest (RF)). Furthermore, the EEG rhythms that are highly correlated with emotions are analyzed and the correlation between different brain areas and emotions is discussed. Finally, they compare different ML and deep learning algorithms for emotion recognition and suggest several open problems and future research directions in this exciting and fast-growing area of AI.

Yu Liu et. al [8] proposed an end-to-end MLF-CapsNet framework for multi-channel EEG emotion recognition. This proposed framework can identify the intrinsic relationship among various EEG

channels well. They combine multi-level features extracted from different convolution layers to form primary capsules. Besides, they add bottleneck layer to reduce the number of parameters and accelerate the speed of calculation. Finally, experiments on DEAP dataset and DREAMER dataset are conducted.

Yongqiang Yin et. al [9] proposed a new emotion recognition method using deep learning model based on EEG's differential entropy, which adopts a novel fusion model of GCNN and LSTM for emotion classification. ECLGCNN utilizes the graph and temporal information, where each EEG channel corresponds to a graph node, and the functional relationship between two channels corresponds to the edge of the graph and LSTM cells' gates are used to extract effective information. Shamim Hossain et. al [10] proposes an emotion recognition system using a deep learning approach from emotional Big Data. The Big Data comprises of speech and video. In the proposed system, a speech signal is first processed in the frequency domain to obtain a Mel-spectrogram, which can be treated as an image. Then this Mel-spectrogram is fed to a convolutional neural network (CNN). For video signals, some representative frames from a video segment are extracted and fed to the CNN. The outputs of the two CNNs are fused using two consecutive extreme learning machines (ELMs). The output of the fusion is given to a support vector machine (SVM) for final classification of the emotions. The proposed system is evaluated using two audio–visual emotional databases, one of which is Big Data. Experimental results confirm the effectiveness of the proposed system involving the CNNs and the ELMs.

Santamaria-Granados et. al [11] applies the deep learning approach using a deep convolutional neural network on a dataset of physiological signals (electrocardiogram and galvanic skin response), in this case, the AMIGOS dataset. The detection of emotions is done by correlating these physiological signals with the data of arousal and valence of this dataset, to classify the affective state of a person. In addition, an application for emotion recognition based on classic machine learning algorithms is proposed to extract the features of physiological signals in the domain of time, frequency, and non-linear. This application uses a convolutional neural network for the automatic feature extraction of the physiological signals, and through fully connected network layers, the emotion prediction is made. The experimental results on the AMIGOS dataset show that the method proposed in this paper achieves a better precision of the classification of the emotional states, in comparison with the originally obtained by the authors of this dataset. Rosa et. al [12] presents a knowledge-based recommendation system (KBRS), which includes an emotional health monitoring system to detect users with potential psychological disturbances, specifically, depression and stress. Depending on the monitoring results, the KBRS, based on ontologies and sentiment analysis, is activated to send happy, calm, relaxing, or motivational messages to users with psychological disturbances. Also, the solution includes a mechanism to send warning messages to authorized persons, in case a depression disturbance is detected by the monitoring system. The detection of sentences with depressive and stressful content is performed through a convolutional neural network and a bidirectional long short-term memory - recurrent neural networks (RNN); the proposed method reached an accuracy of 0.89 and 0.90 to detect depressed and stressed users, respectively.

3. PROPOSED SYSTEM

This research implements facial emotion analysis for personalized music recommendations which combines computer vision, deep learning, and user interface design to create an application that analyses facial expressions in images and recommends music based on the detected emotion.

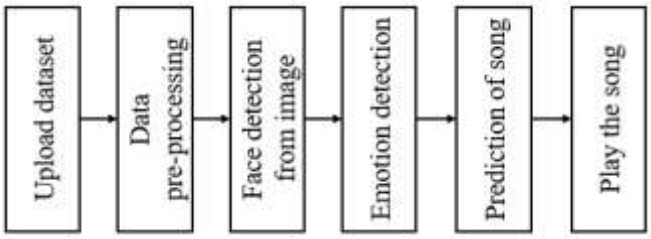

Fig.1: Block diagram of proposed system.

Step 1: User Interface (GUI)

— The project uses the Tkinter library to create a graphical user interface (GUI). The GUI provides a user-friendly way for users to interact with the application.

Step 2: Image Upload and Preprocessing

— Users can upload an image containing one or more human faces.

— The uploaded image is preprocessed, including resizing and face detection using a Haar Cascade classifier.

— The number of detected faces is displayed to the user.

Ste 3: Emotion Detection

— Once faces are detected, the project uses a pre-trained deep learning model (a neural network) to analyze the emotions expressed by the detected faces.

— The model assigns an emotion label to each detected face, such as "angry," "happy," "sad," etc.

— The detected emotion is displayed on the image.

Step 4: Music Recommendation

— Based on the detected emotion(s), the application recommends songs to the user.

— The songs are categorized by emotion, and the user can select a recommended song from a dropdown menu.

Step 5: Music Playback

— The user can play the selected song directly from the application.

— This functionality is achieved using the "playsound" library, which plays audio files.

Step 6: Project Components

— The project involves several components, including image processing (OpenCV), deep learning for emotion recognition (Keras), and GUI development (Tkinter).

— It also relies on pre-trained models for both face detection and emotion recognition.

Step 7: User Interaction

— Users can upload their own images, allowing for real-world applications where users might want to analyze their own facial expressions.

— The system responds to user actions, such as uploading an image, detecting emotions, and playing recommended songs.

Emotion Detection Model

This model is responsible for detecting the emotion expressed in the uploaded image that contains one or more human faces. This function integrates the Haar Cascade face detection and a pre-trained deep learning model for emotion recognition to identify the emotion expressed in the uploaded image's primary detected face. The detected emotion label is then displayed on the image for user visualization. Here's how it works:

- Face Selection: The function begins by checking if there are any faces detected in the uploaded image. If no faces are detected, it displays a message to the user indicating that no faces were found in the image.
- Emotion Detection: If one or more faces are detected, the function proceeds with emotion detection for each detected face. It selects the largest detected face by sorting the faces based on their size (the area of the bounding boxes) in descending order. The largest face is assumed to be the primary face of interest. Then, the coordinates of the selected face (x, y, width, height) are extracted.
- Region of Interest (ROI): A region of interest (ROI) is created by cropping the selected face from the original image. This ROI contains only the facial area to be analyzed for emotion.
- Preprocessing: The ROI is preprocessed before feeding it into the emotion recognition model. Then it is resized to a fixed size (48x48 pixels), which is a common input size for many emotion recognition models. Afterwards, pixel values in the ROI are normalized to be in the range [0, 1] by dividing by 255.0. Finally, the processed ROI is converted into a NumPy array.
- Emotion Prediction: The processed ROI is passed through the pre-trained Haar cascaded classifier model, and it produces a prediction in the form of a probability distribution over different emotions (e.g., angry, happy, sad, etc.).
- Displaying Emotion: The emotion with the highest predicted probability is selected as the detected emotion and the detected emotion label is added as text to the original image. Finally, the image with the detected emotion label is displayed to the user using OpenCV.

4. RESULTS AND DISCUSSION

4.1 Implementation description

This research implements a graphical user interface (GUI) application for facial emotion analysis and personalized music recommendations that allows users to upload an image with faces, detect the emotion in the faces, and recommend and play a song based on the detected emotion. It integrates image processing, deep learning, and audio playback in a user-friendly interface. Below is the step-by-step implementation:

1. Importing Libraries: Various libraries are imported, including Tkinter for GUI components, OpenCV for image processing, Keras for deep learning, NumPy for numerical operations, and others.

2. GUI Initialization: The Tkinter GUI window is created with the title "Facial Emotion Analysis for Personalized Music Recommendations" and a specified geometry (size).

3. Global Variables: Several global variables are declared to store information throughout the program, including filename (to store the path of the uploaded image), faces (to store detected faces), frame (to store image frames), value (to store song recommendations), and others.

4. Model Paths: Paths to the Haar Cascade classifier for face detection and a pre-trained emotion recognition model are defined.

5. Functions: Several functions are defined

— upload(): Opens a file dialog to allow the user to upload an image and displays the selected file's path.

— preprocess(): Preprocesses the uploaded image, resizes it, and detects faces using the Haar Cascade classifier. The number of detected faces is displayed.

— detectEmotion(): Detects the emotion in the uploaded image by cropping the face region, preprocessing it, and using the pre-trained emotion recognition model. The detected emotion label is displayed on the image.

— playSong(): Plays a song based on the detected emotion.

6. GUI Components: Various GUI components, such as labels, buttons, and a text box, are created using Tkinter and positioned on the window.

7. Event Handling: Buttons like "Upload Image With Face," "Preprocess & Detect Face in Image," "Detect Emotion," "Play Song," and the song selection dropdown menu are associated with specific functions to handle events.

8. Text Box: A text box is created to display information, such as the number of detected faces and the detected emotion.

9. GUI Configuration: The GUI components are configured with fonts, colors, and positions.

10. Main Loop: The main loop of the Tkinter GUI application is started with main.mainloop(), which keeps the application running and responsive to user interactions.

4.2 Results Description

Figure 2 shows the graphical user interface (GUI) application designed for the proposed Personalized Music Recommendation (PMR) system. The application utilizes facial emotion analysis to establish a connection between a person's emotional state and music recommendations. Figure 3 displays a collection of various images that were used as input for testing the proposed emotion analysis model. These images represent diverse facial expressions and emotions.

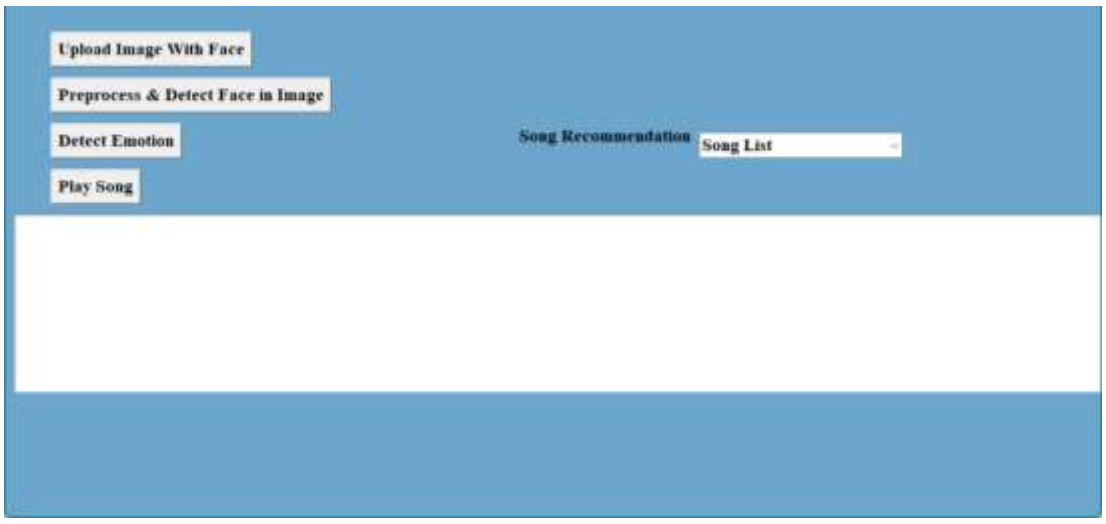

Figure 2: GUI application of proposed PMR using facial emotion analysis.

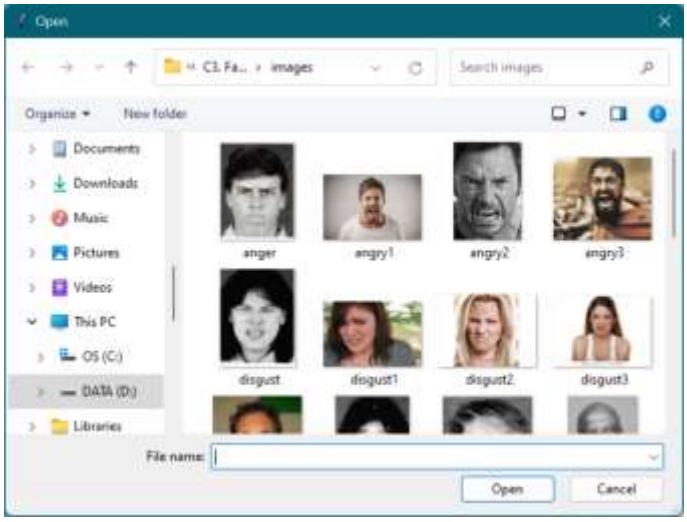

Figure 3: Different images used for testing the proposed model.

Figure 4 illustrates the preprocessing steps applied to a given test image. It depicts how the image is processed to detect and isolate the face region, which is crucial for accurate emotion ana lysis. Figure 5 showcases the outcome of emotion detection from a specific test image. It indicates the identified emotion label as a result of analyzing the person's facial expression.

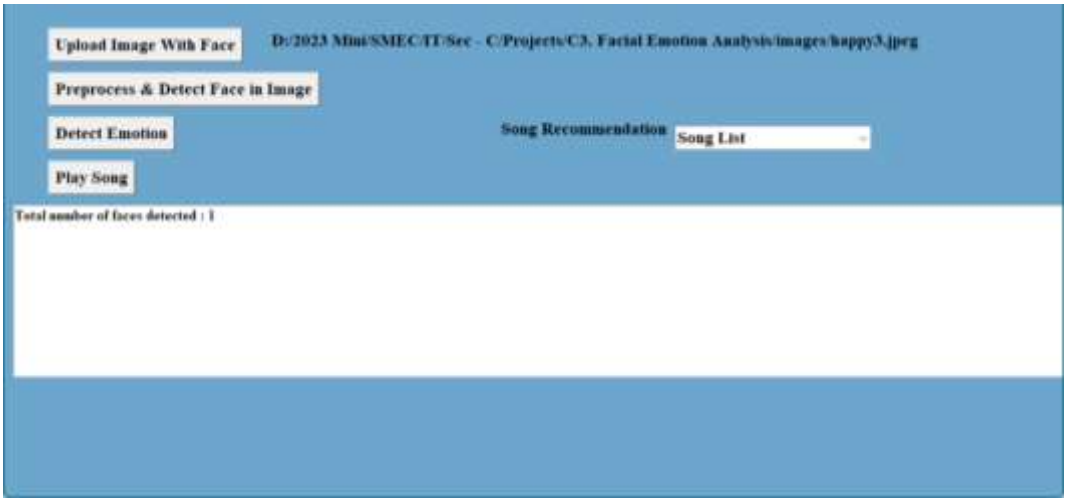

Figure 4: Preprocessing and detection of a face presented in given test image.

Figure 5: Detected emotion from a given test image.

In Figure 6, the detected emotion from a test image is determined to be happy. The figure suggests a song recommendation aligned with the detected emotion, enhancing the user experience. Similar to Figure 5, the Figure 7 depicts the emotion detected from another test image. It emphasizes the importance of emotion recognition in the proposed system.

Figure 6: Recommended happy song for the detected emotion as happy.

Figure 8 here, the detected emotion from a specific test image is identified as sad. The figure presents a song recommendation that corresponds to the detected sad emotion. Figure 9 demonstrates the process of uploading a test image that does not contain a human face. It emphasizes the system's capability to handle non-facial images and react appropriately. Based on a given test image, Figure 10 illustrates the situation where the proposed system fails to detect any human face. This showcases the limitations of the system when dealing with certain image types.

Figure 7: Detected emotion from a given test image.

Similar to Figure 9, this Figure 11 showcases the uploading process of an image that lacks a human face. It highlights the system's response when it encounters images incompatible with its face detection mechanism. Figure 12 reiterates the scenario where the system is unable to detect a human face in a different test image. It emphasizes the variability in the system's performance based on image characteristics.

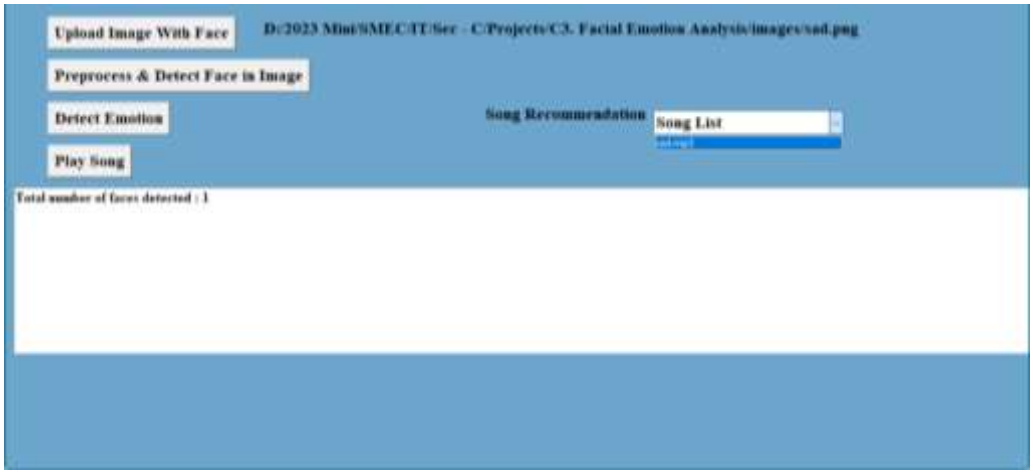

Figure 8: Recommended sad song as the sad emotion is detected from a given test image.

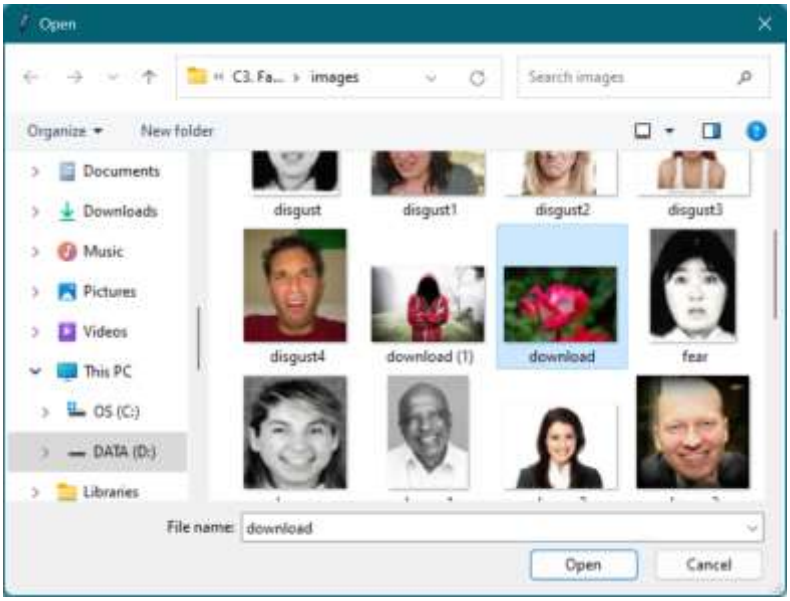

Figure 9: Uploading a flower image as a test image.

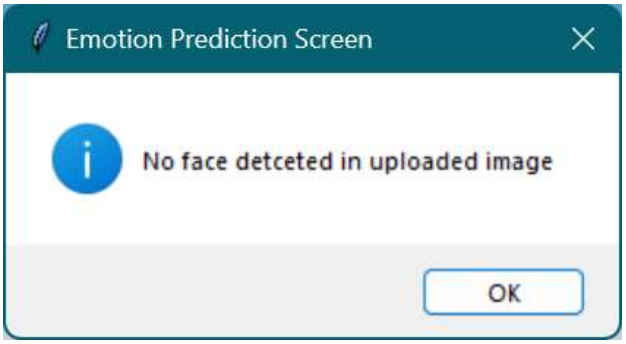

Figure 10: No face is detected in given test image.

Figure 11: Uploading an image without face as a test image.

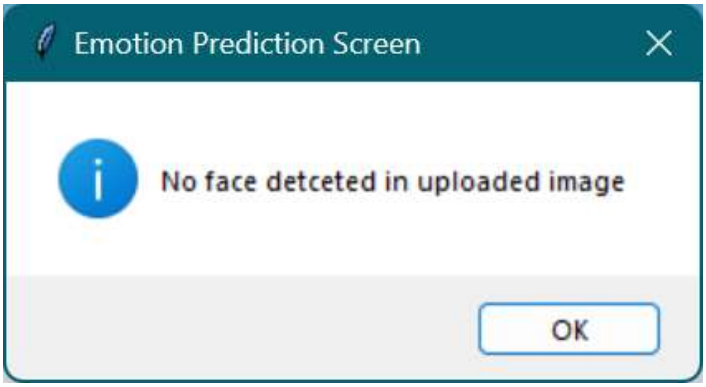

Figure 12: No face is detected in a test image.

5. CONCLUSION AND FUTURE SCOPE

The proposed work is a fascinating application that combines computer vision, deep learning, and user interface design to offer a unique user experience. It successfully implements facial emotion analysis, allowing users to upload images and receive real-time emotion detection results. This can be entertaining and informative for users interested in understanding their own or others' emotions. Additionally, the integration of personalized music recommendations based on detected emotions adds an engaging dimension to the project. Users can enjoy music that matches their emotional state. Further, the Tkinter-based GUI makes the application accessible to a wide range of users, even those without technical expertise. Users can easily upload images, view results, and play recommended songs. Overall, the use of a cascaded Haar Cascade classifier for face detection ensures that the application can quickly identify faces in uploaded images. It also leverages deep learning techniques with a pre-trained emotion recognition model to analyze facial expressions. This demonstrates the practical application of deep learning in real-world scenarios.

REFERENCES

[1] S. Jhajharia, S. Pal, and S. Verma, “Wearable computing and its application,” Int. J. Comp. Sci. and Inf. Tech., vol. 5, no. 4, pp. 5700–5704, 2014.

[2] K. Popat and P. Sharma, “Wearable computer applications: A feature perspective,” Int. J. Eng. and Innov. Tech., vol. 3, no. 1, 2013.

[3] P. Melville and V. Sindhwani, “Recommender systems,” in Encyc. Of mach. learn. Springer, 2011, pp. 829–838.

[4] Yang, X., Dong, Y. & Li, J. Review of data features-based music emotion recognition methods. Multimedia Systems 24, 365–389 (2018). https://doi.org/10.1007/s00530-017-0559-4

[5] J.A. Domínguez-Jiménez, K.C. Campo-Landines, J.C. Martínez-Santos, E.J. Delahoz, S.H. Contreras-Ortiz, A machine learning model for emotion recognition from physiological signals, Biomedical Signal Processing and Control, Volume 55, 2020, 101646, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2019.101646.

[6] Ivana Andjelkovic, Denis Parra, John O’Donovan, Moodplay: Interactive music recommendation based on Artists’ mood similarity, International Journal of Human-Computer Studies, Volume 121, 2019, Pages 142-159, ISSN 1071-5819, https://doi.org/10.1016/j.ijhcs.2018.04.004.

[7] Jianhua Zhang, Zhong Yin, Peng Chen, Stefano Nichele, Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review, Information Fusion, Volume 59, 2020, Pages 103-126, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2020.01.011.

[8] Yu Liu, Yufeng Ding, Chang Li, Juan Cheng, Rencheng Song, Feng Wan, Xun Chen, Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network, Computers in Biology and Medicine, Volume 123, 2020, 103927, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.103927.

[9] Yongqiang Yin, Xiangwei Zheng, Bin Hu, Yuang Zhang, Xinchun Cui, EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM, Applied Soft Computing, Volume 100, 2021, 106954, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2020.106954.

[10] M. Shamim Hossain, Ghulam Muhammad, Emotion recognition using deep learning approach from audio–visual emotional big data, Information Fusion, Volume 49, 2019, Pages 69-78, ISSN 1566-2535,

[11] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay and N. Arunkumar, "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)," in IEEE Access, vol. 7, pp. 57-67, 2019, doi: 10.1109/ACCESS.2018.2883213.

[12] R. L. Rosa, G. M. Schwartz, W. V. Ruggiero and D. Z. Rodríguez, "A Knowledge-Based Recommendation System That Includes Sentiment Analysis and Deep Learning," in IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2124-2135, April 2019, doi: 10.1109/TII.2018.2867174.