Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 43

EMOPLAY- Playlist Recommendation System Based on Facial Emotions

Shubhangi Kale[1], Prajakta Ugale[2]

School of Computer Engineering, MIT Academy of Engineering, Alandi(D.), Maharashtra, India.[1,2]
spkale@mitaoe.ac.in[1],spkale@comp.maepune.ac.in[1], pvugale@mitaoe.ac.in[2], pvugale@comp.maepune.ac.in[2]

***Abstract*— the very commonly found nature of every human being is to decide between things to choose or not. This conflict often arises whenever user has either very few choices which hold an important significance or user might have so many choices to choose from. The mood of a user plays a crucial role in taking up these decisions. This same conflict arises whenever user tries to listen to some music. One can identify an individual's mood by looking at his/her face. More than 50% of users think that their music library is so large that they have difficulty in deciding which music to play. EMOPLAY is developed to play music based on emotions. This paper mainly focuses on developing the concept of emotion detection using the facial expression and music selection and aims to eliminate the hassle of choosing music. There are two parts of this system, first is emotion recognition and second is recommendation of music playlist. If the emotion is sad then provide counselling using chat bot implementation. For this system, CNN, VGG16 and Resnet50 model are trained on FER2013 dataset. The playlist is taken from Spotify and constructed based on 7 different emotions. Then features are extracted to play desire music from recognized emotions. For the proposed system, VGG16 is found better than CNN and Resnet50 which gave approximately 89% accuracy by experimentation.**

I. INTRODUCTION

In daily routine, emotion plays a major role in expressing or communicating human feelings and thoughts. People generally are not aware of their feelings. This paper might help them to check the user's mood in real-time. Understanding/ identifying human emotions will greatly benefit the human computer interaction. Emotion recognition has a very large scope in near future. On the other hand, a music recommendation system brought together with an emotion recognition system will help to get the result in reducing human stress. The proposed system is based on a web application designed specifically to capture the human face, identify the emotion, and suggest a playlist according to it. This proposed system aims to take one problem out of people's lives and help to make their mood lighter. The proposed system mainly takes up the concept of deep learning and web application; by using these two an application is developed where users can see the emotion as well as a playlist is suggested according to the emotion. The playlists which are being used for recommendations are directly taken from Spotify as the playlists are refreshed daily. Hence, the users will not be feeling bored by repetitive songs.

II. LITERATURE SURVEY

S Metilda Florence [1] et. al. discussed about the dilemma a user faces while choosing the music playlist and songs. Hence, suggested a system which assists the users by recommending a playlist based on the user's emotions.

Aditya Gupte et al. discusses the production of android based app made with android SDK (Xbeats) to predict the user's mood using image processing principles and recommend a song on its basis [2].

Jianhua Zhang et. al. analyzed four different methods of facial expression such as facial expressions, speech, behavior (gesture/posture) and physiological signals. Among these, it was found that the physiological signals are most effective and accurate as humans have more control over the other three types of signals and can mislead the emotion recognition system [3].

Byoung Chul Ko [4] et al. described 7 basic types of emotions: happiness, surprise, anger, sadness, fear, disgust, and neutral. He further describes a total of 22 emotions consisting of 7 basic emotions, 12 compound emotions, and three additional emotions.

Renuka R. Londhe et al. grouped facial expressions into 6 common expressions: anger, disgust, fear, happy, sad, and surprised. The paper focused on capturing the changes in various facial properties with the change in facial expression. Further, statistical methods are used to analyze the changes and predict the facial expression.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 43

[5] Kurosh Madani et al. tried to present the potential of artificial neural networks, through some ANN models and ANN based techniques, to solve real world industrial problems dealing with image processing and pattern recognition fields.

Michael Healy et al. first extracted 68-point facial features from the input image and then used a Support Vector Machine algorithm for the purpose of emotional detection [6].

Tuhin Kundu and Chandran Saravanan favored a static approach for emotion classification by using ORB (Oriented Fast and Rotated Brief) as facial signifier for producing information to feed into the SVM (Support Vector Machine). The accuracy achieved by this experiment is about 79.1% after incorporation of the neural expression [7].

Pierluigi Carcagn`ı et al. [8] conducted an experiment to test the effectiveness of Histogram of Oriented Gradients (HOG) descriptor on various publicly available datasets and proved that the HOG configuration is not dependent on the input data. The author mentioned the challenging issues (for having a low recognition rate) such as scene illumination conditions, camera viewpoint, variability in appearance of different subjects [9]. Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade et al. provided the reasoning behind the Cohn-Kanade Dataset and also implemented and provided the baseline results using Active Appearance Models (AAMs) and a linear Support Vector Machine (SVM) classifier for the CK Dataset [10].

In [11], S. Kale et al. used basic resenet50 for for suspect object detection with best accuracy. Abhishek Mutha et Al. [11] proposed CNN model for UAV system to improve the accuracy. In [12], Atharva Gadre et Al. variation of CNN is tested on signature dataset and analyzed with better accuracy.

III. PROPOSED SYSTEM

The project is divided into 2 sections:

a. Face detection and Emotion Recognition by CNN, VGG16 and Resnet50

b. Fetching song playlist based on recognized emotion by best model.

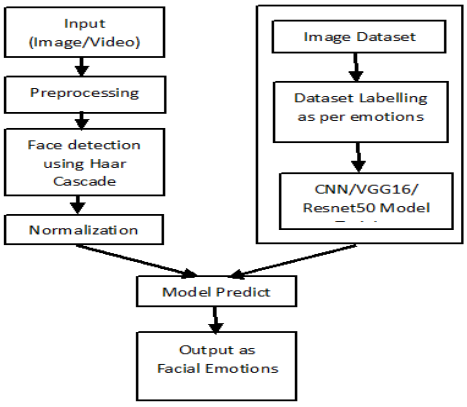

Fig. No. 1 Facial Expression Recognition Flowchart

Fig. no. 1 shows the details of the algorithm of facial expression recognition model is given. Starting with building the model. With use of FER dataset on different models for facial emotion recognition is made. FER 2013 is already a feasible/manageable dataset, so normalization, reshaping and data labeling are the initial steps. The images in dataset are of (48, 48) shape. The gray scaled data is labeled in to X, Y where, X – image matrix and y – emotion. The training the dataset (X, Y) is passed to basic CNN, VGG16 and Resnet50 architecture which are used. Some of the errors were related to small facial features or curves were not identified by the model so the different epoch values are given to CNN layers to check better accuracy. The model was trained with different epochs giving around 60 to 80% accuracy. After saving the trained model it's time to predict the emotion of the given input. Here, when the user is capturing a picture to predict the emotion, the same image is given to detect the face of the user from the captured picture. For face detection the HAAR cascade face detection algorithm is used. A rectangular box is made around the detected face and the next step is to predict the emotion using best model among CNN, VGG16 and Resnet50.

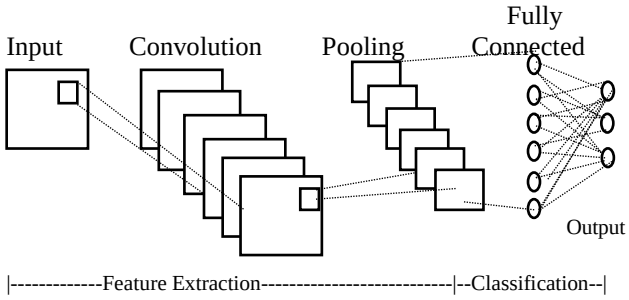

Fig. No. 2 Neural Network Architecture

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 43

Fig no. 2 shows the basic architecture of Convolutional Neural Network. Below fig no. 3 shows flow of to recommend the song for playlist.

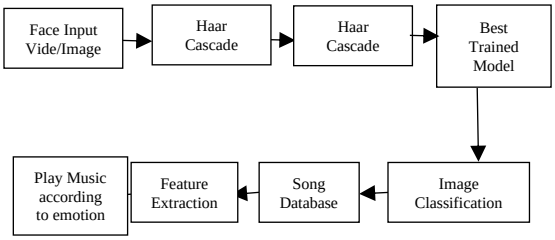

Fig. No. 3 Song Playlist Recommendation Flowchart

IV. Experimental Dataset and its Result

The system requirements for training the model are:

- Central Processing Unit (CPU): Intel Core i5 6th Generation processor or higher. An AMD equivalent processor will also be optimal.
- RAM: 8 GB minimum, 16 GB or higher is recommended.
- Operating System: Ubuntu or Microsoft Windows 10
- RAM: 8 GB minimum, 16 GB or higher is recommended.
- Operating System: Ubuntu or Microsoft Windows 10

For the training facial expression recognition model, the FER2013 dataset is used. The data was put together by Google with help of gathering the results of a Google image search of each emotion and their synonyms. The dataset contains around 28,000 images classified into 7 emotions (sad, angry, surprise, afraid, neutral, happy, disgust). About 24.4% of whole data is classified under 'happy'. The images in the data consists of random posed and unposed images also. The images are in grayscale and 48X48 pixels. Because of randomness in the dataset, it helps the model to not find any unwanted pattern in the dataset and prediction is done only based on given images. The irregularities in reflection and light effects also makes the dataset more real and covering a large portion of possibilities of face imagery.

The accuracy of the model for various epochs on the testing data is increasing as the number epoch are increasing. For the larger epochs, the accuracy is the maximum. The testing is done for the epoch as 10, 20, 30, 40, 50 and 100.

The accuracy of the CNN, VGG16 and Resnet50 model for various epochs on the training, testing and validation data is increasing as the number epoch are increasing. For the larger epochs, the accuracy is the maximum. The testing is done for the epoch as 10, 20, 30, 40, 50 and 100.

The snapshot of the EMPLAY Application Homepage is a follows:

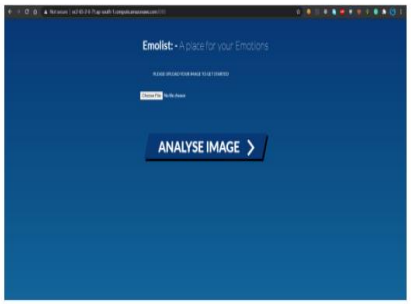

Fig. No. 4: EMOPLAY Application Homepage

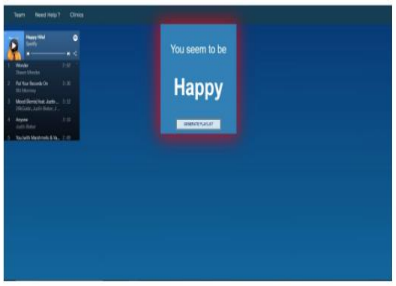

Fig. No. 5: EMOPLAY Display and Playlist Recommendation

There are several features which are added to this proposed system. The first feature which is continuous face detection and output emotion according to it. Take a 2 min video of user as an input and map the emotions throughout the video and take the most dominating emotion to get the prediction. This way the actual emotion of a user in current state can be given. Another feature is added after the recommendation part. If the emotion of a user continuously detects sad motion, then the system will start a chat bot for instant counselling and provide some health clinics references according to it. Two different optimizer are used to train the models such as Adam and Stochastic Gradient Descent (SGD). Table no. 1 show the accuracy and loss of CNN, CGG16 and Resnet50 for the optimizer Adam.

Table No. 1: Model Accuracy and loss for Adam Optimizer

No. of Epoch	CNN		VGG16		RESNET50	
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
10	0.466	2.459	0.860	1.665	0.857	1.784
20	0.589	1.549	0.873	1.510	0.857	1.759
30	0.610	1.495	0.877	1.427	0.857	1.744
40	0.612	1.515	0.880	1.352	0.857	1.738
50	0.629	1.433	0.880	1.365	0.854	1.735
10	0.633	1.418	0.885	1.315	0.857	1.726

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 43

For experimentation purposes, two optimizers are used as Adam and SGD. Fig no. 6 and fig no. 8 show that VGG16 gives better accuracy than CNN and Resnet50as the number of epochs are increased.

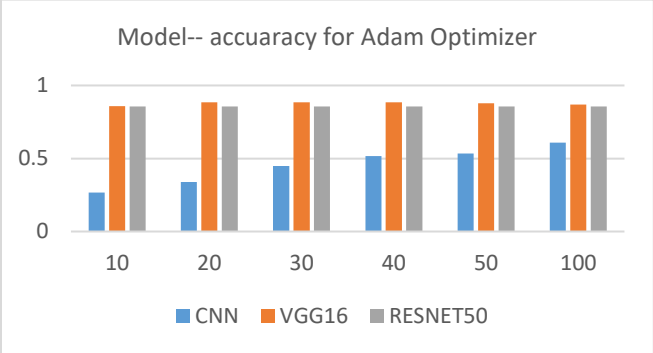

Fig. No. 6: Model Accuracy of CNN, Resnet50 and VGG16 for Adam Optimizer

Fig no. 7 and fig no. 9 show that VGG16 reduces loss than CNN and Resnet50 as the number of epochs are increased.

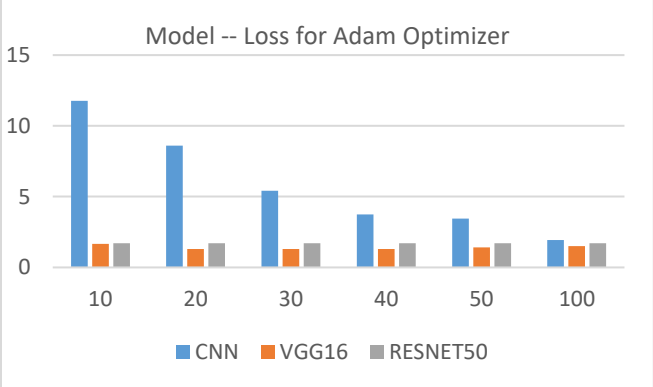

Fig. No. 7: Model Loss of CNN, VGG16 and Resnet50 for Adam Optimizer

Table No. 2: Model Accuracy and loss for SGD Optimizer

Model	CNN		VGG16		RESNET50	
No. of Epoch	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
10	0.267	11.771	0.859	1.666	0.8570	1.719
20	0.339	8.594	0.885	1.302	0.8571	1.719
30	0.448	5.418	0.885	1.307	0.8571	1.717
40	0.516	3.742	0.885	1.298	0.8571	1.717
50	0.534	3.457	0.877	1.421	0.8571	1.718
10	0.609	1.939	0.869	1.511	0.8571	1.717

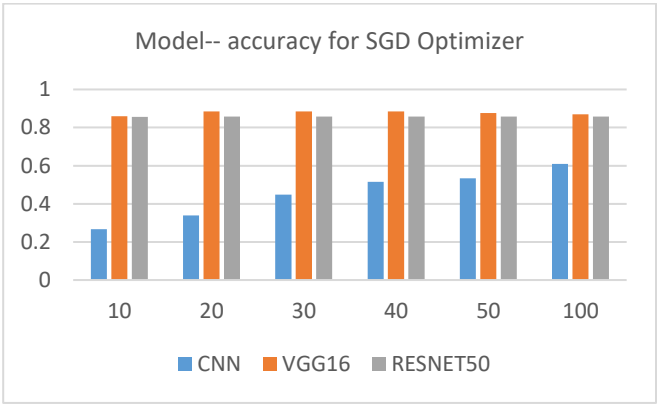

Fig. No. 8: Model Accuracy of CNN, VGG16 and Resnet50 for SGD Optimizer

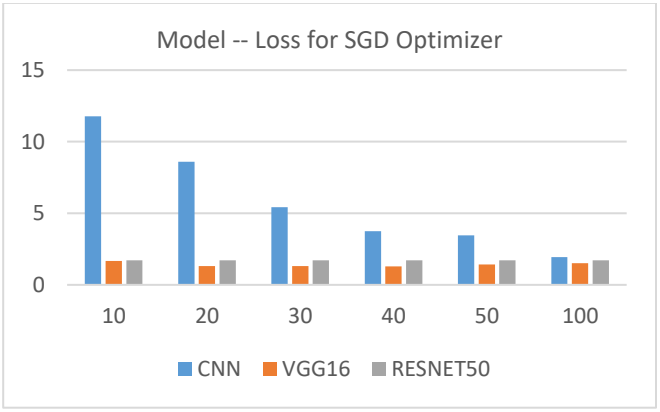

Fig. No. 9: Model Loss of CNN, VGG16 and Resnet50 for SGDs Optimizer

V. Conclusion

This proposed system aims to make people's life simpler and helps to keep focus on the current work. The currently world is witnessing the rise in mental health issues and this problem is growing seriously. This idea has potential to solve some really major problems in real life. To achieve this goal, CNN, VGG16 and Resnet50 models are trained on FER2013 dataset for facial expression and emotion recognition. VGG16 gave better accuracy and less loss as compared to other models. Based on emotion recognition, the features are extracted to play the music as per user's mood.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1, January 2024
Paper: 43

REFERENCES

[1] S Metilda Florence and M Uma. 'Emotional Detection and Music Recommendation System Based on User Facial Expression'. IOP Conference Series: Materials Science and Engineering, vol. 912, 2020, doi:10.1088/1757-899X/912/6/062007.

[2] Aditya Gupte, 'Emotion Based Music Player -XBeats', vol. 9, Sept. 2016, pp. 217–23. ijaers.com, doi:10.22161/ijaers/3.9.30.

[3] Zhang, Jianhua & Yin, Zhong & Cheng, Peng & Nichele, Stefano, 'Emotion Recognition Using Multi-Modal Data and Machine Learning Techniques: A Tutorial and Review', Information Fusion, 2020. 59. 10.1016/j.inffus.2020.01.011.

[4] Ko BC, 'A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors (Basel)', 18(2):401. Published 2018 Jan 30. doi:10.3390/s18020401.

[5] Renuka R. Londhe and Dr. Vrushshen P. Pawar, 'Analysis of Facial Expression and Recognition Based On Statistical Approach', International Journal of Soft Computing and Engineering (IJSCE), vol. 2, no. 2, May 2012, pp. 391–94.

[6] K. Madani, 'Artificial Neural Networks Based Image Processing Pattern Recognition: From Concepts to Real-World Applications', 2008 First Workshops on Image Processing Theory, Tools and Applications, Sousse, Tunisia, 2008, pp. 1-9, doi: 10.1109/IPTA.2008.4743797.

[7] M. Healy, R. Donovan, P. Walsh and H. Zheng, 'A Machine Learning Emotion Detection Platform to Support Affective Well Being', 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 2694-2700, doi: 10.1109/BIBM.2018.8621562.

[8] T. Kundu and C. Saravanan, 'Advancements and recent trends in emotion recognition using facial image analysis and machine learning models', 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT),Mysuru, India, 2017, pp. 1-6, doi: 10.1109/ICEECCOT.2017.8284512.

[9] Carcagnı P, Del Coco M, Leo M, Distante C., 'Facial expression recognition and histograms of oriented gradients: a comprehensive study', Springerplus. 2015;4:645. Published 2015 Oct 26. doi:10.1186/s40064-015-1427-3

[10] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, 'The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,s', 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

[11] Kale, S., Shriram, R. (2021). Suspicious Activity Detection Using Transfer Learning Based ResNet Tracking from Surveillance Videos. In: , et al. Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020). SoCPaR 2020. Advances in Intelligent Systems and Computing, vol 1383. Springer, Cham. https://doi.org/10.1007/978-3-030-73689-7_21

[12] Muttath, A., Veer, A., Pandey, M., Gawai, M., Wangikar, V. (2022). UAV System Using Convolutional Neural Network (1. Angle Prediction Model). In: Fong, S., Dey, N., Joshi, A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 314. Springer, Singapore. https://doi.org/10.1007/978-981-16-5655-2_45

[13] A. Gadre, P. Pund, G. Ajmire and S. Kale, "Signature Recognition Models: Performance Comparison," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2021, pp. 1-6, doi: 10.1109/ICAECA52838.2021.9675598.