Facial emotion based recommender system

Abhijith P
MTech CSE, IIITD
abhijith21059@iiitd.ac.in

Aman Srivastava
MTech CSE, IIITD
aman21007@iiitd.ac.in

Jatin Agarwal
MTech CSE, IIITD
jatin21032@iiitd.ac.in

Mona Singh
MTech CSE, IIITD
mona21053@iiitd.ac.in

Nehal Chourasia
MTech CSE, IIITD
nehal21056.ac.in

Rishabh Kumar Pundhir
MTech CSE, IIITD
rishabh21071@iiitd.ac.in

1 MOTIVATION AND PROBLEM STATEMENT

Emotions play an important role in human behavior and action. We are highly influenced by the music we listen to and the movies we watch. Mostly the music and movie service providers make use of genre, title and album to search music and movies and user's listening history for retrieving the user's preference. However this system does not focus on extracting the user's preference based on the user's emotion. So our work will recommend movies and music preference by focusing on the user's emotions. There are many recommendation systems in the market but what makes our recommendation system stand out is, we even recommend movies based on user's emotion as novelty. The dataset for facial emotions is FER 2013 and for movie recommendation is IMDB dataset from kaggle.

2 LITERATURE REVIEW

In [1], the authors have suggested the steps for a movie recommendation system from facial expressions. Based on the previous works in the field, it is suggested that CNN is a better model for facial expression recognition. Then based on the emotions obtained it is fed to a recommender system which will recommend movies based on the emotion. Although the research was thorough,and the facial emotion detection was implemented with an accuracy of 99.81%, still the movie recommendation system part was not implemented by the team.

In [5] Human facial expressions are linked to emotions,which in turn are linked to communication. In real life, facial expressions play a vital role in non verbal communication. To build an emotion detection system, the research presents a method for recognising human emotions using the convolutional neural network (CNN) algorithm, a deep learning technique. Majorly, there are two phases to the system implemented : Training a model using the dataset proposed and then once trained, it can be used to classify the new images. The training phase comprises procedures such as face detection and feature extraction. The Viola-Jones algorithm is used for face detection, which employs Haar Feature Selection to crop out undesired parts of the image, such as the backdrop, before converting the image to grayscale.An analytical approach is used for feature extraction, which solely employs the precise features of the face. A three-layer Convolutional network is used to learn: n input layers, seven output layers for each emotion, and a hidden layer. The Face Expression Recognition dataset was utilized for this (FER2013). The accuracy for the proposed model was 79

In [3], the paper aims at scanning human emotions for building an emotion-based music player based on emotions. One of the most prominent tasks was using Deep Neural Networks to learn the most relevant features in real-time and handling the limitations from handcrafted attributes. The model used was VGG16 CNN to detect the facial expressions of the individual user after which the most relevant song based on the user's expression was played. Implantation could be broadly classified into three tasks, including emotion-based detection which makes predictions from the user's emotion and forwarding the emotion target, then Spotify implementation for sending requests to the active Spotify accounts to track down the audio features and tracking playlists, at last implementing the results on the server using flask framework.

In [4], the authors have proposed a real-time emotion-based recognition system using the CNN model. For tuning the model, techniques such as Batch Normalization, Dropout Regularization and Max Pooling have been used. Shapes of 48*48*64 and 24*24*64 were used for Batch Normalization and Dropout Regularization respectively. They have also used the IMDB dataset for extracting emotion from movies using movie reviews and description with the help of the text2emotion library. After extracting the emotion (and its intensity value) from the user's captured image, they have matched and recommended 10 movies on the basis of content based filtering approach using cosine similarity.

In this proposed work [2], the most promising CNN model for emotion detection task of the FER-2013 dataset was described along with face detection methodologies like Haar Cascade and Viola Jones Face detectors. Here, 3 experiments were conducted using Neural Networks for emotion recognition task of the FER-2013 dataset. In the first experiment of networking programming, three convolution and two connected layers were aggregated with max-pooling layers. In the second one, 3 convolution layers instead of 5 were used and the nodes were reduced to 1024 from 4096. In the last one, 48 by 48 layers in the input layer were taken and after this input layer, the model contained a convolution layer that was followed by a contrast normalization layer which was again followed by a max-pooling layer. At the end of the network, there was another two-convolution layer and an output layer that in turn was connected to a softmax layer. After the final evaluation, the performance of the last network was found to be better for the emotion recognition task.

In paper [6], Integrate computer vision and machine learning techniques for extracting emotions and suggest music based on that. Using the camera, facial emotions have been extracted using a point detection algorithm.OpenCV has been used to train input images.Apart from that, they have used the Canny edge detection algorithm in image pre-processing, tensor flow for severe computations, and Pygame for music recommendation techniques. Their work proved to recommend music with a good precision level.

In paper [7], Initially, Dataset has divided into 7 classes after splitting into 80-20 ratio and then engaged in a neural network-based approach. The system has been divided into 3 parts 1) face detection using HAAR cascades which scan an image and return face coordinates. 2) Then they have utilized 6 layers CNN model for emotion detection into 7 classes.Based upon the detected emotion, 3) finally, music has played out of seven folder maps mood and music.

3 PLAN OF WORK

Below figure explains the working of our project in the form of Block diagram.

Initially, We take input as an image which will contains the emotion of a person. We then preprocess the image by converting into pixel array and reshaping and then divided into training, validation and test set. We then recognise emotion using certain techniques then we split emotions into 2 categories and based on that we fetch the corresponding playlists and display to the user.

4 BASELINES

[1] The first baseline that we went with for the emotion recognition system was SVM (Support Vector Machine).SVM are supervised learning algorithms that help in classification and regression. The dataset was divided into three categories the training model, the testing model and the validation model. We trained the SVM with the training dataset, which contains 28709 images and then tested on the test dataset containing 3589 images. The parameters taken were 'C': 1000,'gamma':0.01. The accuracy that we achieved was 47.19%.

[2] The second baseline model used was KNN which is a supervised learning algorithm. The model was trained on the dataset using hyperparameter tuning with different k values ranging from 1 to 5. The best accuracy obtained was 40.23% at k = 1.

5 METHODOLOGY

Convolutional neural networks work using a kernel of different size and multiplying the filter to the image. Here, we have used 3*3 and transform the original image to a feature map. Each cell of the image is multiplied by the kernel to form a feature map, Using these feature maps, the network is able to detect different patterns in the image. The input shape of the image we're taking here is 48*48*1, so that it is easier for CNN to compute. Relu helps make the model non linear, i.e it converts <1 to 0 and leaves others as it is. Max pooling is to convert the feature map to reduced size. It takes a max of (2,2) window. It helps in position invariant i.e doesn't matter where the features the model will detect. Dropout layer helps to prevent the overfitting which means the model is more accurate for training dataset but not with testing.

Our best CNN model has a different number of filters at each layer:

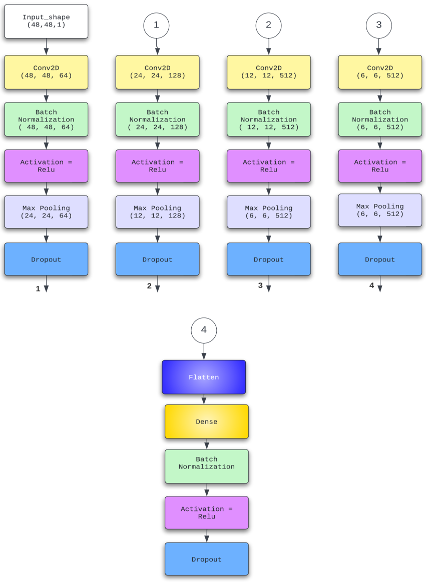


Figure 1

- First layer with 128 filters of kernel size (3,3), using a max pooling of size (2,2).
- Second layer with 256 filters of kernel size (3,3), using a max pooling of size (2,2).
- Third layer consisting of 512 filters with kernel size (3,3), using a max pooling of size (2,2).
- All of these layers used the activation function of relu and a dropout of 0.25.
- The model had an accuracy of 63.28%

Another one of the models we tried was with the following:

- First layer with 32 filters of kernel size (3,3), using a max pooling of size (2,2).
- Second layer with 64 filters of kernel size (3,3), using a max pooling of size (2,2).
- Third layer consisting of 64 filters with kernel size (3,3).
- Fourth layer consisting of 128 filters with kernel size (3,3), followed by a max pooling of size (2,2).
- Fifth layer consisting of 128 filters with kernel size (3,3), followed by a max pooling of size (2,2).
- All of these layers used the activation function of Relu and the last and fully connected layer with a dropout of 0.6.
- The model had an accuracy of 57.70%.

Another Model was trained and tested with Under sampled data with the following configuration.

- First layer with 2 convolution layer with 32 filters each of kernel size(3,3) using batch normalization, maxpool and dropout.

- Second layer with 2 convolution layer with 64 filters each of kernel size(3,3) using batch normalization, maxpool and dropout.
- Third layer with 2 convolution layer with 128 filters each of kernel size(3,3) using batch normalization, maxpool and dropout.
- All of these layers used the activation function of relu and a dropout of 0.3 and the last and fully connected layer with a dropout of 0.5.
- Adam optimizer is used with learning rate = 0.001.
- The model had a validation accuracy of 46.59%.

Another Model was used with the following configuration.

- First layer with 2 convolution layer with 128 filters each of kernel size(3,3) using batch normalization, maxpool and dropout.
- Second layer with 2 convolution layer with 256 filters each of kernel size(3,3) using batch normalization, maxpool and dropout.
- Third layer with 2 convolution layer with 512 filters each of kernel size(3,3) using batch normalization, maxpool and dropout.
- All of these layers used the activation function of relu and a dropout of 0.3 and the last and fully connected layer with a dropout of 0.25.
- Adam optimizer is used with learning rate = 0.001.
- The model had a validation accuracy of 62.55%.

The best model for emotion detection, one with the accuracy of 63.28%. This is the final model that we have considered for recommendation.

6 EVALUATION

When compared with the baselines, all the models are performing better. The baseline results were with accuracies 47%. But the models we used had accuracies of 63.28%, 57.70% and 46.59%. We also saw that there was an imbalance in the dataset and we tried to correct it using undersampling. As we can see that disgust is fairly low in number in the original dataset, but after undersampling, we were able to remove the imbalance in the dataset. The models still need to be improved and one of the factor could be that the model requires more number of images, as we can see that the model overfits on higher epochs.We can also use grid search to find the best set of parameters and activation functions for better performance.

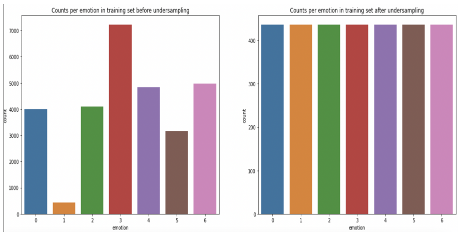


Figure 2

6.1 Metric for evaluation

6.1.1 Accuracy. Accuracy is one metric for evaluating classification models. Informally, accuracy is the number of predictions that our model got correct. Formally, accuracy has the following definition: Or we can also say that:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

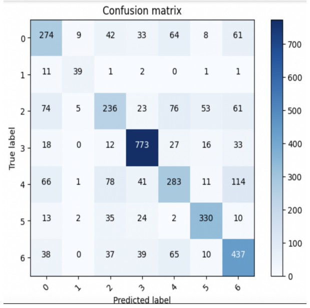
Figure 3

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Figure 4

Accuracy is a good measure when all the classes are balanced. It is not a good measure when one of the label class is in majority

6.1.2 Confusion Matrix. A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm. The confusion matrix consists of four basic characteristics (numbers) that are used to define the measurement metrics of the classifier. These four numbers are:



	0	1	2	3	4	5	6
0	274	9	42	33	64	8	61
1	11	39	1	2	0	1	1
2	74	5	236	23	76	53	61
3	18	0	12	773	27	16	33
4	66	1	78	41	283	11	114
5	13	2	35	24	2	330	10
6	38	0	37	39	65	10	437

Figure 5

1.TP (True Positive): You predicted positive and it's true. You predicted that a woman is pregnant and she actually is.

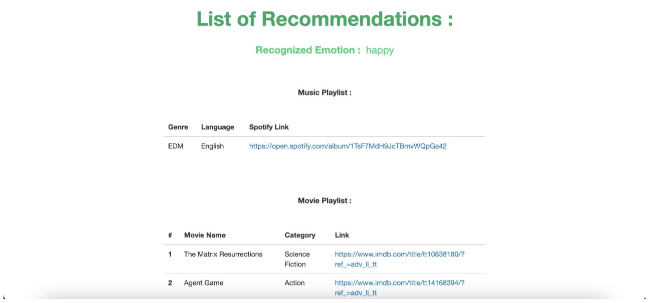
2.TN (True Negative): You predicted negative and it's true. You predicted that a man is not pregnant and he actually is not.

3. FP (False Positive): You predicted positive and it's false. You predicted that a man is pregnant but he actually is not. FP is also known as a Type I error.

4. FN (False Negative): You predicted negative and it's false. You predicted that a woman is not pregnant but she actually is. FN is also known as a Type II error.

7 RECOMMENDATION SYSTEM

We have created a dataset comprising playlists of various genres for music, such as 'Instrumental', 'EDM', 'Rock', 'Pop', and so on, as well as movie recommendations based on user sentiment, such as 'Romantic Comedy', 'Drama', 'Epic', 'Action,' and so on. In order to improve the mood of users, their emotions were linked to various genres. Depending on the severity of the user's feelings, different playlists were suggested based on emotion and genre mapping. For this, we created a website where we can upload an image that we have captured using the camera option on the portal. Then, depending on the image collected, we presented the user with a Spotify playlist and best 10 IMDB movie recommendations, as shown in the figure below



List of Recommendations :

Recognized Emotion : happy

Music Playlist :

Genre	Language	Spotify Link
EDM	English	https://open.spotify.com/album/1TsF7MdH9JcTBmvWQpGa42

Movie Playlist :

#	Movie Name	Category	Link
1	The Matrix Resurrections	Science Fiction	https://www.imdb.com/title/tt10838180/?ref_=adv_li_tt
2	Agent Game	Action	https://www.imdb.com/title/tt14168394/?ref_=adv_li_tt

Figure 6

8 MEMBER'S CONTRIBUTION

All of the authors agreed on the project's concept and completed the necessary preparatory research. Jatin created the dataset and performed the preprocessing. Nehal and Mona were in charge of putting the models into action. The results of the models were analysed by all of the writers. The movie and music playlists for the recommendation algorithm were built by Aman, Abhijith, Jatin and Rishabh. The website was built and deployed by Aman, Abhijith and Rishabh. The report was written by all of them. All of the contributors contributed constructive criticism and assisted in the development of the research, analysis, and report.

9 CONCLUSION

This paper demonstrates how an emotion-based recommendation system for music and movies works. It employs a variety of deep learning approaches to accurately anticipate emotions. We first preprocess the image by transforming it to a three-dimensional image and reshaping it, after which we have chosen the best model as CNN by varying the number of layers to achieve the highest accuracy. The user's image is then presented to the website in order to make this process more convenient.

REFERENCES

[1] Shavak Chauhan, Rajdeep Mangrola, and D. Viji. 2021. Analysis of Intelligent movie recommender system from facial expression. (2021). https://doi.org/10.1109/ICCMC51019.2021.9418421

[2] Banpreet Singh Chhabra. 2020. Emophony – Face Emotion Based Music Player. *International Research Journal of Engineering and Technology (IRJET)* 07 (2020), 8. https://www.irjet.net/archives/V7/i6/IRJET-V7I6112.pdf

[3] G.KIRAN P.SHIVESH KARTHIC ABDUL KAIYUM G.CHIDAMBARAM, A.DHANUSH RAM. 2021. MUSIC RECOMMENDATION SYSTEM USING EMOTION RECOGNITION. Volume: 08 (2021). https://doi.org/10.1109/ZINC50678.2020.9161445

[4] Dr. Rajesh R P Harish. 2021. MOVIE RECOMMENDATION BASED ON HUMAN EMOTION USING CNN. *International Research Journal of Computer Science* Vol.08 (2021). https://doi.org/10.26562/irjcs.2021.v0807.005

[5] Sandhya Armoogum Ravi Foogooa Phavish Babajee, Geerish Suddul. 2020. Identifying Human Emotions from Facial Expressions with Deep Learning. (2020). https://doi.org/10.1109/ZINC50678.2020.9161445

[6] Leelavathy. S Raviraghul. R Ranjitha. J Saravanakumar. N ShanthaShalini. K, Jaichandran. R. 2021. Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques. *Turkish Journal of Computer and Mathematics Education* 12, 1 (2021), 6.

[7] Deevesh Chaudhary Sunil kumar Shikha Sharma Vijay Prakash Sharma, Azeem Saleem Gaded. 2021. Emotion-Based Music Recommendation System. *9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (2021), 5. https://doi.org/10.1109/ICRITO51393.2021.9596276