2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)

Facial Emotion Based Automatic Music Recommender System

Mahesh Kumar Singh
Computer Science Department, ABES Engineering College Ghaziabad, UP, India
Email- maheshkrsg@gmail.com

Pushpendra Singh
Information Technology Department
IMS Engineering College, Ghaziabad, UP, India
Email-pushpendra.singh1@gmail.com

Amit Sharma
Information Technology Department
IMS Engineering College, Ghaziabad, UP, India
Email-amit.faculty@gmail.com

Abstract— Listening to music can change the mood of any individual based on type of song like classic, pop or hip-hop. Music can have very strong influence on an individual's emotions. These emotions can be recognized by extracting various facial image features like Eye or Lip movement. Using facial expressions, meaning and context of words can be clearly understood by Humans. With the quick improvement of versatile systems and advanced interactive media innovations, advanced music has gotten to be the standard customer substance looked for by numerous youthful people. People frequently utilize music for mood regulation, more specifically to change a bad mood, to get motivated, to focus on their work, to increase energy levels or to reduce tension. Moreover, tuning in to the correct kind of music at the proper time may progress mental wellbeing. Recommendation of different songs based on an individual's current facial expressions could be latest research problem within the field of image processing. Very few researchers have proposed the solution of this research problem through automatic facial recognition using convolution neural network (CNN) models. In this paper, the proposed system based on CNN model focused on detecting human emotions from face image to identify list of songs to be played through music player to swing user's mood. The objective of this study is to improve the accuracy and reduce time and cost of proposed system for music recommendation.

Keywords—Face Recognition, Computer Vision, Convolution Neural Network, Deep Learning, Open CV, Tensor Flow, Keras

I. INTRODUCTION

Everything revolves around ―data‖. Therefore, data mining, classification, and prediction have all occurred throughout the evolution of computers. Data science has much new technological advancement with the availability of various mining tools, deep learning and computer vision techniques. So why not employ these approaches to make our lives easier.The human face is crucial in determining an individual's mood [1]. Using a camera, the required input is taken straight from the human face [2]. One use for this input is to extract data to determine a person's mood and generate a playlist matching the "mood" of individual generated from the last input. It can save you from the trouble of manually categorizing music and arranging it into different lists and helps you create appropriate playlists according to a person's emotional characteristics.Artificial intelligence is a vast, distinct and fascinating field that has recently attracted many researchers and programs [3]. This domain quickly conquered the world. It is integrated into our daily lives in the form of Chabot's, digital assistants like Siri and Alexa, and several other technology-based systems [4]. It applies machine learning to improve solutions' efficiency, accuracy, and performance to bring effective and reliable solutions to the market [5]. One of the most prominent powers of all these is facial recognition technology. A basic example of its use is to group Google photos of a particular person.

There have been many technological breakthroughs in the area ofmusic suggestion or music information retrieval [6], but we still don't have this problem solved. The retrieval ofmusic information is still a hot topic in academia. As there is a wide choice ofmusic available, we may argue that there is variety. Sorting them gets challenging for the machine. There is a wide range of music, tempos, and amplitudes available, as well as a wide range ofemotions associated with it. In this era of fast growing technology, the traditional methodologies used are comparatively less accurate, high cost and slow in speed. So, it is the need of time to move from old approach to new technological methodology, which is more reliable, fast, and secure.

II. LITERATURE REVIEW

Face recognition, face surveillance, emotion recognition, and sentiment analysis are among the most widely researched topics in the computer vision literature, manually segregating playlists as per emotions of an individual. Suggesting that annotating a song takes a lot of effort and time [7,8]. Automatic song annotation was proposed by a lot of researchers through latest models with low accuracy and higher hardware (EEG Systems, Sensors, etc.) cost. Song selection can be accurate by considering personality and facial emotion of a person [10]. We believe that integrating these psychological aspects will improve the accuracy of recommendations [11,12,13]. People often express their emotions through facial expressions, hand gestures, and pitch of voice, but they usually express emotions through their faces. User temporal complexity is reduced by using an emotion-based music player. People's playlists generally contain a good number of songs. Random song selection could not be a good choice for any individual's mood [14]. This method allows users to autoplay songs based on their mood. Various advances in human emotion recognition have been studied in [15, 16,17], focusing on multiple models for processing audio- visual recordings of different emotions. This paper provides an in-depth survey of audio-visual computer technology. Happiness, sadness, disgust, fear, surprise and anger are among the categories of emotions considered in [18]. The purpose of this article [18] was to address the difficulties in developing automatic and spontaneous emotion recognition capabilities to aid in emotion recognition. We also identified some issues that need to be addressed. A program was written in Eclipse and OpenCV to implement the face recognition method [19, 20].

A novel algorithm proposed in paper [21] is suggested a three-stage procedure for recognizing face emotion: pre-processing of image, feature extraction and recognition, feature classification. Paper [22] developed "Smart music

979-8-3503-2919-3/23/$31.00 ©2023 IEEE

2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) | 979-8-3503-2919-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSEIET58677.2023.10303573

player incorporating facial emotion recognition and music mood recommendation". The Classification Module uses audio features to categorize songs into four different mood classes and reached a stunning result of 97.69 percent. The Based on an individual‘s preferences, feelings of a person were mapped with type of song considering emotion by recommendation Module. Paper [23] developed an "Emotion recognition and music recommendation system using machine learning," which generates an automated playlist based on the user's emotion.

III. Proposed Methodology

In this paper, a web application has been proposed to detect facial expression of a person. A CNN model is used to implement a system for facial expression recognition with other ML algorithms. Facial expressions can be divided into five different groups as _Anger‘, _Sad‘,

_Happy‘, _Surprise‘, _Neutral‘and images for training and testing are taken from Kaggle repository. Integration of different modules was implemented through Flask Framework.

In today‘s world, computing devices have built-in webcam facility which can be utilized to capture emotions to design music recommendation systems. After capturing the image through webcam, various image features are extracted using different feature descriptors based on CNN model. The result of CNN model is an emotion out of five emotions used by the algorithm. The emotion once detected is then used to prepare the list of songs for the user based on the emotion detected. We have used Spotipy module for establishing connection and getting the music tracksfrom Spotify using Spotipy wrapper.

A. Proposed Convolution Neural Network Model

A convolutional neural network consisting several layers to retrieve essential information from data that has a grid-like structure is a neural network. One of the biggest advantages of employing CNNs is that you don't have to conduct a lot of image pre-processing.The filters for most image processing algorithms are often built by an engineer using heuristics. CNNs can automatically figure out most significant filter properties. This saves a lot of time and trial and error labor.

It may not appear to be a significant savings until you work with high-resolution photographs with thousands of pixels. The basic goal of the CNN algorithm is to transform data into easier-to-process formats while preserving crucial representation properties of input. This makes them excellent candidates for dealing with large datasets.The use of convolutions to conduct arithmetic behind the scenes distinguishes a CNN from a standard neural network. In at least one layer of the CNN, convolution is utilized instead of matrixmultiplication. Convolutions return a function after taking two functions.CNNs process data by adding filters to it. The ability of CNNs to modify the filters while training is what makes them unique. Even when working with large data sets, such as photographs, the outcomes can be fine-tuned in real time this way.

A shortage of data is one of the factors that prohibit many issues from being solved with CNNs. While networks can be trained with a small number of data points (10,000 or less), the more data provided, the better the CNN will be tuned.Convolutional neural networks are based on scientific findings in neuroscience. They are made up of nodes, which are layers of artificial neurons. These nodes are functions that return an activation map after calculating the weighted sum of the inputs. This is the neural network's convolutional layer. The weight values of each node in a layer define it. When you feed a layer data, such as an image, it takes the pixel values and extracts some of the visual characteristics.Each layer returns activation maps when working with data in a CNN. These maps highlight key aspects of the data collection. If you provide CNN with an image, it will highlight aspects based on pixel values, like colors, and provide an activation function. The classification layer is the final layer of a CNN, and it determines the predicted value based on the activation map. If you give CNN a sample of handwriting, the classification layer will tell you which letter is present in the image. This is how self-driving cars assess whether an object is another car, a person, or something else.Many additional machine learning techniques are equivalent to training a CNN. You'll start with some training data apart from your test data and adjust your weights based on the predicted values' accuracy. Just make sure your model isn't too tight.

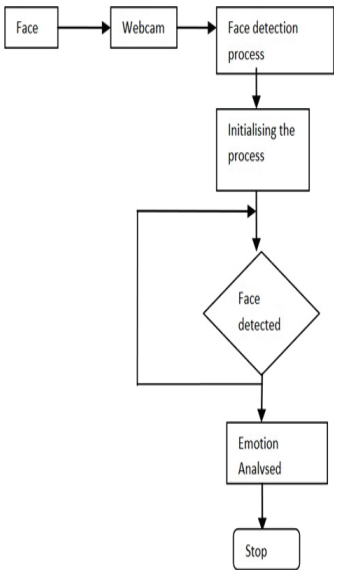

Fig. 1. Flow Chart of Proposed Methodology

A. Dataset Description

FER2013 [24] dataset from a Kaggle is used for the training and testing.It consists gray scale labeled facial images of size [48 x 48] each. The seven categories of emotions are taken into account as [Happiness, Surprise, Anger, Fear, Disgust, Sadness, and Neutral].Training data contains set of 35887 labeled facial images with expressions. Each class of emotion contains around 5000 images with minimum 600 images for disgust expression. The whole dataset has class imbalance problem but still suitable for facial emotion analysis in comparison with other existing facial expression datasets.

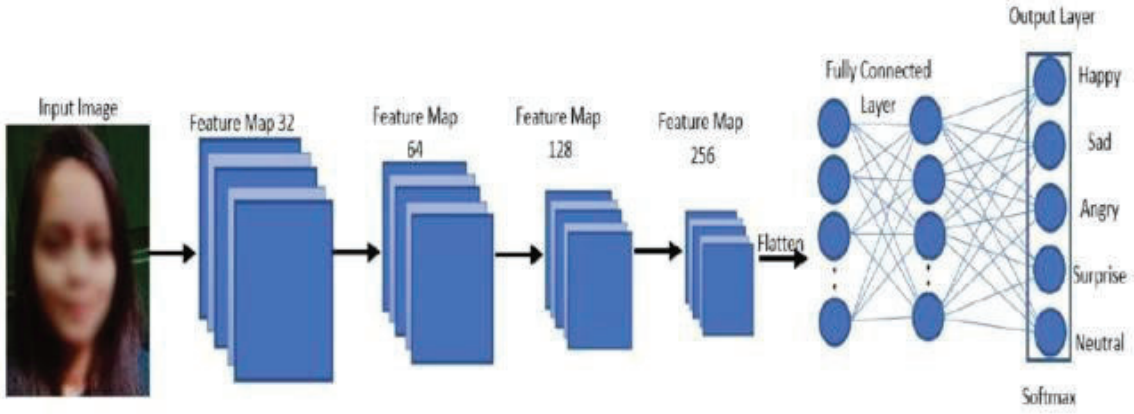

Fig. 2. Proposed CNN Model Architecture

Fig. 3. Retrieved Sample of data

B. Spotify API

The Spotify API is used to fetch the playlists to recommend it to the user according to the mood detected using the model [25]. To access the Spotify API, we need to register on Spotify developer account. Once we are provided with client ID and client secret ID, we can then use it for further information. We need to use the key for authorization so that we can fetch songs from Spotify.

A dictionary is set up that can hold all the data we want to crawl. Next, we looped through each of the genres and then through each track in the genre. For each of these tracks, we crawled metadata and audio information and stored them in the dict we created earlier. Then, we transform the dictionary to a Pandas data frame.

C. Face Detection using Open CV

For computer vision applications, an OpenCV library containing many inbuilt functions related to image is being very much popular these days.

It is an open source software tool to provide basic infrastructure with good accuracy and speed. It is best suited for every business application due to its capability to adapt changes and strong community of contributors to enhance functionality of the OpenCV tool.

Around 2500 ML based algorithms are already included in the tool to be readily available for use in computer vision applications. Various tasks on images such as face detection and recognition, object classification and human action analysis in videos, tracking camera movement and moving object recognition, eye movement tracking in videos and sensor image detection and recognition etc. At present around fifty thousand users connected through community groups of OpenCV tool over 18 million number of downloads. The tool

is widely used by different business houses, government agencies and research groups.

IV. Model Implementation and Result

There are following steps are used to implement the proposed model.

Step-1: First, the webcam will be used.

Step-2: The system will begin gathering the available video frames in front of it. Now, the face of the person will be detected out of the whole video frame available.

Step-3: The frame will then be cropped to remove the facial portion.

Step-4: At this point, the cropped image will be transformed to grayscale.

Step-5: Now the taken image will be compared to the available dataset.

Step-6: The person‘s facial expression will now be predicted, and you will receive the expected emotion name as a text on the facial rectangle.

Step-7: Now this text (predicted emotion name) will be taken as input to the music classification part. On the other hand, music will be classified according to mood at the same time.

Step-8: A playlist will be created for each mood after the music has been classified.

Step-9: When the music section receives the predicted mood/expression name as input, the playlist created for that mood will be shuffled, the machine‘s music player will be activated, and random songs will be displayed in accordance with the person‘s mood.

The following processes are used to design and implementation of the proposed model:

- The Facial Expression Recognition 2013 Data set, or FER-2013 dataset [24], is used in the training procedure, with the Convolution Neural Network (CNN) approach for feature extraction and a suitable facial prediction process.
- Real-time facial expression identification, facial object detection using the HAAR Cascade approach, and facial expression classification using the Convolution Neural Network (CNN) Methods.
- While the facial expression identification procedure is in action, the system‘s findings will appear in the information viewer on the expression display board.

A. Integrating the Model with the Website

We have all encountered online applications that employ machine learning. Netflix and YouTube, for example, employ machine learning to tailor your experience by proposing new material based on your viewing history. This allows them to make better decisions and have a better understanding of their users browsing habits. We

may even design an application that takes your input data and predicts the output using the machine learning model. Incorporating the machine learning models into an application, we can put them to better use. This not only demonstrates our machine learning expertise, but also our app development abilities.

Flask is a Python API that allows us to create web- based applications. Armin Ronacher was the creator. Flask framework is more explicit than Django‘s, and it also easier to learn because it requires less basic code to create a simple web application. A Web-Application Framework, often known as a Web Framework, is a set of modules and libraries that allow programmers to create apps without having to write low-level code such as protocols and thread management. The Flask template engine is built on the WSGI (Web Server Gateway Interface) toolkit and the Jinja2 template engine. We utilize the Flask framework in Python to host local servers and route web pages. We utilize it to install our Machine Learning model locally in this case.

B. Results

The machine‘s webcam will be used to obtain real- time photographs of the person. As soon as the webcam is activated, it will begin taking video frames. Haar cascading filter is utilized for face detection from available video frames. It recognizes objects in the video frame that is accessible. The human face is used as the object to detect the face in this case. When the webcam is activated, it displays whatever is in front of it. The figure below depicts the individual in front of the camera. The face spotted out of the available video frame is indicated by the rectangle in the image. The facial component detected from the video frame will now be converted to grey scale in this step, making facial expression recognition easier. (Because gray scale is simple to process.) a gray scale image rather than a color image. As a result, determining the face expression is simple.) The facial part of the previous image is recorded, cropped, and then transformed to grey scale for easier facial expression recognition in the figure below. The emotion prediction will be done based on this facial photograph. In the emotion detection algorithm, when the input image is captured through the webcam then the detected face is highlighted by a frame across it. The emotion associated with that face is predicted and displayed to the user. The purpose of this is to help the user as well as the system become aware of the current mood.

The mood predicted is then analyzed by the system to recommend the song genre. The webpage is divided into two parts (see fig 4), one part mainly displays the emotion of the user, and the other part recommends the music playlist to the user based on the emotion detected.

V. Conclusion and Future Scope

The approach you use for music suggestion is mostly determined by your specific needs and the data you have available. A hybrid technique that incorporates characteristics of collaborative filtering, content-based filtering, and context- aware suggestions is an overarching trend. However, all sectors are constantly evolving, and advancements distinguish each method. What works for one music library may or may not work for another. Access to large enough data sets and understanding how diverse musical aspects influence people‘s perception of music are two typical issues in the subject.

REFERENCES

[1] Parasar, D., et.al, ―Artificial Intelligence and Sustainable computing‖, Algorithms for Intelligent Systems. Springer, Singapore, (2022), https://doi.org/10.1007/978-981-19-1653- 3_3.

[2] Zebrowitz LA, Monte pare JM, ―Social Psychological Face Perception: Why Appearance Matters‖, Soc Personal Psycho Compass, (2008) May 1;2(3):1497. doi: 10.1111/j.1751- 9004.2008.00109.x.

[3] Emmanuel Gbenga Dada, et.al, ―Machine learning for email spam filtering review, approaches and open research problems‖, Heliyon,Volume 5, Issue 6, (2019), ISSN 2405-8440,https://doi.org/10.1016/j.heliyon.2019.e01802.

[4] Schedl, Markus, et.al, ―Music Information Retrieval: Recent Developments and Applications‖, Foundations and Trends in Information Retrieval. 8. 127-261, (2014), 10.1561/1500000042.

[5] Mian, A., Bennamoun, M. and Owens, R., ―An efficient multimodal 2D-3D hybrid approach to automatic face recognition‖, IEEE transactions on pattern analysis and machine intelligence, 29(11), pp.1927-1943, 2007.

[6] Preema J. S, Rajashree, Sahana M, Savitri H, Shruthi S. J, ― Review on Facial Expression Based Music Player‖, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) , (Volume 06 – Issue 15), 2018.

[7] Preema J. S, Rajashree, Sahana M, Savitri H, Shruthi S. J, 0, ―Review on Facial Expression Based Music Player‖, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICRTT – 2018 (Volume 06 – Issue 15),

[8] S. Alizadeh and A. Fazel, "Convolutional Neural Networks for Facial Expression Recognition", arXivpreprint arXiv:1704.06756, 2017.

[9] Zhang, D.X., An, P. and Zhang, H.X., 2018. Application of robust face recognition in video surveillance systems. Optoelectronics Letters, 14(2), pp.152-155.

[10] Awais, M., Iqbal, M.J., Ahmad, I., Alassafi, M.O., Alghamdi, R., Basheri, M. and Waqas, M., Real-time surveillance through face recognition using HOG and feed forward neural networks. IEEE Access, 7, pp.121236- 121244, 2019.

[11] M. Mandila and V. Gerogiannis, "The effects of music on customer behavior and satisfaction in the region of Larissa– the cases of two coffe bars", International conference on contemporary marketing issues (ICCMI), 2012.

[12] Wang, Nannan & Gao, Xinbo & Tao, Dacheng & Liu, Wei. (2014). Facial Feature Point Detection: A Comprehensive Survey. Neurocomputing. 275. 10.1016/j.neucom.2017.05.013.

[13] Cheng, Xiaotong & Wang, Xiaoxia & Ouyang, Tante & Feng, Zhengzhi. (2020). Advances in Emotion Recognition: Link to Depressive Disorder. 10.5772/intechopen.92019.

[14] Spyrou, Evaggelos, Rozalia Nikopoulou, Ioannis Vernikos, and Phivos Mylonas. 2019. "Emotion Recognition from Speech Using the Bag-of-Visual Words on Audio Segment Spectrograms" Technologies 7, no. 1: 20.https://doi.org/10.3390/technologies7010020

[15] Yong-Cai, Pan & Wen-chao, Liu & Xiao, Li. (2010). Development and Research of Music Player Application Based on Android. 23 - 25. 10.1109/ICCIIS.2010.28.

[16] Singh Pushpendra, Hrisheekesha P.N. and Singh K. Vinai,―Ensemble Visual Content Based Search and Retrieval for Natural Scene Images‖, Recent Advances in Computer Science and Communications 2021;14(2),https://dx.doi.org/10.2174/2213275912666190327175712

[17] Metilda Florence and M Uma(2020): ―Emotional Detectionand Music Recommendation System based on User Facial Expression‖ 3rd International Conference on Advances in Mechanical Engineering (ICAME 2020). IOP Conf. Series: Materials Science and Engineering 912 (2020) 062007 IOP Publishing doi:10.1088/1757-899X/912/6/062007

[18] S. Gilda, H. Zafar, C. Soni and K. Waghurdekar, "Smart music player integrating facial emotion recognition and music mood recommendation," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017, pp. 154- 158, doi: 10.1109/WiSPNET.2017.8299738.

[19] Florence, S & Mohan, Uma. (2020). Emotional Detection and Music Recommendation System based on User Facial Expression. IOP Conference Series: Materials Science and Engineering. 912. 062007. 10.1088/1757- 899X/912/6/062007.

[20] Singh Pushpendra, Hrisheekesha P.N. and Singh Kumar Vinai, ―CBIR-CNN:Content-Based Image Retrieval on Celebrity Data Using Deep Convolution Neural Network‖, Recent Advances in Computer Science and Communications 14(1), 2021; https://dx.doi.org/10.2174/2666255813666200129111928.

[21] Kamehkhosh, I., Bonnin, G. & Jannach, D. Effects of recommendations on the playlist creation behavior of users. User Model User-Adap Inter 30, 285–322 (2020). https://doi.org/10.1007/s11257-019-09237-4.

[22] Adiyansjah, Alexander AS Gunawan, Derwin Suhartono,―Music Recommender System based on Genre using Convolutional Recurrent Neural Network‖, Procedia Computer Science 157(2019), pp-99-109.

[23] Willian Garcias de Assuncao and Vania Paula de Almeida Neris, ―An algorithm for music recommendation based on the user's musical preferences and desired emotions. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM '18). Association for Computing Machinery, New York, NY, USA, 205–213. https://doi.org/10.1145/3282894.3282915