RESEARCH ARTICLE OPEN ACCESS

Emotion Detection and Music Recommendation System

Dr. S. Padmapriya[1], S. Parvadhavarthini[2], C. Kiruthiga[3], V. Supriya[4]

Professor, Department of Computer Science and Engineering, A.V.C College of Engineering, Mayiladuthurai, India[1], padmapriyas@avccengg.net

U.G. Student, Department of Computer Science and Engineering, A.V.C College of Engineering, Mayiladuthurai, India[2], parvadha.official@gmail.com

U.G. Student, Department of Computer Science and Engineering, A.V.C College of Engineering, Mayiladuthurai, India[3], kiruthigacanagaradjou@gmail.com

U.G. Student, Department of Computer Science and Engineering, A.V.C College of Engineering, Mayiladuthurai, India[4], supriyasasha2019@gmail.com

ABSTRACT: We propose a new approach to automatically play music based on facial emotion detection, which differs from existing approaches that involve manual selection, wearable computing devices, or audio feature classification. We utilize Haar cascade to detect the face, Convolutional Neural Network to recognize emotions and TensorFlow & Tkinter for music recommendations. We tested the system on the FER2015 dataset by capturing facial expressions through an inbuilt camera and extracting features to detect emotions such as happiness, anger, fear, sadness, surprise, and neutrality. The system generates an automatic music playlist based on the user's current emotion which the user can play. It performs better than existing algorithms in terms of computational time.

KEYWORDS: Emotion detection, Music recommendation, Smart music player, Emotion-based music.

I. INTRODUCTION

Recent studies have shown that humans respond to music and that it has a significant impact on the brain. Music is known to affect arousal and mood and has been found to be a useful tool for achieving a positive mood and greater self-awareness. It has also been observed that musical preferences are closely linked to personality traits and moods.

The elements of music, such as meter, timbre, rhythm, and pitch, are processed in areas of the brain that are involved in regulating emotions and mood. Interaction between individuals is a crucial aspect of human life and is conveyed through various means, including body language, speech, facial expressions, and emotions. Emotion detection has become an important technique in many applications, such as surveillance, criminal investigations, and human-computer interface systems.

With the advancement of digital signal processing technology and feature extraction algorithms, automated emotion detection in multimedia content like music and movies is becoming more prevalent. One possible application is a recommender system for music based on facial expressions, which detects the user's emotions and recommends appropriate playlists to enhance their mood. The system uses a Haar Cascade, Convolutional Neural Network to accurately detect facial expressions with a 96% accuracy rate. The dataset used for emotion detection is from Kaggle Facial Expression Recognition, while the music playlist dataset has been created using Kollywood Tamil songs.

II. RELATED WORK

The review is done to get insights into the methods, their shortcoming which we can overcome. A literature review, a literature survey is a text of a scholarly paper, which includes the current understanding along with great findings, as well as theoretical and methodological contributions to a particular topic. The latent qualities of humans that can provide inputs to any system in various ways have brought the attention of several learners, scientists, engineers, etc. from all over the world.

The current mental state of the person is provided by facial expressions. Most of the time we use nonverbal clues like hand gestures, facial expressions, and tone of voice to express feelings in interpersonal communication. Preema et al [6] stated that it is very time-consuming and difficult to create and manage a large playlist. The paper states that the `music player itself selects a song according to the current mood of the user. The application scans and classifies the audio files according to audio features to produce mood-

based playlists. The application makes use of the Viola-Jonas algorithm that is used for face detection and facial expression extraction. Support Vector Machine (SVM) was used in the classification extracted features into 5 major universal emotions like anger, joy, surprise, sad, and disgust.

Yusuf Yaslan et al. proposed an emotion-based music recommendation system that learns the user's emotion from signals obtained through wearable computing devices that are integrated with galvanic skin response (GSR) and photoplethysmography (PPG) physiological sensors in their paper [3]. 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. In [7] Ayush Guidel et al stated that a human being's state of mind and current emotional mood can be easily observed through their facial expressions. This system was developed by taking basic emotions (happy, sad, anger, excitement, surprise, disgust, fear, and neutral) into consideration. Face detection in this project was implemented by using a convolutional neural network. Music is usually told as a "language of emotions" throughout the planet. The paper proposed by Ramya Ramanathan et al [1] conveyed the intelligent music player using emotion recognition. Emotions are a very basic part of human nature. They play the most important role throughout life. Human emotions are meant for sharing feelings and mutual understanding. The user's local music selection is initially grouped based on the emotion conveyed by the album. this is often calculated taking into consideration the song's lyrics. The paper specifically makes a specialty of the methodologies available for detecting human emotions for developing emotion-based music players, the approach a music player follows to detect human emotions, and the way it is ideal to apply the proposed system for emotion detection. It additionally offers a brief idea about our systems working, playlist generation, and emotion classification.CH Radhika et al [8] advised manual segregation of a playlist and annotation of songs, following the current emotional state of a user, as a labor-intensive and time-consuming task. Numerous algorithms had been proposed to automate this manner. However, the prevailing algorithms are slow, increase the overall cost of the system by using additional hardware (e.g., EEG structures and sensors), and feature much less accuracy. The paper presents an algorithm that automatically does the process of generating a playlist of audio, based on the facial expressions of a person, for rendering salvage of time as well as labor, invested in performing this process manually. The algorithm given in the paper directs at reducing the overall computational time and the cost of the designed system. It additionally aims at growing the accuracy of the system design. The system's facial expression recognition module is validated by comparing it to a dataset that is both user-dependent and user-impartial.

III. METHODOLOGY

The system is to capture the face of the user with the user's face as input. Haar-cascade is used for face detection in the system. The CNN is programmed to evaluate the features of the user image. User image features are extracted to detect the emotion depending on the user's emotions. After detecting the emotions, the playlist will be recommended based on the category in the dataset and the songs can be played.

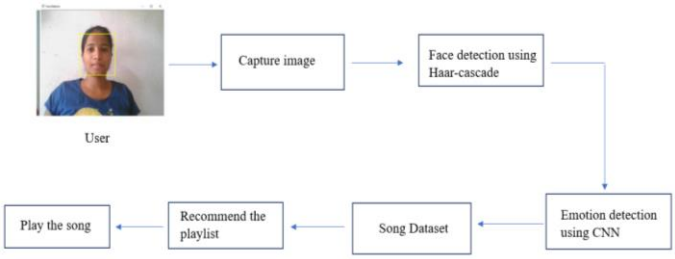

Figure 1: Architecture Diagram

A. Database Description

We created a Convolutional Neural Network (CNN) model using the FER2015 Kaggle dataset, which is divided into two parts: a training set consisting of 24,176 grayscale images and a testing set containing 6,043 grayscale images of faces with a resolution of 50x50 pixels. Each image is labelled with one of five emotions: happy, fear, sad, angry, surprise, or neutral. The images are automatically registered and centered in each image, taking up approximately the same amount of space. The dataset includes both posed and unposed headshots, all in grayscale and with a resolution of 50x50 pixels.

Figure 2: Samples from FER2015 dataset.

The FER-2015 dataset was created by collecting images of emotions and their synonyms through a Google image search. However, using this imbalanced dataset for training facial expression recognition (FER) systems can result in poor performance for the under-represented emotions, such as disgust and fear. One common approach to address this issue is the weighted-SoftMax loss, which assigns a weight to each emotion class based on its proportion in the training set. However, this approach relies on the SoftMax loss function, which may cause features of different classes to stay apart without considering intra-class compactness. To overcome this, an auxiliary loss can be used to guide the neural network. Additionally, to handle missing and outlier values, the categorical cross entropy loss function was employed for each iteration to measure the error.

B. Face Detection

Face detection is a prominent application of computer vision technology, involving the development and training of algorithms that can accurately locate faces or objects in images. This process can be performed in real-time on either video frames or images. Classifiers, which are algorithms designed to identify whether an object in an image is a face or not, are used in face detection. These classifiers are trained using numerous images to improve accuracy. OpenCV utilizes two types of classifiers, namely LBP and Haar Cascades, with Haar classifiers being specifically used for face detection. These classifiers are trained using pre-defined face data to accurately detect different faces. The primary goal of face detection is to identify faces within the image frame by minimizing the impact of external noises and other factors. This process involves machine learning techniques, where the cascade function is trained using a set of input files. Additionally, Haar Wavelet technique is utilized to partition pixels in the image into squares, thereby enhancing the accuracy of the model. Ultimately, the accuracy of the face detection algorithm relies on the quality of the "training data" used in the machine learning process.

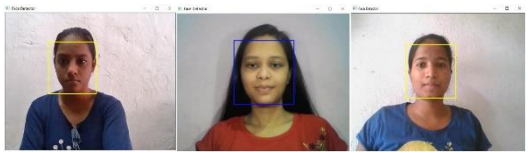

Figure 3: Face Detection

C. Feature Extraction

When performing feature extraction, a pre-trained sequential model can be used as a feature extractor. The input image is passed through the model, and the output of a pre-specified layer is taken as the extracted features. The initial layers of a convolutional network extract high-level features using only a few filters, while deeper layers use a larger number of filters, often double or triple the size of the previous layer's filter. Deeper layers extract more complex features, but this comes at the cost of increased computational complexity.

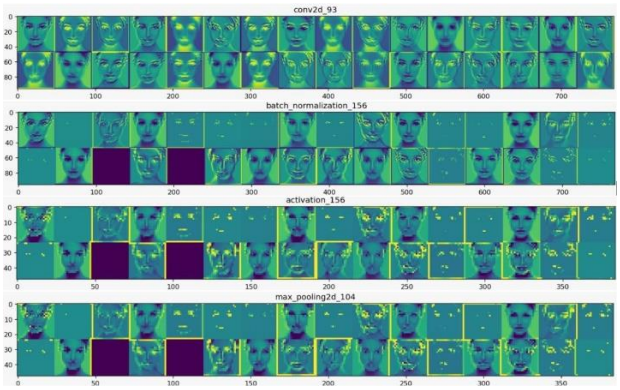

Figure 4: Visualization of Feature Map

In order to extract informative and discriminative features, we employed a Convolutional Neural Network (CNN) [10]. The CNN outputs feature maps, which represent the intermediate activations of all layers after the first layer. To gain an understanding of the prominent features used to classify a given input image, we can load the input image and visualize its corresponding feature map. The feature maps are obtained by applying filters or feature detectors to either the input image or the feature map output from the previous layers. The visualization of feature maps provides valuable insights into the internal representations of the CNN for specific inputs, at each of the Convolutional layers within the model.

D. Emotion Detection

In a Convolutional neural network architecture, filters or feature detectors are applied to input images to produce feature maps or activation maps using the ReLu activation function. These feature detectors help identify different features present in the image, such as edges, horizontal and vertical lines, and bends. Pooling is then applied over the feature maps to ensure invariance to translation. This means that when we slightly change the input, the pooled outputs remain unchanged. There are various types of pooling, such as min, average, and max pooling, with max-pooling performing better than the others. Finally, the inputs are flattened and given to a deep neural network that outputs the class of the object.

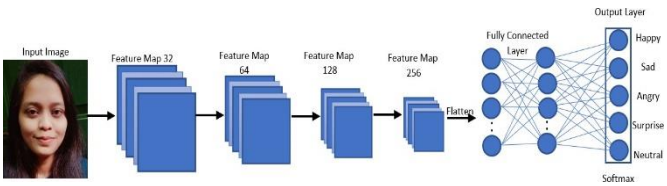

Figure 5: Convolution neural network Architecture

The image can be classified as binary or multi-class depending on the purpose, such as identifying digits or distinguishing between different clothing items. Neural networks function as a black box, and the learned features within them cannot be easily interpreted. In essence, when we provide an input image to a CNN model, it returns the result [10]. To detect emotions, a pre-trained CNN model with weighted parameters is loaded and the real-time image is sent to it for emotion prediction. The model adds a label to the image indicating the detected emotion.

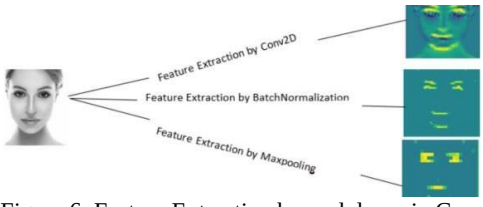

Figure 6: Feature Extraction by each layer in Convolutional Neural Network.

E. SONG DATABASE

A database of Kollywood Tamil songs was developed, comprising 70 to 80 songs for all emotions. Music has a significant impact on our mood, and as such, the system can suggest a playlist of songs that will enhance a user's mood based on their emotional state. For example, if a user is feeling sad, the system will recommend songs that are likely to uplift and motivate them, resulting in an automatic improvement in their mood.

1	Angry	https://op https://op https://op https://op https://open.spotify.com/track/1RQw9ru5PJ5Is7OzgG09Zs?si=63be1ca529704fbd
2	Fear	https://op https://op https://op https://op https://open.spotify.com/track/7LGSPUrp1l8IEuUqM6NG39?si=7fef2bc3b863483c
3	Happy	https://op https://op https://open.spotify.com/track/27N6lqXxjDylUqHa84fA6?si=c79b77c432e04cc2
4	Neutral	https://op https://op https://op https://op https://open.spotify.com/track/5DYfTyqUNBmOnQtJlQYk2?si=2b9acb09ba8a4b67
5	Sad	https://op https://op https://op https://op https://open.spotify.com/track/5xwM2yTshGrx5azFpwIkgG?si=1dd7ecee385c4712
6	Surprise	https://op https://op https://op https://op https://open.spotify.com/track/6Ro16W3cg8dg7um9aRLWAj?si=b6d73fa900e04999

Figure 7: Song Playlist Database

F. MUSIC PLAYLIST RECOMMENDATION

The recommended playlist for the user is displayed on the screen based on the detected emotions of the user. Then, the user will be able to play the song which is being displayed on a shuffled manner from the song playlist database. The Song database contains the songs from Spotify.

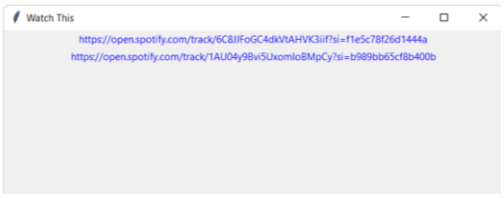

Figure 8: Recommendation of Playlist

IV. EXPERIMENTAL RESULTS

The below figure shows the detection of happy emotion from the system. The 'watch this' pop up recommends the randomly shuffled suggested songs from the song database (in this image shown only the sample of two songs). The same output will be produced for other emotions like fear, sad, surprised, angry and neutral also.

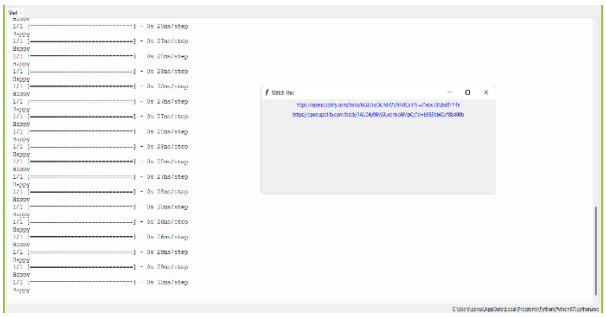

Figure 9: Output and Recommendation of Playlist

Several studies were analysed that implemented using Haar Cascade, a convolutional neural network (CNN) for emotion detection. The performance of these algorithms was compared, and the results are summarized along with the corresponding accuracy values reported in each study. The findings revealed that using a CNN resulted in higher accuracy for emotion detection compared to the other algorithms.The CNN network's hyperparameters are presented in Figure 10. The learning rate controls the weight update at the end of each batch, while the number of epochs represents the iterations of the complete training dataset fed to the network during the training process. The batch size refers to the number of patterns that the network processes before updating the weights. Activation functions enable the model to learn nonlinear prediction boundaries.

Hyperparameters	Values
Batch size	128
No. of classes	5
Optimizer	Adam
Learning rate	0.001
Epoch	48
No. of Layers	28
Activation function	ReLu, SoftMax
Loss function	Categorical-cross entropy

Figure 10: Hyperparameter for trained CNN network.

Adam is an alternative optimization algorithm for stochastic gradient descent used for training deep learning models. Categorical cross-entropy is the loss function employed to quantify the deep learning model errors, typically in single-label, multi-class classification problems.

V. CONCLUSION

After conducting a thorough review of existing literature on music recommender systems, we have established that there are various approaches to implementing such systems. We have determined the objectives of our own system and have chosen to utilize AI-powered technology to stay on trend. Our system aims to help users improve their moods through music by detecting their emotions and selecting appropriate music

tracks. The emotions our system can detect include happy, sad, angry, neutral, and surprised. Once the user's emotion is detected, our system generates a playlist with music that matches the mood. Developing this system will be challenging due to the large dataset it requires and the intensive CPU and memory usage involved. However, we aim to keep the development cost low and make the system easily accessible on standard devices. Ultimately, our music recommendation system will reduce user efforts in creating and managing playlists by utilizing facial emotion recognition.

REFERENCES

[1] Ramya Ramanathan, Radha Kumaran, Ram Rohan R, Rajat Gupta, and Vishalakshi Prabhu, an intelligent music player based on emo- tion recognition, 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2017. https://doi.org/10.1109/CSITSS.2017.8447743

[2] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar, Smart music player integrating facial emotion recognition and music mood recommendation, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, (IEEE),2017. https://doi.org/10.1109/WiSPNET.2017.8299738

Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, Emotion-based music recommendation system using wearable physiological sensors, IEEE transactions on consumer electronics, vol. 14, no. 8, May 2018.https://doi.org/10.1109/TCE.2018.2844736

[3] Ahlam Alrihail, Alaa Alsaedi, Kholood Albalawi, Liyakathunisa Syed, Music recommender system for users based on emotion detection through facial features, Department of Computer Science Taibah University, (DeSE), 2019. https://doi.org/10.1109/DeSE.2019.00188

[4] Research Prediction Competition, Challenges in representation learning: facial expression recognition challenges, Learn facial expression from an image, (KAGGLE).

[5] Preema J.S, Rajashree, Sahana M, Savitri H, Review on facial expression-based music player, International Journal of Engineering Re- search & Technology (IJERT), ISSN-2278-0181, Volume 6, Issue 15, 2018.

[6] AYUSH Guidel, Birat Sapkota, Krishna Sapkota, Music recommendation by facial analysis, February 17, 2020.

[7] CH. sadhvika, Gutta.Abigna, P. Srinivas reddy, Emotion-based music recommendation system, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad; International Journal of Emerging Technologies and Innovative Research (JETIR) Volume 7, Is-sue 4, April 2020.

[8] Vincent Tabora, Face detection using OpenCV with Haar Cascade Classifiers, Becominghuman.ai,2019.

[9] Zhuwei Qin, Fuxun Yu, Chenchen Liu, Xiang Chen. How convolutional neural networks see the world - A survey of convolutional neural network visualization methods. Mathematical Foundations of Computing, May 2018.

[10] Ahmed Hamdy AlDeeb, Emotion- Based Music Player Emotion Detection from Live Camera, ResearchGate, June 2019.

[11] Frans Norden and Filip von Reis Marlevi, A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition, TRITA-EECS-EX-2019:143.

[12] P. Singhal, P. Singh and A. Vidyarthi (2020) Interpretation and localization of Thorax diseases using DCNN in Chest X-Ray. Journal of Informatics Electrical and Elecrtonics Engineering,1(1), 1, 1-7

[13] M. Vinny, P. Singh (2020) Review on the Artificial Brain Technology: BlueBrain. Journal of Informatics Electrical and Electronics Engineering,1(1), 3, 1-11.

[14] A. Singh and P. Singh (2021) Object Detection. Journal of Management and Service Science, 1(2), 3, pp. 1-20.

[15] A. Singh, P. Singh (2020) Image Classification: A Survey. Journal of Informatics Electrical and Electronics Engineering,1(2), 2, 1-9.

[16] A. Singh and P. Singh (2021) License Plate Recognition. Journal Management, Service Science, 1(2), 1, pp. 1-14.