Emotion Based Music Recommendation - Emosic

Arvind Srinivas K G*, Shrutika Pawar S*, Haresh V*

*Department of Information Technology

Coimbatore Institute of Technology
Coimbatore - 641014

{2007006it,2007015it,71762107206it}@cit.ed.in

***Abstract*— Music has a profound impact on our emotions and cognitive functions, and recent research has shown that facial expressions can provide valuable insight into a person's emotional state. In this study, we propose a personalized music recommendation system that utilizes computer vision and machine learning to analyse a user's facial expressions and recommend music tracks that match their mood. The system extracts facial landmarks in real-time, allowing for accurate analysis of the user's emotional state, and advanced algorithms are used to train the system to recognize different human emotions or moods. This system has the potential to improve the emotional well-being of users and has broad applications across various industries, including healthcare, entertainment, and education.**

***Keywords*— facial expressions, computer vision, machine learning, emotion recognition, personalized music recommendations.**

I. Introduction

As humans, we spend a significant amount of time observing and interpreting each other's behaviours, which is an essential aspect of our social interaction. A significant portion of our communication is nonverbal and involves the expression of emotions through our facial expressions. Decoding facial expressions is a critical component of understanding and interpreting human emotions. With the advancements in technology and machine learning, it is now possible to develop models that can recognize facial expressions accurately. This research paper aims to demonstrate how a model can recognize facial expressions and assist in changing the user's mood.

Facial expression recognition is a crucial aspect of human emotion recognition. The face is a vital tool for communication, and facial expressions are universal across different cultures and races worldwide. Machine learning algorithms have been developed to analyse high-dimensional facial features and accurately recognize different emotional states.

The ability to accurately recognize facial expressions has significant implications across various industries, including healthcare, entertainment, and education. By leveraging the power of machine learning, it is now possible to develop personalized systems that can cater to individual emotional needs. The ability to recognize and interpret facial expressions is an essential part of human social interaction. It allows us to understand the emotional state of others and adjust our behaviours accordingly. With the recent advancements in computer vision and machine learning, it is now possible to develop systems that can recognize and analyse facial expressions in real-time. These systems can be utilized to provide personalized recommendations, such as music or entertainment, based on the user's emotional state.

This research paper aims to demonstrate the effectiveness of facial expression recognition and its potential applications in various industries. By leveraging the power of machine learning, we can develop intelligent systems that can help improve the emotional well-being of individuals.

II. Related Work

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

A. Literature Survey

This paper [1] introduces a system that focuses on real-time face emotion classification and recognition using a deep learning model. The authors present their research and implementation of a model that utilizes deep learning techniques to accurately classify and recognize emotions from facial expressions. The system operates in real-time, providing efficient and timely results. The paper highlights the use of deep learning algorithms in training the model, allowing it to learn complex patterns and variations in facial expressions. The authors discuss the methodology employed, including the dataset used for training and evaluation. They also present the results of their experiments, demonstrating the effectiveness of their deep learning model in accurately classifying and recognizing facial emotions.

In this research [2], a Facial Expression Recognition (FER) system is developed using Histograms of Oriented Gradients (HOG) and various machine learning algorithms. The study focuses on effectively classifying human facial expressions using the Extended Cohn-Kanade (CK+) dataset. To handle the high-dimensional data, Principal Component Analysis (PCA) is applied for dimensionality reduction. The FER system incorporates three classifiers (Support Vector Machine, Neural Network, and K-Nearest Neighbors) and undergoes 10-fold validation. Experimental results reveal that SVM outperforms NN and K-NN classifiers, achieving 93.53% accuracy.

In this system [3], they proposed an unique approach where the system utilizes human facial input to detect emotions, subsequently playing music based on the detected emotional state. To extract relevant features, they employed a point detection algorithm, and for effective machine learning training, we leverage the capabilities of OpenCV. Through the proposed system, a commendable level of accuracy in recognizing emotions from real face images is achieved, leading to reliable and promising results.

In this system [4], an algorithm has been developed for a real-time emotion recognition system using virtual markers and an optical flow algorithm. It demonstrates effectiveness in recognizing emotions from facial expressions and EEG signals, even in challenging conditions such as uneven lighting, subject head rotation, different backgrounds, and various skin tones. It aims to assist physically disabled individuals and benefit children with Autism, while also having applications in business outcomes and personalized e-learning. The system achieves high accuracy in real-time facial landmark recognition and offline EEG signal analysis. However, data collection was limited due to scheduling constraints at Kuwait University.

This paper [5] introduces FaceNet2ExpNet, an innovative two-stage training algorithm for expression recognition. In the initial stage, we utilize a probabilistic distribution function to model the high-level neuron response, leveraging the fine-tuned face net and incorporating feature-level regularization. This approach effectively exploits the wealth of facial information within the face net. In the subsequent stage, we employ label supervision to enhance the discriminative capability of the system. The results demonstrate that FaceNet2ExpNet significantly improves visual feature representation and surpasses the performance of several state-of-the-art methods across four publicly available datasets.

This paper [6] investigates the effectiveness of the proposed backpropagation neural networks combined with a hidden Markov model for face emotion recognition. The study utilizes different face databases, including Jaffe and Cohn Kanade, to evaluate the performance of the recognition system. The process involves detecting geometric facial points in captured digital images and reducing noise through the application of a non-local median filter. Facial features are then extracted by segmenting the images using local binary patterns and identifying key points using a progression invariant subspace learning method. Optimized global features are selected using the particle swarm optimization technique. The extracted features are trained using backpropagation neural networks and classified using the proposed classifiers. The performance analysis is conducted on the Jaffe and Cohn Kanade datasets, demonstrating reduced error rates and improved classification accuracy compared to previous approaches. These findings highlight the enhanced sensitivity and accuracy of the proposed classification method.

In one study [7], a preliminary approach focusing on Hindi music mood classification was presented. The study utilized simple audio features and achieved an average accuracy of 51.56% using the 10-fold cross-validation technique with MIREX mood taxonomy. Additionally, an article [10] highlights the current state of music recommendation research, emphasizing the need for systematic investigations into user behavior and needs, improved feature extraction methods, and the utilization of multiple evaluation indices. The article further emphasizes the significance of considering contextual factors, such as the situation, in personalized music recommendation systems. Notably, it was observed that assigning equal weights to all contextual factors significantly reduced the accuracy of recommendation results.

B. Existing Systems

- **EMO Player:** Emo player (an emotion-based music player) is a novel approach that helps the user to automatically play songs based on the emotions of the user.
- **SoundTree:** Sound Tree is a music recommendation system which can be integrated to an external web application and deployed as a web service. It uses people-to-people correlation based on the user's past behaviour such as previously listened, downloaded songs.
- **lucyd:** lucyd is a music recommendation tool developed by four graduate students in UC Berkeley's Master of Information and Data Science (MIDS) program. lucyd lets the user ask for music recommendations using whichever terms they want.
- **Reel Time.AI:** This system works by having the user subscribe to them. The user can then upload images of large gatherings such as shopping malls, movie theatres and restaurants. The system then identifies the moods happy and sad. It recognizes which faces portray happy emotion and which faces portray sad emotion, and gives the verdict of the situation from the faces of the people present.
- **Music.AI:** It uses the list of moods as input for mood of the user and suggests songs based on the selected mood. It is a combination of Collaborative filtering based and Content based filtering models. Emotion, time, ambience and learning history are the features taken into account for music recommendation.

- **Pandora, Jango, and Last.fm-** These applications recommend music based on user preferences and radio stations, but do not offer any customization based on user's mood.
- However, an application like **Moodagent** incorporates options like manual selection of songs and playlist creation. This application uses mood-based algorithms to generate playlists that match the user's current mood. Additionally, it allows the user to adjust the intensity of their mood selection, creating a more personalized experience.

- Other music player applications like **Apple Music, Tidal, and Amazon Music** also offer personalized playlists, but these are typically based on the user's listening history or preferences, and not on their mood.

III. METHODOLOGY

The emotion-based music recommendation system is an application that focusses on implementing real time emotion detection. It is a prototype which consists of two main modules: Facial expression recognition/emotion detection and Music Recommendation

A. Emotion Detection Module

This module is divided into two parts:

- Face Detection - Ability to detect the location of face in any input image or frame. The output is the bounding box coordinates of the detected faces. For this task, OpenCV is used along with the dlib library for face detection instead of the traditional HaarCascade classifier because, dlib is faster and more accurate than HaarCascade classifier.

- Emotion Detection — Classification of the emotion on the face as happy, angry, sad, neutral, surprise, fear or disgust. For this task, the traditional Keras module of Python is used. low computational efficiency without compromising the accuracy of the results. It uses depth wise separable convolutions to build light weight deep neural networks. The dataset used for training was obtained by combining FER 2013 dataset from Kaggle. The FER 2013 dataset contained grayscale images of size 48x48 pixels. Keras was used to train and test our model for seven classes - happy, angry, neutral, sad, surprise, fear and disgust. We trained it for 15 epochs and achieved an accuracy of approximately 65%

B. Music Recommendation Module

For this module, packages such as tkinter, pygame of Python are used to build a basic music player which plays music from the local storage based on the emotion label that is returned by the previous module.

C. Integration

For the integration of these two modules, the trained model architecture and the model weights are stored in .json and .h5 formats respectively, and it is served in the driver program.

The driver program uses packages such as OpenCV, numpy, dlib, tkinter, pygame, PIL etc for various tasks.

Fig.1, shows the system architecture of our project. the Driver program captures the user's video using OpenCV, dlib package is used for the face detection module and the faces are sent to the emotion detection module inorder for emotion prediction. After the emotion is predicted, suitable songs from the local storage are shown as in the form of a Music Player GUI using the tkinter package. This is a short description of the architecture.

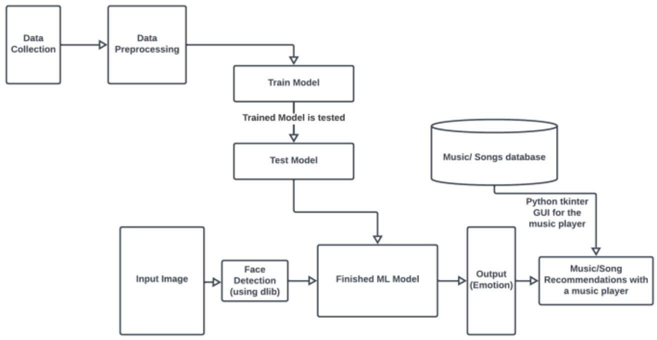

Fig. 1 System Architecture Diagram

The data flow diagram of the system can be seen in Fig.2.

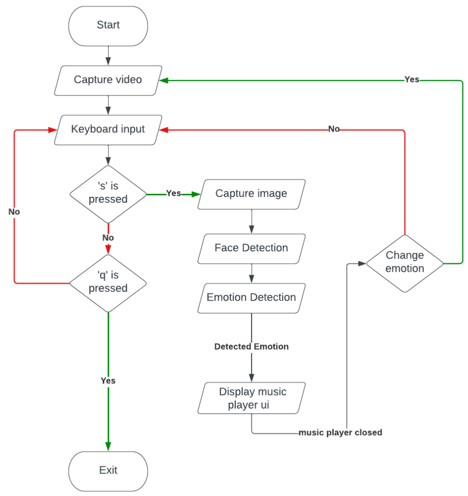

Fig. 2 Data Flow Diagram of the system

The program starts by importing necessary libraries and loading the pre-trained model for emotion detection using Keras. The program then initializes a dlib face detector to detect faces in the images. The main function of the program is get_image() which starts by capturing frames from the camera continuously. It then converts the frames to grayscale, detects faces using the dlib detector, and crops the face regions for each detected face. These face regions are resized to the input size of the model, and the pre-trained model is used to predict the emotion for each face.

The predicted emotions are stored in a list, and the mode of the predicted emotions is calculated. The mode emotion is displayed as text on the frame along with a bounding box around each detected face. When the 's' key is pressed, the program stops capturing frames and starts playing songs using Pygame based on the emotion that was detected. The songs are stored in a folder named based on the emotion that was detected, and the program loads the songs from the corresponding folder.

The Pygame mixer is initialized and the songs folder is defined. A list of all the songs in the folder is created. The current song index is initialized, and a function is defined to play the current song. Other functions are defined to pause, resume, play the next or previous song, and to stop the current song. The program then waits for user input, and the Pygame mixer is used to play songs based on the detected emotion when the 's' key is pressed.

The working of our model is explained with the help of Fig.2, step-wise below:

Step 1: The get_image is called, this function is used to capture frames from the video feed and process them. The frames are converted to grayscale and faces are detected using Dlib's detector.

Step 2: The faces are then cropped and resized to match the input size of the pre-trained Keras model. The model is used to predict the emotion for each face, and the predicted emotion is added to a list.

Step 3: The list of predicted emotions is then used to calculate the mode of the emotions, which is used as the final predicted emotion. The final predicted emotion is then displayed on the video feed and stored in the final_label variable.

Step 4: Set up Pygame mixer and define a folder containing songs for each emotion.

Step 5: Create a list of all the songs in the folder and initialize the current song index.

Step 6: Define several functions to play, pause, resume, skip to the next, and go back to the previous song.

Step 7: When a new emotion is predicted, the song folder is changed to the folder containing the songs to the appropriate folder based on the predicted emotion and plays the first song in the folder, also update the name of the song being played on the Pygame window.

IV. Hardware and Software Requirements

A. *Hardware Requirements*

The hardware requirements required for this project are:

- Minimum 4 Gigabyte (GB) RAM (used for processing)
- Webcam (for testing on laptop/desktop)
- 30 MB Memory space (approximate value)

B. *Software Requirements*

Software Requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application. These requirements or pre-requisites are generally not included in the software installation package and need to be installed separately before the software is installed. The software requirements that are required for this project are:

- Python 3.9
- OpenCV 4.7
- Visual Studio Code
- Dlib 19.22.99
- pygame 2.0.1

V. Results and Discussion

As every person has unique facial features, it is difficult to detect accurate human emotion or mood. But with proper facial expressions, it can be detected up to a certain extent. The model that we have created runs successfully and following are some of the screenshots captured while using it.

Fig.3. displays “Happy” emotion being detected, Fig.4. displays “Neutral” emotion being detected, Fig.5. displays “Sad” emotion being detected, Fig.6. displays “Surprise” emotion being detected, Fig.6. display “Fear” emotion being detected, Fig.8. displays “Angry” emotion being detected, Fig.9. displays “Disgust” emotion being detected.

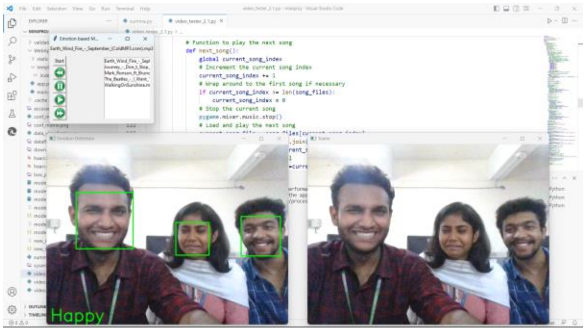

Fig. 3. “Happy” – detected successfully

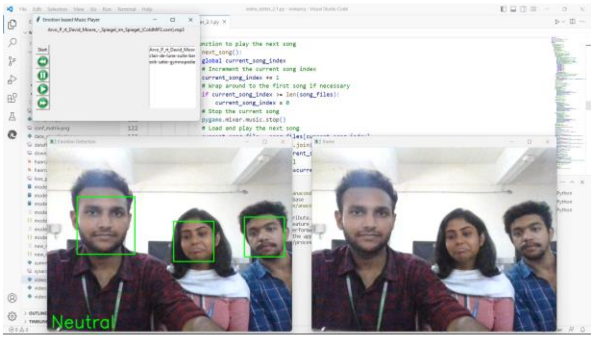

Fig. 4. “Neutral” – detected successfully

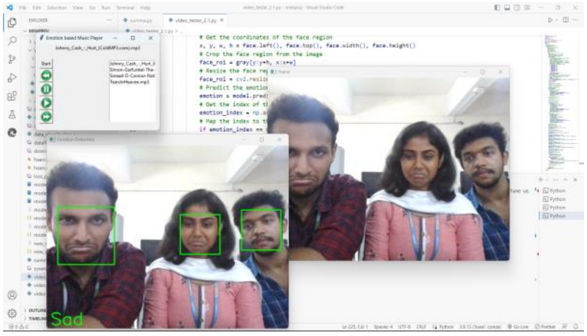

Fig. 5. “Sad” – detected successfully

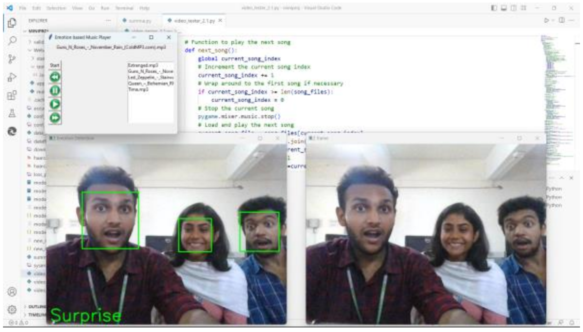

Fig. 6. “Surprise” – detected successfully

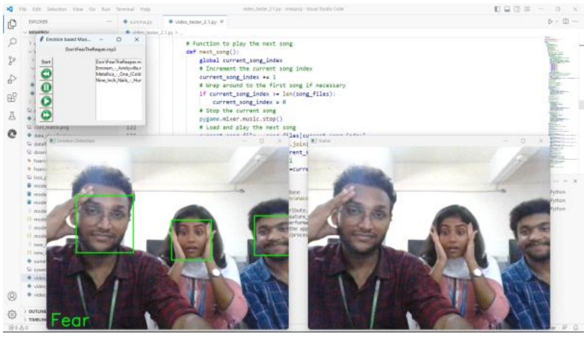

Fig. 7. “Fear” – detected successfully

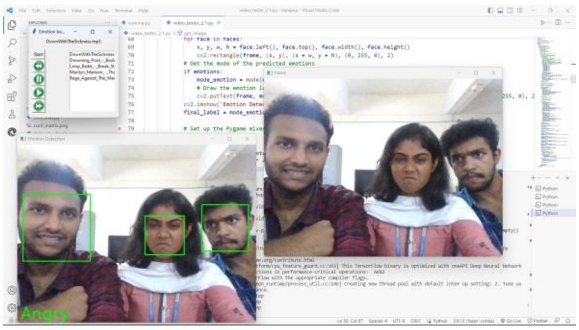

Fig. 8. “Angry” – detected successfully

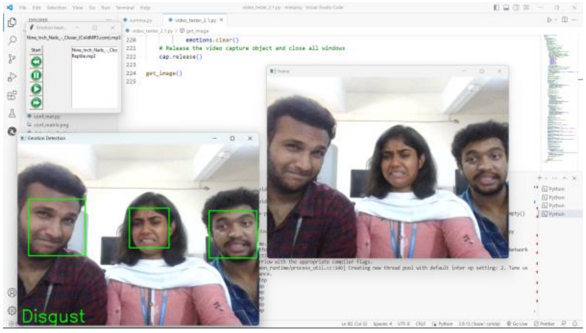

Fig. 9 “Disgust” – detected successfully

The above screenshots also show a music player GUI made using Tkinter which displays the songs according to the emotion label that is predicted by the model and lists the available songs in that playlist in the music player UI, the player has functionalities such as play, pause, resume, previous and next.

The following is the screenshot where the model is able to predict the emotions for a group of people. We can observe that the model is able to predict multiple faces in the captured image and predict the overall emotion of the group. We can note in Fig.10. that there are total of 6 people in the picture and three of them show “Happy” emotion, the other three show different emotions. The algorithm detects five faces and predicts the emotion as “Happy” because majority of the people in the picture show that particular emotion.

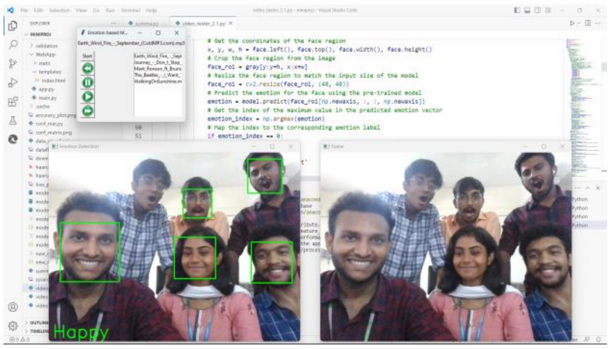

Fig. 9 Group emotion – detected successfully

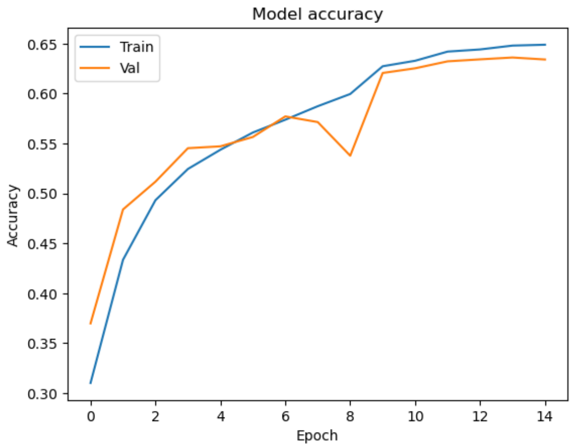

Fig. 10(a). Training and Validation Accuracy

In Fig.10(a), the graph displays the accuracy of our model, where the x-axis specifies the number of epochs and the y-axis specifies the accuracy. As it can be seen in the figure, our model has achieved approximately 65% accuracy. Since it is a fully computer-based system, it understands emotions in the way it has been trained.

In Fig.10(b), the graph displays the loss of our model, where the x-axis specifies the number of epochs and the y-axis specifies the loss during the training and validation phases.

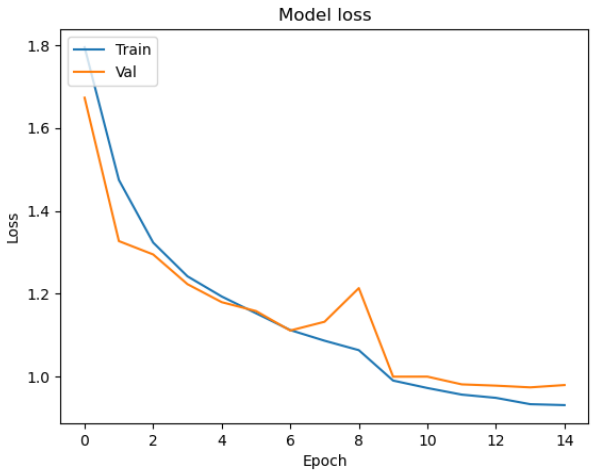

Fig. 10(b). Training and Validation Loss

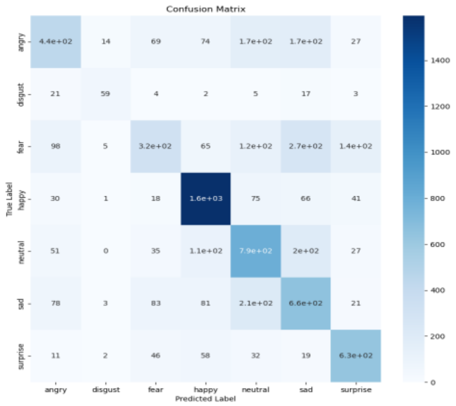

Fig. 11. Confusion Matrix of the Model

Fig.11 plots the confusion matrix for our model. As we can observe, emotions such as Happy, Neutral, Surprise, Sad, Angry and Fear are predicted right most of the times. It is due to the dataset's constituent images for each emotion class. Images for the above mentioned emotion labels are in abundant so the model has learnt well about this emotion and the model is able to predict these emotions accurately.

VI. Conclusions

Our machine learning model is designed to identify various emotions based on facial expressions with an accuracy rate of around 65%. The model can recognize seven distinct emotions, including anger, disgust, fear, happiness, sadness, surprise, and neutrality. This accomplishment is noteworthy because human emotions are complex and subtle, making it difficult to discern them accurately. However, we recognize that the current model's accuracy can be improved, especially in detecting fear and disgust, which can be aided by incorporating additional parameters like heart rate or body temperature alongside facial expressions.

The primary objective of our project is to recommend music that aligns with the user's current emotional state. Based on the analysis of the facial expressions, the model suggests music that suits the user's mood accurately. However, finding appropriate music for certain emotions, like fear or disgust, poses a challenge, which presents opportunities for future development. Enhancing the model's ability to detect these emotions accurately will help identify suitable music for these emotions.

One of the main limitations of our current model is the likelihood of overfitting. Overfitting occurs when the model's ability to recognize specific emotions is highly influenced by the dataset it was trained on, leading to occasional fluctuations in accurate detection. The "disgust" mood is often misclassified as "anger" because the facial features of these two emotions, such as eyebrows and cheeks, are similar. To improve the model's precision, it is necessary to conduct further training with a larger dataset and an increased number of training epochs.

Expanding the scope of our project, we are considering the possibility of recommending movies and TV series based on mood detection. This prospect opens up exciting possibilities for future improvements. By analyzing the user's emotional state, the model can recommend movies and TV shows that align with their mood accurately. This feature could revolutionize the entertainment industry and help users find content that suits their emotional state.

REFERENCES

[1] Dr. Shaik Asif Hussain, Ahlam Salim Abdallah Al Balushi "A real time face emotion classification and recognition using deep learning model "Department of Electronics and Communication Engineering, Middle East College, Muscat. Journal of Physics: Conference Series, ICE4CT 2019.

[2] Hivi Ismat Dino, Maiwan Bahjat Abdulrazzaq "Facial Expression Classification Based on SVM, KNN and MLP Classifiers" 2019 International Conference on Advanced Science and Engineering (ICOASE), University of Zakho, Duhok Polytechnic University, Kurdistan Region, Iraq

[3] ShanthaShalini. K, Jaichandran. R , Leelavathy. S A, Raviraghul. R , Ranjitha. J and Saravanakumar. N "Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques" Turkish Journal of Computer and Mathematics Education Vol.12 No.1 (2021), 912-917

[4] Aya Hassouneh, A.M. Mutawa, M. Murugappan "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods" Informatics in Medicine Unlocked 20 (2020) 100372

[5] Hui Ding , Shaohua Kevin Zhou and Rama Chellappa "FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition" arXiv:1609.06591v2 [cs.CV] 22 Sep 2016

[6] V. Sathya, T.Chakravarthy "AUTOMATIC FACIAL EXPRESSION RELATED EMOTION RECOGNITION USING MACHINE LEARNING TECHNIQUES" International Journal of Computer Engineering & Technology (IJCET) Volume 8, Issue 5, Sep-Oct 2017, pp. 126–135, Article ID: IJCET_08_05_014

[7] Patra, Braja & Das, Dipankar & Bandyopadhyay, Sivaji. (2013). Automatic Music Mood Classification of Hindi Songs.