MUSIC RECOMMENDATION THROUGH FACIAL EMOTION DETECTION USING DEEP LEARNING

M. Jahnavi, A. Keerthi Akshaya, M. Lasya, G. Devi Varaprasad, Students,
M. Sunil Babu Professor5
Dept. of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology, AP,

Abstract— *We cannot imagine our lives without music. Only commercially produced music is played for users. The selection of the main features is an enormously important issue for systems like facial expression recognition. The recommended strategy helps individuals in their musical listening by providing recommendations based on emotions, feelings, and sentiments. The seven facial emotion categories that have been considered are angry, disgusted, fear, pleased, sad, surprise, and neutral—are meant to be specifically allocated to each identified face. To classify the emotion, the object should be detected from an inputted image. The object can be recognized in the image using the Haar-Cascades technique. This algorithm can be defined in different stages: Calculating Haar Features, Creating Integral Images, BiLSTM, and Implementing Cascading Classifiers. A deep learning model called BiLSTM (Bidirectional Long Short-Term Memory) is used to categorize human emotion. Based on the predicted emotion the music is mapped and the playlist is recommended to the user. The k-means clustering algorithm is used to map the music to the expected emotion, as compared to the existing models the deep learning model BiLSTM will give the best performance and 86.5% accuracy.*

Keywords: Haar-Cascades algorithm, BiLSTM model, k-means clustering algorithm, Deep Learning.

I. INTRODUCTION

Generally, facial expressions are the primary means through which people communicate their sentiments. Although one may conceal their words, one cannot conceal their expressions. People have long been aware that music may influence their emotions. 90 out of 100 people like to listen to music. Considering that, this work aims to recommend music based on the user’s emotions. Sensing and recognizing the emotion being detected and displaying appropriate songs can increasingly calm the user's mind and the overall pleasant end up giving a pleasing effect.

At first, the image is given as input and the detection of an object can be done. Later, the emotion of a person is predicted. Based on the predicted emotion the music is mapped. The model is programmed to analyze an image using segmentation method and algorithms for image processing in order to extract information from the target person's face and attempt to determine the emotion the person is attempting to convey. This article seeks to lighten the user's mood by playing music that suits the user's requirements. The greatest method to interpret or infer someone else's feelings is through their facial expressions. Facial expression recognition has been the most effective method of expression analysis known to mankind since the dawn of time. Occasionally, altering one's mood might aid in overcoming challenges like melancholy and despair.

The model is trained with the FER-2013 dataset which contains images with all seven emotions. The image is given as input then object detection, and image pre-processing can be done. Then the emotion can be predicted, and now the music is recommended based on the predicted emotion. To do this the Deep Learning models Haar-Cascades, BiLSTM, and k-means clustering are used. The object can be recognized in an image or video using the Haar-Cascades technique. After detecting the object, by using the BiLSTM model the emotion of an inputted image is predicted. The model is used to recognize facial emotion expressions. Recognizing the expression of a person is very difficult because everyone cannot express their emotions or feelings in the same way, everyone will have their own fashion. So, this work aims to predict emotion accurately. Then based on the predicted emotion, the respective music will be recommended. To do this k-means clustering algorithm the mapping of emotion and music.

II. LITERATURE SURVEY

A smartphone-based mobile system developed by Hyoung-Gook Kim, Gee Yeun Kim, and Jin Young Kim included two essential modules for recognizing human activities and then making music recommendations based on those actions. Their approach uses a deep residual bidirectional gated recurrent neural network to extract high activity detection accuracy from smartphone accelerometer. The results are supported by extensive tests using data from the real world. Extensive trials using real-world data demonstrate the suggested activity-aware music recommendation framework's correctness.

JIANNAN YANG 1, TIANTIAN QIAN 1, FAN ZHANG 2, AND SAMEE U. KHAN, senior IEEE members, presented facial action unit (AU) identification, which detects facial emotions by examining cues relating the movement of atomic muscles in the immediate face area. They might construct AU values based on the observed facial feature points and then use them to classify algorithms for emotion recognition. With the edge devices, they have optimized and customized algorithms to directly interpret the raw picture data from each camera, allowing them to send the identified emotions more readily to their end-user. As a result, they used Raspberry Pi to create a lightweight edge computing-based distributed system.

Deep learning algorithms were used by KORNPROM PIKULKAEW, EKKARAT BOONCHIENG, WARAPORN BOONCHIENG, and VARIN CHOUVATUT4 to study the utilization of 2D facial expressions and motions to evaluate pain. Their method divides pain into three categories: not painful, becoming painfully painful, and becoming excruciatingly painful. To sum up, their research offers a different method of assessing pain before hospitalization that is quick, affordable, and simple for both the general public and medical experts to understand. This analytical method might also be used to other screening methods, such the identification of pain in infectious disorders. An Xception-inspired model used residual blocks and depth-separable convolutions to achieve an accuracy rate of 81% for unforeseen events, but only 51% for neutral emotion recognition.

ZIFAN JIANG, SAHAR HARATI, ANDREA CROWWELL, SHAMIM NEMATI, AND GARI D. CLIFFORD predicted that an automated facial expression detection system based on convolutional neural networks (CNN), pre-trained on a massive auxiliary public dataset, can improve generalizable approaches to MDD automatic assessment from videos and classify remission or response to treatment. They tested a new deep neural network framework on 365 video interviews (88 hours) from a group of 12 depressed patients before and after DBS therapy. A Regional CNN detector and an ImageNet pre-trained CNN were used to extract seven primary emotions. The Open face toolkit was also used to extract facial action units. The classifier achieved 63.3% accuracy in the Affectnet evaluation set.

PRANAV E, SURAJ KAMAL, SATHEESH CHANDRAN C, and SUPRIYA M.H have presented an emotion recognition system that can be deployed with high accuracy. The main is to categorize five different human face emotions. Using the manually gathered image dataset, the model is trained, tested, and verified. With an accuracy of 84.33%, the model can forecast various emotions. The method tested attained an accuracy of about 83%. The accuracy of the model, which utilizes an Adam optimizer to reduce the loss function, was tested and found to be 78.04% accurate.

Hui Zhang, Kejun Zhang, and Nick Bryan-Kinns constructed an emotional map between task and song for the two nations based on emotional preferences, cultural differences, and an examination of the emotional preference of music in daily activities through a cross-cultural survey in China and the UK. Then they unveiled EmoMusic, a ground-breaking emotion-based music suggestion service for everyday tasks that lets users see and manage music emotion through an interactive interface. User research is offered to assess the app. This project looked at the emotional preferences of music for different types of activity and employed emotional cues in a recommender service.

III. EXISTING SYSTEM

It is essential to consider how emotions affect a person's thoughts, behaviors, and emotions. An emotion detection system may be developed by utilizing the benefits of deep learning, and numerous applications, such as feedback analysis and face unlocking, may be carried out with high accuracy. The primary goal of this system is to build a Deep Convolutional Neural Network (DCNN) model that can distinguish 5 (five) different forms of emotional expressions that individuals utilize on their faces. The model is developed, tested, and validated using a hand-gathered image dataset. This gives an accuracy of 78.04 percent.

Although the uses in automatic music production and videography, the issue of music recommendation from dancing movements has not been studied. To solve this problem, the system recommends and assesses a deep music selection algorithm based on dancing motion analysis. For quantitative assessment, this model uses an LSTM-AE-based music recommendation technique that learns the correspondences between motion and music. Comparative testing of the two methods reveals that the motion analysis-based approaches perform noticeably better. Also, a quantitative evaluation of the most appropriate musical genre is proposed.

IV. PROPOSED SYSTEM

The proposed system employs a music recommendation through facial emotion detection using deep learning. To identify an object in an image or a video, the Haar-Cascades approach is chosen. After detecting the object, by using the BiLSTM model the emotion of an inputted image is predicted. The model is used to recognize facial emotion expressions. Recognizing the expression of a person is delicate because everyone cannot express their feelings or passions in the identical way, everyone will have their own fashion. Hence, this work aims to predict emotion accurately. also based on the predicted emotion, the identical music will be recommended. To suit this k- means clustering algorithm is applied. The technique separates the unlabeled dataset into clusters with different attributes, ensuring that each dataset only corresponds to one group. The image is given as input, then the object spotting and image pre-processing can be done. Then the emotion can be predicted, and now the music is recommended based on the predicted emotion. The design is divided into different ways:

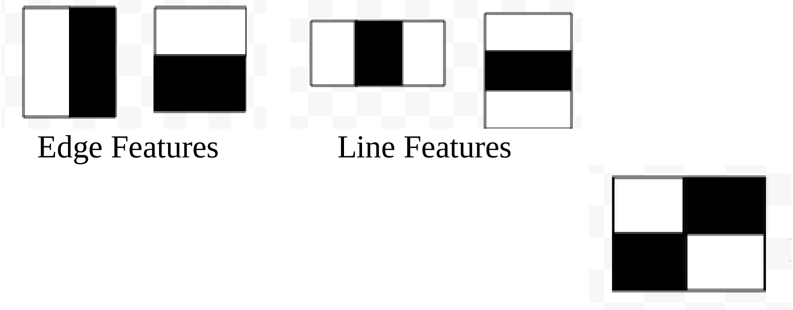
a. Haar Feature Selection:

Most of the human faces exhibit a few traits or similar characteristics that we can recognize or notice, including:

- A deeper area around the eyes than the upper cheeks.
- a brighter area of the nasal bridge than the eyes.

- Some specific regions of the lips, nose, and eyes.

Computation on adjacent rectangular regions at certain points in a discovery window provide the core of a Haar feature. Below are a few illustrations that demonstrate Haar characteristics.



Edge Features

Line Features

Four-rectangle Features

Fig. 1 Types of Haar Features

Feature Extraction:

The feature extraction process will seem like below:

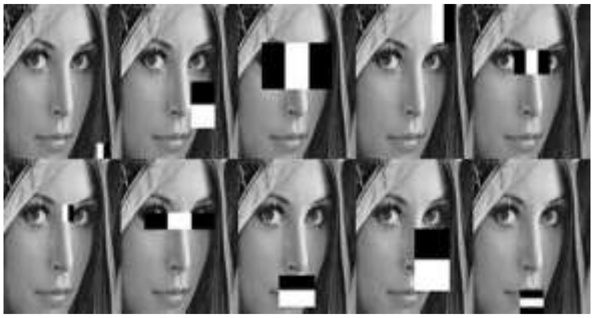


Fig. 2 Feature Extraction

In this example, the first characteristic evaluates the contrast in brightness between the region across the tops of the cheeks and the area around the eyes. Simply adding the pixels in the black area and subtracting the pixels in the white area yields the feature value.

$$\text{Rectangle Feature} = (\text{pixels}_{\text{blackarea}}) - (\text{pixels}_{\text{whitearea}})$$

Determining these characteristics for a husky image can be tricky. Although the number of methods shrinks when applying the integral image, this is where integral visuals are useful.

b. Create Integral Image:

The goal is to turn an input image into an added-area table, where the value at any point p in that table is equal to the sum of all the pixels in the block to the left and above of p, inclusive. Where i(p) is the value of the integral picture pixel at location p and i(p) is the corresponding intensity in the initial image.

$$\text{sum} = I(C) + I(A) - I(B) - I(D)$$

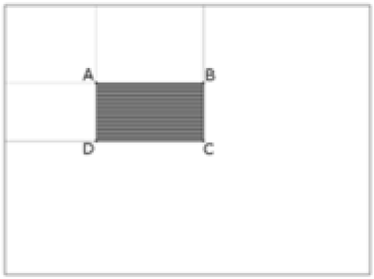
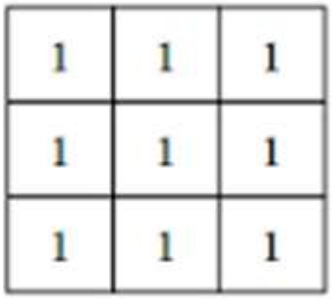


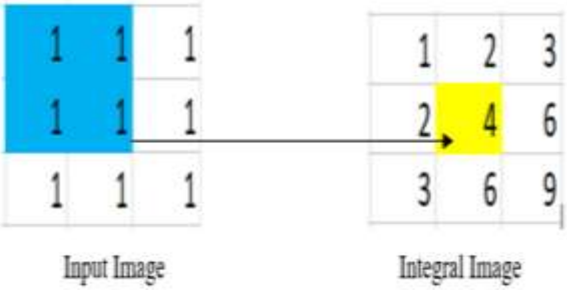
Fig. 3 Integral Image

Consider an example to understand things better. Suppose that our image is represented in the matrix.



1	1	1
1	1	1
1	1	1

Fig. 4 Image in Matrix representation



Input Image

1	1	1
1	1	1
1	1	1

Integral Image

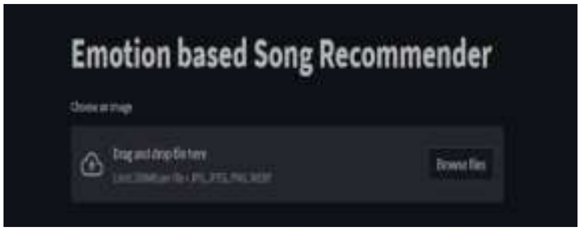
1	2	3
2	4	6
3	6	9

Fig. 5 The Integral Image for the input image

Adding the rectangular region will result in the highlighted region in the Integral Image i.e., 1+1+1+1(highlighted region in input image) = 4 (highlighted region in integral image) and so on.

c. BiLSTM Training:

Recognition of emotions is one of the most investigated disciplines right now. Technology for detecting sentiments or emotions can help humans and machines communicate with one another. Additionally, it will help improve the way decisions are made. There are multiple Deep Learning Models that were developed built to extract emotions from images and videos. But the primary objective of this work is on the Bidirectional LSTM Model. A modification of conventional LSTMs known as Bidirectional LSTMs, or simply BiLSTMs, is used to enhance the model's performance on problems involving sequence classification. Two LSTMs are used in BiLSTMs to train on consecutive input. The initial LSTM uses the input sequence exactly as it is. The second LSTM is applied to a representation of the input sequence in reverse. This adds additional context and speeds up the model. These are the input images:



Emotion based Song Recommender

Choose an image

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG, WEBP

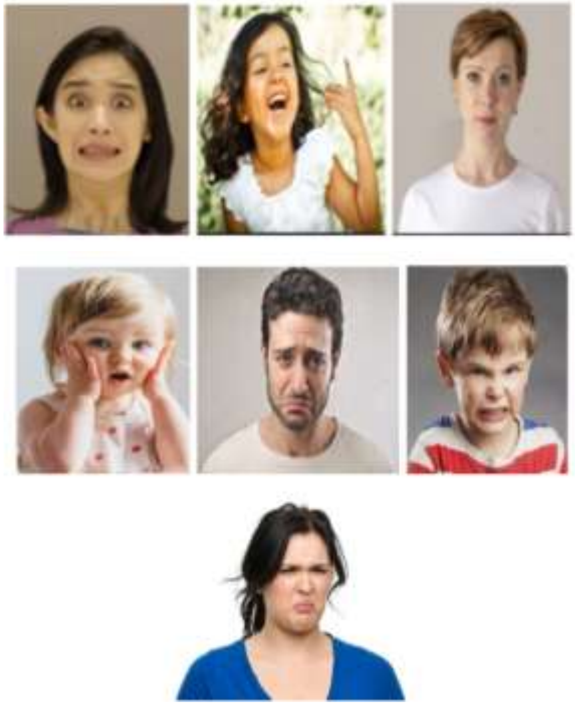


Fig. 6 Input Images

d. Implementing Cascading Classifiers:

The cascade classifier is made from several stages, each of which has a collection of poor individuals. Boosting is used to train weak learners, and the average prediction of all weak learners is utilized to create a highly accurate classifier.

This forecast influences weather the classifier selects to go on to the adjacent zone or chooses to notify its find of an object (positive). (negative). Stages are designed to exclude unsatisfactory samples in as fast as possible because almost all the frames are empty of anything of significance. A low rate of false positives must be maximized because identifying a substance as a non-object will seriously damage your object detection system. The Haar cascade technique is one of the multiple techniques used recently for recognizing objects.

K-means Clustering Algorithm

It is an iterative approach that separates the unlabelled dataset into k distinct clusters, each of which contains just one dataset and shares a set of characteristics. Here, k determines how many pre-defined clusters must be created as part of the process; for instance, if k=2, there will be two clusters, if k=3, there will be three clusters, and so on.

It enables us to divide the data into different groups and offers a workable technique for automatically recognising the groups in the unlabelled dataset without the need for any training.

Because the technique is centroid-based, each cluster has a centroid assigned to it. The main objective of this approach is to minimise the overall distances between each data point and its matching clusters. As an outcome, every group stands apart from the others and has some shared data elements. The picture below explains the k-means clustering algorithm:

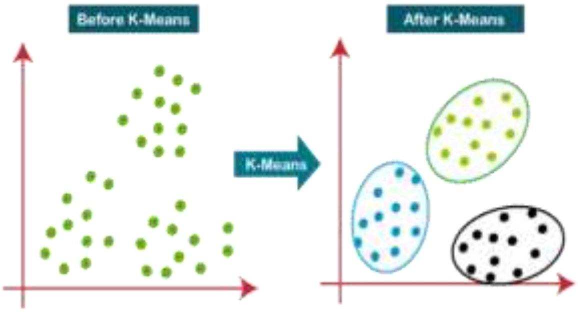


Fig. 7 K-means clustering

The k-means clustering technique has two primary functions:

• selects the ideal value for the k-centre points or centroids using an iterative method.

• The closest k-centre is matched with each data point. The data points that are close to a certain k-centre group together to create a cluster.

Algorithm:

The following stages illustrate how the K-Means algorithm functions:

Step 1: Pick K to get the total amount of clusters.
Step 2: Select K centroids in order or places at random. The input dataset may not reflect that.
Step 3: Construct the K present groups by assigning every point of data to the nearest centroid.
Step 4: Compute the variance and move the centroid of each cluster.
Step 5: Perform the third step to reassign every statistic to the new median for each cluster.
Step 6: If there is a relocation, proceed to phase 4, if not go to FINISH.

Step 7: The finished model.

The K-means clustering algorithm's effectiveness rests on the incredibly effective clusters that it creates. Yet, calculating the ideal number of clusters is a complex process. There are a few additional approaches to choosing the optimum number of clusters, but the best technique is the Elbow method.

ELBOW METHOD:

One of the most often used techniques for determining the ideal number of clusters is the Elbow approach. The WCSS value idea is used in this technique. The term "total variations inside a cluster" is abbreviated as "WCSS," which stands for Within Cluster Sum of Squares. The following formula may be used to get the value of WCSS:

$$\text{WCSS}=\sum\nolimits_{P_{i\ \text{in Cluster1}}} \text{distance}(P_i\ C_1)^2+\sum\nolimits_{P_{i\ \text{in Cluster2}}} \text{distance}(P_i\ C_2)^2$$

The elbow approach can be implemented through the steps listed below to find the optimal value of clusters:

- To use a collected data, it runs K-means clustering for various K values.
- Calculate the WCSS value for each value of K.
- traces a curve from the estimated WCSS values to the K-cluster count.
- When a bend's sharp tip or a plot point resembles an arm, that point is regarded as having the highest K value.

The elbow technique is so named because the graph depicts a steep bend that resembles an elbow. The number of clusters can be chosen to match the amount of data points. The plot's endpoint will be reached if the number of clusters chosen is the same as the number of data points, in which case the WCSS value will be zero.

The graph using the elbow approach resembles the figure below:

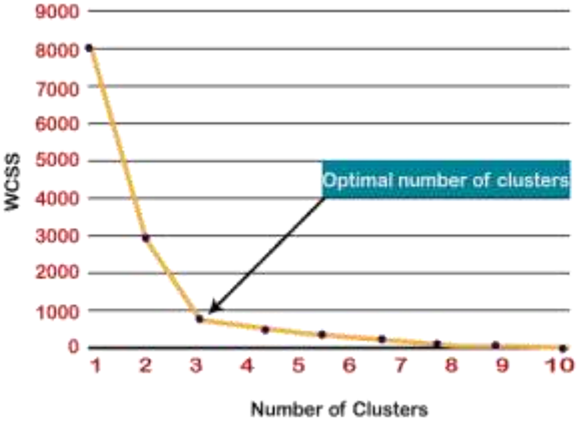
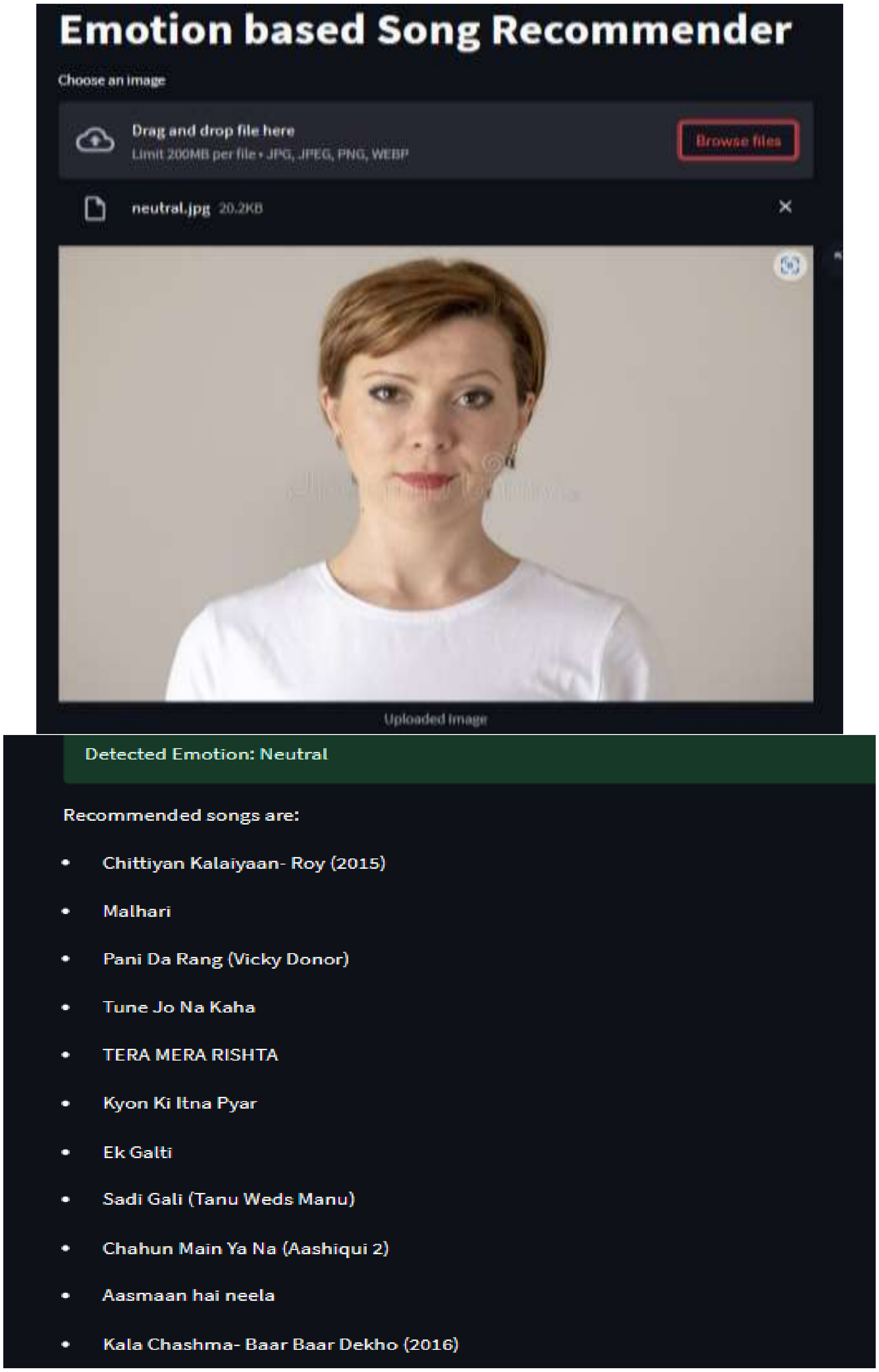


Fig. 8 Graph using the Elbow approach

V. RESULT/OUTPUT:

Uploading the image should be done as input. To find the object, an integral image is created on the uploaded input. The emotion can be categorized among the seven emotions using the BiLSTM technique. To recommend the songs, the K-means clustering algorithm is used. This algorithm will group similar elements or clusters. The emotion-based music recommended to the user.



Emotion based Song Recommender

Detected Emotion: Neutral

Recommended songs are:

- Chittiyan Kalaiyaan- Roy (2015)
- Malhari
- Pani Da Rang (Vicky Donor)
- Tune Jo Na Kaha
- TERA MERA RISHTA
- Kyon Ki Itna Pyar
- Ek Galti
- Sadi Gali (Tanu Weds Manu)
- Chahun Main Ya Na (Aashiqui 2)
- Aasmaan hai neela
- Kala Chashma- Baar Baar Dekho (2016)

Fig. 8 Output

VI. CONCLUSION:

The proposed system results in the classification of seven different facial emotions using some of the deep learning techniques to detect the emotion of the user and retrieve the music genre information by recommending perfect music. A model is created that can be coupled with other electronic devices for efficient control and has equivalent training and validation accuracy, which indicate that the model has the best fit and is generalised to the data. The use of the BiLSTM algorithm reduces the errors to give better accuracy and a computer vision system that automatically recognizes facial expressions with subtle differences.

VII. FUTURE SCOPE:

Future applications of this system have enormous prospects. This strategy can be developed for a big crowd. It is easy as people show their emotions, this technology detects them and recommends music to them. This might also be classified into many other emotions and categorized to people. This work can also be extended with the work of Automatic Pain Detection technology making it available and usable for non-communicative people.

VIII. REFERENCES:

[1] "Akriti Jaiswal, A. Krishnama Raju, Suman Deb, "Facial Emotion Detection Using Deep Learning", The construction of an artificial intelligence (AI) system that can recognize emotions from a person's expressions is shown in the article (INCET), DOI: 10.1109/INCET49848.2020.9154121"

[2] Heechul Jung, Sihaeng Lee, Sunjeong Park, Byungju Kim, Junmo Kim, Injae Lee, and Chunghyun Ahn, Development of Deep Learning-based Facial Expression Recognition System by was presented at the 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision.(FCV), DOI: 10.1109/FCV.2015.7103729"

[3] "Wenjun Gong, Qingshuang Yu, "A Deep Music Recommendation Method Based on Human Motion Analysis", IEEE Access, Volume: 9, PP(s): 26290 – 26300, 05 February 2021, DOI: 10.1109/ACCE SS. 2021.3057486"

[4] "Yuedong Chen, Jianfeng Wang, Shikai Chen; Zhongchao Shi; Jianfei Cai, Using Facial Motion Prior Networks for Face Emotion Recognition, 2019 IEEE Visual Communications and Image Processing (VCIP),DOI:10.1109VCIP4 7243.2019.8965826"

[5] "Shan Li, Weihong Deng, "Deep Facial Expression Recognition: A Survey", IEEE Transactions on Affective Computing, Volume: 13, Issue: 3, 01 July-Sept. 2022,DOI: 10.1109/TAFFC.2020.2981446"

[6] "Jiannan Yang, Tiantian Qian, Fan Zhangg, Samee U. Khan, "Real-Time Facial Expression Recognition Based on Edge Computing", IEEE Access, Volume: 9, PP(s): 76178 – 76190, 21 May 2021, DOI: 10.1109/ACCESS.2021.3082641"

[7] "S L Happy, Aurobinda Routray, "Automatic Facial Expression Recognition Using Features of Salient Facial Patches," IEEE Transactions on Affective Computing, Volume: 6, 01 Jan.-March 2015, DOI: 10.1109/TAFFC.2014.2386334"

[8] "Hari Prasad; P. Swarnalatha, "Facial expression detection using facial expression model," 017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS),DOI: 10.1109/ICECDS.2017.8389644"

[9] "Xuan Zhu, Yuan-Yu an Shi; Hyoung-Gook Kim; Ki-Wan Eom, "An Integrated Music Recommendation System", IEEE Transactions on Consumer Electronics (Volume: 52, Issue: 3, August 2006), DOI: 10.1109/TCE.2006.1706489"

[10] Thanapong Khajontantichaikun, Saichon Jaiyen, Siam Yamsaengsung, Pornchai Mongkolnam, and Unhawa Ninrutsirikun, International Computer Science and Engineering Conference (ICSEC), Volume: 14, Issues: October 27, 2020, "Emotion Detection of Thai Elderly Facial Expressions using Hybrid Object Detection," DOI: 10.1109/TAFFC.2020.3034215"