Emotion Based Music Recommendation System

Praveen K
Btech Artificial Intellignce
IT Dept
National institute of technology
Karnataka
India
Praveen@nitk.edu.in

Aditya Narayansetti
Btech Artificial Intellignce
IT Dept
National institute of technology
Karnataka
India
aditya.211ai005@nitk.edu.in

***Abstract*—In this fast paced and competitive world it may be true that we have reached a point with great technological advancements, However the question is at what cost? We would like to do all we can to fix this. We have come to understand that music used right is a good stress buster. Researchers at Stanford University have said that "listening to music seems to be able to change brain functioning to the same extent as medication. The truth is music has some sort of magic and that is what we want to harness.**

The thing is the current convention of listening to music is not that efficient, in other words making playlists according to your mood adding songs to them and then selecting them depending on how you are feeling. Other than this there is no convention being used for the reading of emotions in the market. The majority used convention right now is based on language and frequently listened to artists which doesn't help our case.

So we would like to propose a solution using convolutional neural networks, In other words we would like to train a model with vast datasets of images in order to enable it to first determine emotions and then creates program that is able to input live video feed, then upon application of the model it will detect the emotion being portrayed by the user and then recommend songs to them.

Keywords—CNN, emotion, stress, model,

I. INTRODUCTION

Mental health is something we all need to maintain in order to ensure a happy life. Mental health is not just the lack of mental diseases but also the wellbeing of an individual mentally. WHO estimates that the burden of mental health problems in India is 2443 disability-adjusted life years (DALYs) per 100 00 population; the age-adjusted suicide rate per 100 000 population is 21.1. The economic loss due to mental health conditions, between 2012-2030, is estimated at USD 1.03 trillion. [4] These stats are prone to even more change due to the recent fluctuations of the mental health of the general population due to the whole Covid-19 lockdown situation. The study made by TCOH(the center of healing) implied that stress and anxiety levels have been on the rise with 74 per cent and 88 per cent Indians suffering from stress and anxiety respectively.[5] The study also showed that 68.6 per cent therapists reported an increase in the number of people they see and in the hours they spend taking therapy, and 55 per cent therapists said the number of first-time therapy seekers has risen since the outbreak of Covid-19.

People are generally stressed due to the consequences of this hectic world. There are many reasons for one to be stress, from the dissatisfaction of their job, excessive workload to conflicts within their family, and we'd like to do all we can help no matter how little it may seem. Another reason why we'd like to help people get rid of stress is because according to many sources it leads to them picking up one or the other bad habit.

Music is not only useful for getting rid of stress but is also used to motivate one self, turns them a bit upbeat, reduces the chances of depression, keeps the heart healthy and increases workout endurance. A few interesting but unique applications of music are that it decreases the pain felt and is able to stimulate memories, in other words help in the cure of those with Alzheimer disease or at the very least calm down such patients. The battle against Depression actually is one of the major reasons that pushed us to pursue this topic. Depression is not something that has a straight forward approach or some sort of medicine that can directly cure it, only through mood enhancers and constant mental health surveillance can it be avoided and hence came the motivation for our project.

So in order to address the above issues we decided that a viable solution was to create a program that is able to detect what emotions the human is portraying and then accordingly decide what songs the human should listen to. It will decide based on the model we trained beforehand.

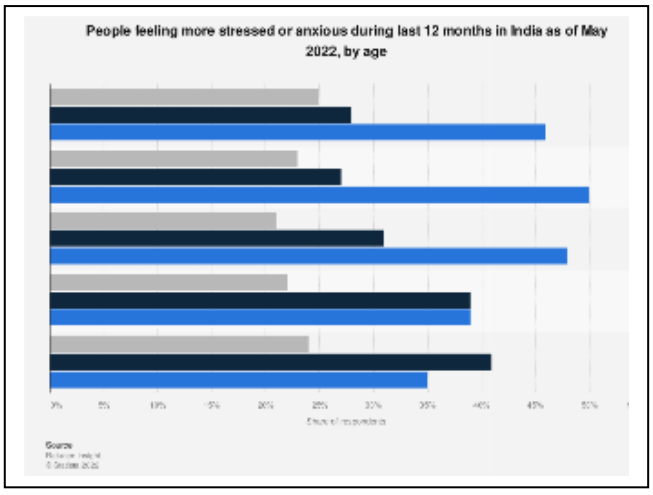


Fig. 1. This is a bar graph showing the percentage of effected people over the past 3 years 2020, 2021 and 2022 for different age groups

II. Related Work

Facial Action Coding System (FACS) [11] is one of the best psychological frameworks which describe muscle movements that the face can produce. CNN shows better performance in Yu and Zhang earlier work of Emotion 2015. Yu and Zhang achieved state-of-the-art results in EmotiW in 2015 using CNNs to perform FER. They used an ensemble of CNNs with five convolutional layers each [16]

There are certain characteristics of a human face that make our emotions more communicable. The automatic face analysis [12] was an amazing method in order to analyze human facial expressions based on both permanent facial features (brows, eyes, mouth) and transient facial features (deepening of facial furrows) in a nearly frontal-view face image sequence.

The effect of the efficiency and accuracy was well explained [13] i.e. difference between shallow and deep neural networks.

A multidetector approach to facial feature localization is utilized to spatially sample the profile contour and the contours of the facial components such as the eyes and the mouth. From the extracted contours of the facial features, we extract ten profile-contour fiducial points [14] and 19 fiducial points of the contours of the facial components. Based on these, 32 individual facial muscle actions (AUs) occurring alone or in combination are recognized using rule-based reasoning. With each scored AU, the utilized algorithm associates a factor denoting the certainty with which the pertinent AU has been scored. A recognition rate of 86% is achieved.

Methods like fractional max pooling and fine tuning resulted in an accuracy of 0.48 in a seven-class classification task [15]

A unique a hybrid feature strategy was proposed, [17] where they trained a novel CNN model with the combination of raw pixel data and Histogram of Oriented Gradients (HOG) features.

Application of a convolutional neural network for an extremely large dataset was proposed [18] i.e. 1.2 million images into 1000 different classes with one of the best accuracy rates so far.

We also saw the proposal of a model that was trained and based on the intensity of pixels and certain facial curvatures that were predefined in order to recognize emotions. [19]

Robert E. Thayer [18] applies rhythm, tempo, intensity, pitch, and timbre to distinguish music emotions. He defines emotion types based on energy and stress. The energy of the music is a range between calm and energetic, while the stress is a range of happy and anxious/sad. This research divides the music emotions into 8 types; namely, exuberance, anxious/frantic, contentment, depression, calm, energetic, happy, and anxious/sad

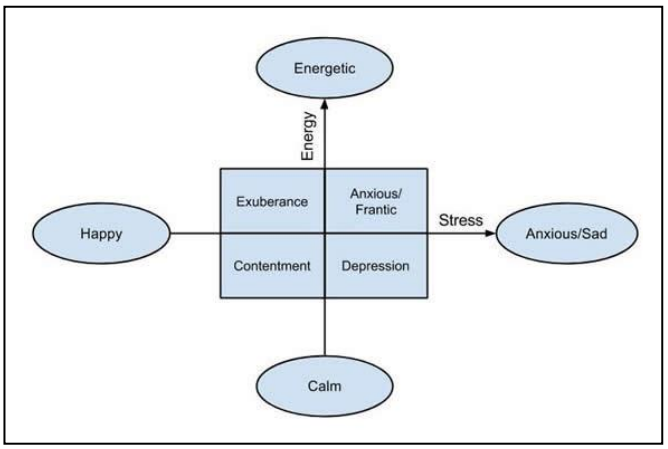


Fig 2. Robert Thayer's classification of music

III. System Design and Anaylsis

A. *Existing music suggestion techniques*

Music suggestion or recommendation is currently used on many platforms like amazon music, Spotify, YouTube music, Apple music, Tidal and Qobuz.

However, one can easily recognize that the technique used by all of these streaming services is quite similar. Generally, they put up some sort of survey during the installation of the application in order for them to get some raw data. They ask questions about your preferred artists or genre and which language of songs you listen to. Other that this as you continue to use their application, they gather that data and further fine tune it in order to improve the suggestions they give you which is based on your music history and the first song you played in this music streaming session.

However, this is useless in solving the problem that we are targeting for, as it doesn't recommend music based on the current feeling expressed by the user

B. *Proposed System Design*

Our system used three steps in order to function i.e. first we must train the model in order to be able to recognize a person emotion based on live video feed or images, secondly we must create an application that is able to receive live video feed and used the previous model and third we must create an

application that uses the emotion we determined from the previous step to select songs from a list of songs or from some online service. Also, the emotions that we have classified are angry, disgusted, fearful, happy, sad, surprised, neutral.

CNN Layer 1	32 Filters Size 3X3 Activation relu
CNN Layer 2	64 Filters Size 3X3 Activation relu
Max Pooling 2D	Size 2X2
Dropout Layer	Parameter 0.25
CNN Layer 3	128 Filters Size 3X3 Activation relu
Max Pooling 2D	Size 2X2
CNN Layer 4	128Filters Size 3X3 Activation relu
Max Pooling 2D	Size 2X2
Dropout Layer	Parameter 0.25
Flattened Layer	No Parameters
Dense Layer	Activation relu Feature size 1024
Dropout Layer	Parameter 0.5
Dense Layer	Activation SoftMax Features 7

Table 1. Layers in the proposed Neural Network

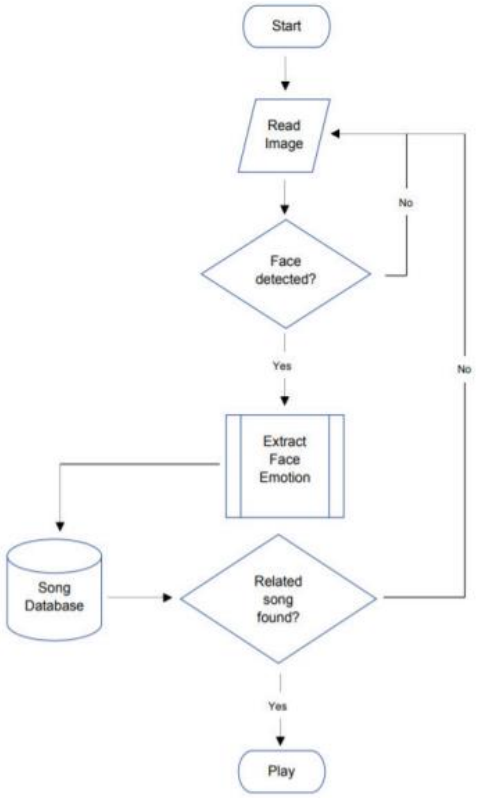


Fig 3. Flowchart for proposed Algorithm

IV. Implementation and Results

For our project we used python to code the entire thing. For the first part i.e. training the model in order to determine emotions from an image we took in a dataset and preprocessed the images into batch size of 64 and greyscale image. Then we start adding layers with activation function as relu and we add max pooling and drop out layers at the end of every two hidden layers, then we add final dense layer of SoftMax activation, then we compile the model with learning rate 0.0001 which can be changed based on the computing power of your system. After this we will be training it for 50 epochs and then we save the model structure and save the model weights and biases.

Upon training the model we move on to the second step ie an application that takes live feed and utilizes the previously made model. For this we used CV2 module in order to video capture until the emotion is detected, now until one emotion is not portrayed in majority the code will continue to run. Once the emotion is detected from our classification of emotions, we will then send this info to the music recommender.

Then the third part takes place, however in order to make it even more efficient depending on whether the user has a language and artist specification they can input that as well so that our program comes to the best possible recommendation.

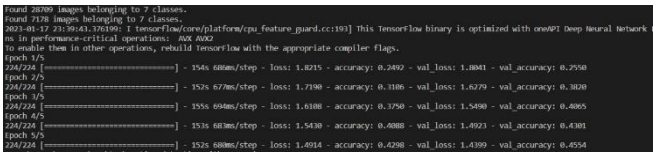


Fig 4. Training model with batch size 128 and epochs 5

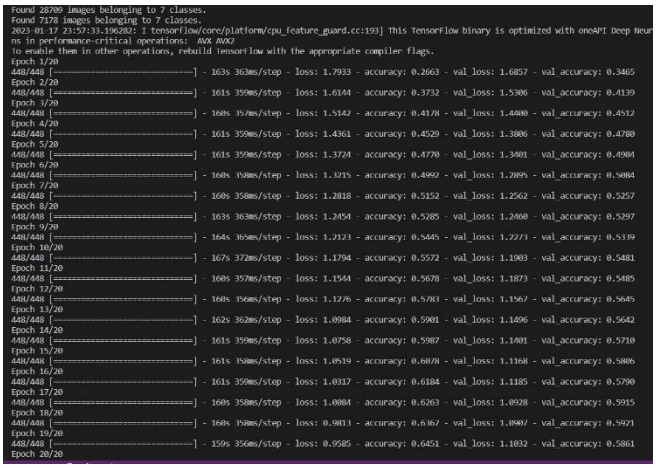


Fig 5. Training model with batch size 64 and epochs 20

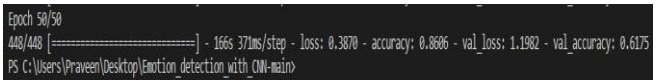


Fig 6. Training model with batch size 64 and epochs 50

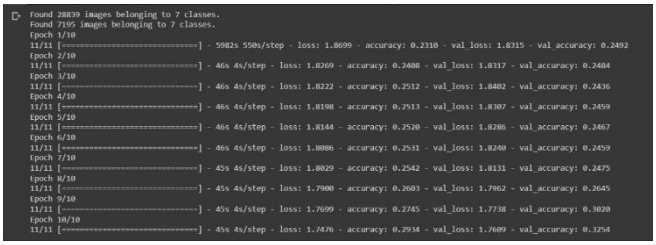


Fig 7. Training model with batch size 2560 and epochs 10

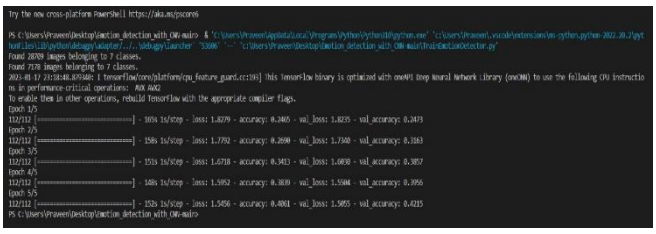
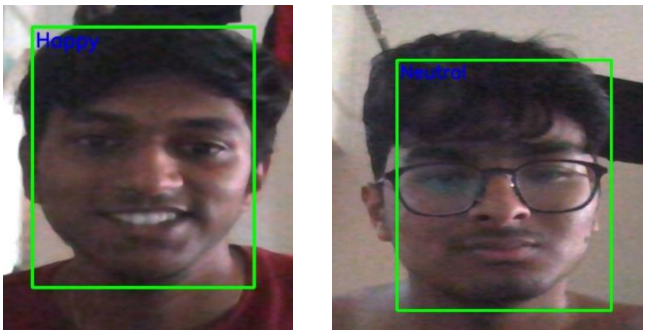


Fig 8. Training model with batch size 256 and epochs 5



V. DATASET

The dataset we used contains 28000 images and are classified into 7 classes. The memory size of each image was around 2Kb each. First, Dataset split into train and test set of 80:20 ratio. The number of images in the training and test class is shown in fig 4 and fig 5.

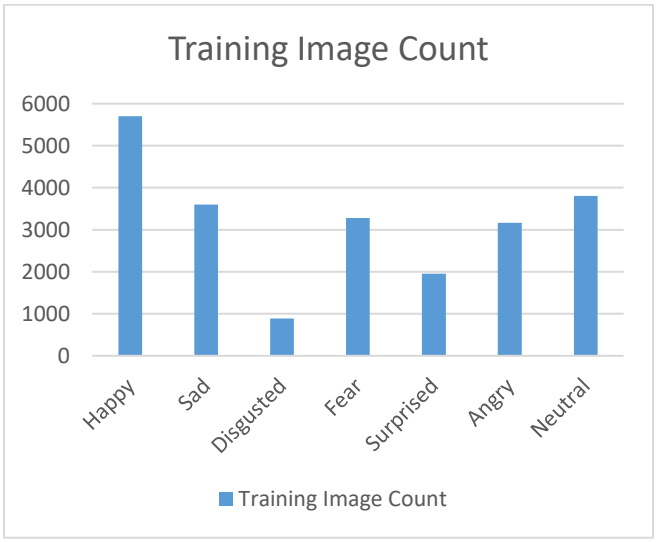


Fig 4. Bar graph for Training Image Count

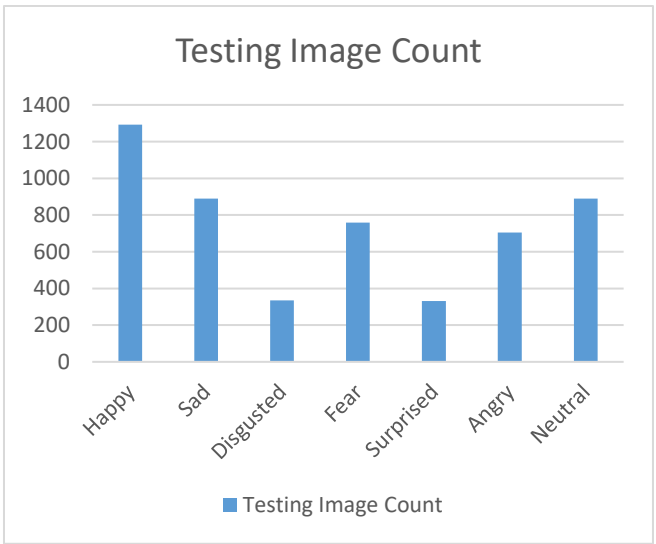


Fig 5. Bar graph for Testing Image Count

VI. Conclusion and Further Scope

In this paper we proposed a Project in order to successfully determine what emotions are perceived by someone in live time and then based on which we recommend songs to them. The proposed system mainly contains one CNN model in order to determine which emotion is being portrayed and a few algorithms in order to ensure that the video feed is taken until a face is seen and another algorithm such that the best possible outcome for each emotion along with the user's artist and language input is given. We can further extend this project sometime in the future for movie recommendation.

Another very useful thing we learnt from this project is how changes to certain things in the program effect the accuracy i.e. If we increase the number of epochs the accuracy increases but if we keep the epochs too high over fitting on the training and testing set takes place which results in poor accuracy when a new dataset is used. Also, when the number of images passed during each iteration reduces the accuracy increases and the time taken for taken increases.

Epochs	No. of images per Set	Accuracy %
10	2560	32.54
5	256	40.61
5	128	48.61
20	128	63.61
50	64	86.06

We also were curious to how the 4th CNN layer really effects our entire model and hence we ran tests into how accurate our model would be without it and got the following results.

Epochs	No. of images per Set	Accuracy%
10	2560	29.21
5	256	36.43
5	128	45.26
20	128	59.02
50	64	82.95

References

[1] https://www.unr.edu/counseling/virtual-relaxation-room/releasing-stress-through-the-power-of-music#:~:text=Upbeat%20music%20can%20make%20you,these%20personal%20experiences%20with%20music.

[2] https://www.webmd.com/balance/guide/causes-of-stress

[3] https://www.statista.com/statistics/1320246/india-people-feeling-more-stressed-by-age/

[4] https://www.who.int/india/health-topics/mental-health#:~:text=It%20includes%20mental%20well%2Dbeing,100%20000%20population%20is%2021.1.

[5] https://indianexpress.com/article/lifestyle/health/indians-stress-anxiety-mental-health-study-7101237/

[6] https://www.northshore.org/healthy-you/9-health-benefits-of-music/

[7] https://en.wikipedia.org/wiki/Emotion_classification

[8] https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/#:~:text=As%20could%20be%20expected%2C%20higher,%2Fsad%2C%20calm%20or%20depressed.

[9] https://www.cnet.com/tech/services-and-software/best-music-streaming-service/

[10] https://www.coursera.org/learn/neural-networks

[11] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J.R. Movellan. Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia, 2006.

[12] Y. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), 2001.

[13] R. Pathar, A. Adivarekar, A. Mishra and A. Deshmukh, "Human Emotion Recognition using Convolutional Neural Network in Real Time," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Chennai, India, 2019, pp. 1-7, doi: 10.1109/ICIICT1.2019.8741491.

[14] M. Pantic and J.M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. IEEE Transactions on Systems, Man and Cybernetics, 34(3), 2004

[15] Arushi Raghuvanshi, Vivek Choksi, Stanford University: "Facial Expression Recognition with Convolutional Neural Networks"

[16] Zhiding Yu and Cha Zhang. Image based static facial expression recognition with multiple deep network learning. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pages 435–442. ACM, 2015.

[17] Shima Alizadeh, Azar Fazel, Stanford University "Convolutional Neural Networks for Facial Expression Recognition" arXiv:1704.06756v1 [cs.CV] 22 Apr 2017

[18] R. Thayer, "Music Mood Classification," 2014.Available: https://sites.tufts.edu/eeseniordesignhandbook/2015/music-moodclassification/.

[19] R. R. Londhe, P. Vrushshen, and Pawar, "Analysis of Facial Expression and Recognition Based On Statistical Approach," Jan. 2012

[20] F. Mozafari and H. Tahayori, "Emotion Detection by Using Similarity Techniques," 2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Bojnord, Iran, 2019, pp. 1-5, doi: 10.1109/CFIS.2019.8692152.

[21] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.

[22] D. Dai, "An Introduction of CNN: Models and Training on Neural Network Models," 2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR), Shanghai, China, 2021, pp. 135-138, doi: 10.1109/ICBAR55169.2021.00037.

[23] A. Ajit, K. Acharya and A. Samanta, "A Review of Convolutional Neural Networks," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, pp. 1-5, doi: 10.1109/ic-ETITE47903.2020.049.