ANN Based Facial Emotion Detection and Music Selection

Dr. Merin Thomas
Associate Professor
Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jain (Deemed-to-be University)
Bangalore, India,
merin.thomas@jainuniversity.ac.in

Shreenidhi H S,
Asst. Prof.
Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jain (Deemed-to-be University)
Bangalore, India,
hs.shreenidhi@jainuniversity.ac.in

***Abstract*:Everybody like to listen to music, depending on their mood. But choosing a music manually based on mood is a task that has to be addressed because it takes time and effort. When anticipating a person's emotions and mood, the face is crucial. We create a prototype for a dynamic music recommendation system based on human emotions in this system that is being suggested. Songs for each emotion are taught based on human listening patterns. The emotion on a real person's face is recognised using an integration of feature extraction and machine learning techniques. Once the mood is determined from the input image, the appropriate music will be played to keep the users' attention. It consists of two stages: training and real-time emotion recognition and music selection. The suggested study has demonstrated a notable level of performance in terms of recognition and music choice.**

Keywords—, emotion recognition, facial recognition.

I. Introduction

The human face is an important organ of the body, and it plays an especially vital function in determining an individual's behaviour and emotional condition. It is a highly laborious, time-consuming, and time-consuming operation to manually segregate the list of songs and generate a suitable playlist depending on an individual's emotional qualities. However, this is something that must be done.

A number of different algorithms have been proposed and developed in order to automate the process of creating playlists. However, the proposed existing algorithms that are now being used are computationally slow, have a lower level of accuracy, and at times even necessitate the usage of supplementary gear such as EEG or sensors. This proposed method, which is based on the extraction of facial expressions, will automatically build a playlist, hence decreasing the amount of time and effort that are required to render the process manually.

The integration of Facial Emotion Recognition (FER) and Music Information Retrieval (MIR) into conventional music players made it possible to automatically categorise the playlist according to a wide range of feelings and states of mind. MER is a technique that is used to identify a facial extraction that has been received by taking into consideration the numerous facial features of the extracted face and how they correspond to different categories of feelings and states of mind. Even though both MER and MIR featured the capabilities of avoiding the manual segregation of songs and development of playlists, it is still unable to fully incorporate a human emotion-controlled music player. This is the case despite the fact that both MER and MIR included these capabilities. Even though human voice and gesture are typical ways of communicating feelings, the most natural and ancient mer of doing so is through facial expression. Feelings, emotions, and mood may all be communicated through the face.

Fear, disgust, anger, surprise, sad, happy and neutral, and a state of neutrality are the basic human emotions that can be broken down into more specific categories. These feelings can also be used as an umbrella term to describe a wide range of different states of mind, including concept and cheerfulness, amongst others. These feelings are perfectly suited for the situation. Because of this, facial muscular contortions are quite subtle, yet being able to differentiate between them can result in a wide range of expressions. Because of how strongly an emotion is influenced by its surrounding context, different people — or even the same people — may express the same feeling in different ways..

Machine learning and neural networks have shown promising outcomes when applied to such categorised tasks. Machine learning algorithms have already been put to good use in the fields of pattern identification and classification; this suggests that they may also be applicable to the detection of emotional states. Because of the rise of digital music, it is crucial to create a system that can propose songs based on individual tastes.

II. Literature Survey

Interaction with a human being is necessary for the traditional method[5] of playing music that corresponds to a person's emotional state. The transition to technologies based on computer vision will make the automation of such systems possible. In order to do this task, an algorithm is utilised to categorise human expressions, and then a music track is played in accordance with the current emotion that has been identified. It saves time and effort that would otherwise be required to manually browse through a collection of songs in order to find one that corresponds to a person's current mood. The facial features of a person are extracted with the use of an algorithm called Principle Component Aanalysis and a classifier called Euclidean Distance. This allows the expressions of the person to be recognised. When compared to alternative ways, the use of an integrated camera to record a person's facial expressions results in a reduction in the amount of money spent on the system's design.

A wide variety of human experiences and emotions can be captured and understood through the medium of song. Emotion-based classification systems that can be relied

2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC) | 978-1-6654-5687-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/IIHC55949.2022.10060593

978-1-6654-5687-6/22/$31.00 ©2022 IEEE

upon would greatly aid in understanding their significance. Studies attempting to categorise music according to how it makes a listener feel have not been particularly successful thus far. Users shouldn't have to spend a lot of time manually selecting and reordering songs when using a playlist to do so. Automatically creating a playlist based on the user's mood is made possible by a new technology that uses facial expressions to determine what kind of music will make them feel good.

One of the most fascinating and mysterious aspects of music [10] is its capacity to evoke feelings in its audience. Music has the power to not only change the listener's emotional state but also their physical state. In this paper, we'll look at a number of different classification-based algorithms in order to outline a clear methodology for doing two things: I categorising songs into four distinct moods, and ii) detecting a user's mood based on facial expressions so that a playlist can be generated specifically for that person. The geometry or principal promiment points of key facial features like the lips and eyes were the only things taken into account by the feature extraction method that relied on geometry.

Over the past few years, as a result of the development and application of big data, deep learning has garnered an increasing amount of attention. Convolutional neural networks, which are a type of deep learning neural networks, are an extremely significant component in facial picture identification. In this paper, a combination of the micro-expression identification technology of convolutional neural networks and the automatic music recommendation algorithm is built to identify a model that identifies facial micro-expressions and suggests music according to related moods.

A novel artificial neural network based facial emotion detection system have been proposed in the real time and based on the mood of the person and music are recommended based on the detection of the mood.

III. PROPOSED SYSTEM

Real-time mood recognition is the main goal of the application known as the mood-based music recommendation system. It is a prototype for a brand-new product that has two primary modules: music suggestion and facial expression recognition/mood detection.

A. *Facial Mood Detection*

This process consists of the two stages:

• ***Face Detection :***

A camera is utilised in order to record an image of the user as they appear in real time. Once the photo has been taken, the frame of the image that was acquired from the webcam feed is transformed to a grayscale image in order to increase the performance of the classifier that is used to identify the face that is present in the picture. After the conversion has been finished, the image is delivered to the classifier algorithm, which, with the assistance of feature extraction techniques, may extract the face from the frame of the web camera stream. The extracted face is broken down into its component parts, which are then fed into an artificial neural network () model that has been trained to identify the user's emotional state. The classifier will be trained using these images so that when it is presented with an entirely new and unknown set of images, it will be able to extract the position of facial landmarks from those images based on the knowledge that it had already acquired from the training set and return the coordinates of the new facial landmarks that it detected. These images will be used to train the classifier so that it will be able to do so successfully.

• ***Mood Detection*** **:**expression of an emotion on the face can be classified as either happy, angry, sad, neutral, surprise, fear or disgust. MobileNet, an architecture model for Image Classification and Mobile Vision, is utilised here for the purpose of carrying out this assignment. Running MobileNet or applying transfer learning to it takes an extremely low amount of computation power. Because of this, it is an excellent choice for mobile devices, embedded systems, and computers with limited computing efficiency, without sacrificing the quality of the findings in any way. Convolutions that are depth-wise separated are used in its construction of lightweight deep neural networks. Combining the FER 2013 dataset and the MMA Facial Expression Recognition dataset from Kaggle resulted in the creation of the dataset that was utilised for training purposes. The FER 2013 dataset included grayscale images that were 48 pixels on a side and comprised 48 total pixels. The dataset used for MMA Facial Expression Recognition contained photos that varied in terms of their criteria. Therefore, all of these photos were processed in the same mer as the images in the FER 2013 dataset, and then they were pooled to produce an even larger dataset that included 20,045 images for training and 7,724 images for testing. In order to train and validate our model, MobileNet was combined with Keras. for seven classes -happy, angry, neutral, sad, surprise, fear and disgust. The proposed block diagram is shown in the figure 1.

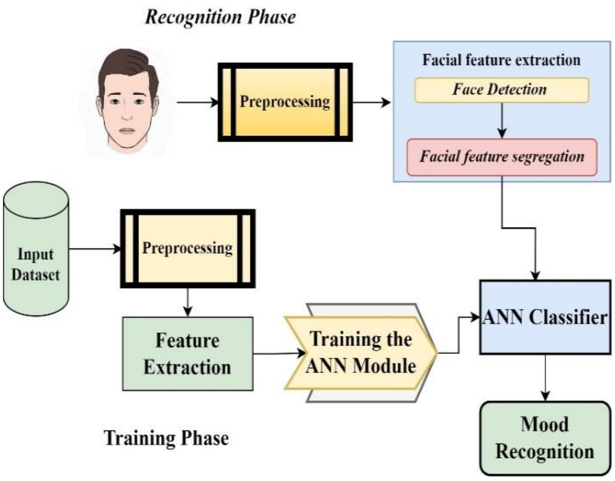

Fig 1:Facial Mood training and recognition

B. *Music Recommendation Module*

Both Hindi and English versions of the dataset including songs that had been categorised according to their mood were discovered on Kaggle. An investigation was carried out to find a reliable cloud storage platform that could save, retrieve, and query this music data in response to specific user requests.Once the facial expression is identified by classifier, it is considered as the mood of the person, based on those songs stored in the database or selected and played as shown in the figure 2.

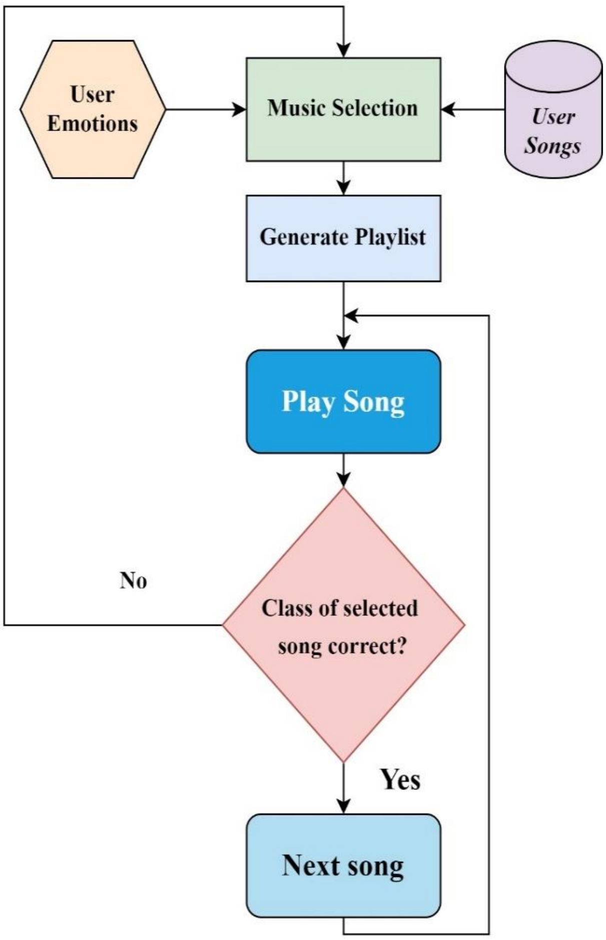

Fig 2: Music recommendation Module

IV. SYSTEM ARCHITECTURE

flow diagram represents the functional formation of the music player as shown in Figure 3.

Initially the model is trained by using the input dataset and the preprocessing model takes place and then the model trains the dataset and then it tests the dataset and captures the facial expression using the webcommand the emotion detection is formed using themodel. The model classifies the facial expressions whereas, happy, sad, angry, depressed etc. and it analyses the exact emotion and plays the music. A behavioral diagram is the same thing as an activity diagram, which means that it displays the behavior of a system. An activity diagram depicts the control flow from a starting point to an ending point, outlining the many decision routes that are available while the activity is being carried out and demonstrating how they connect to one another.

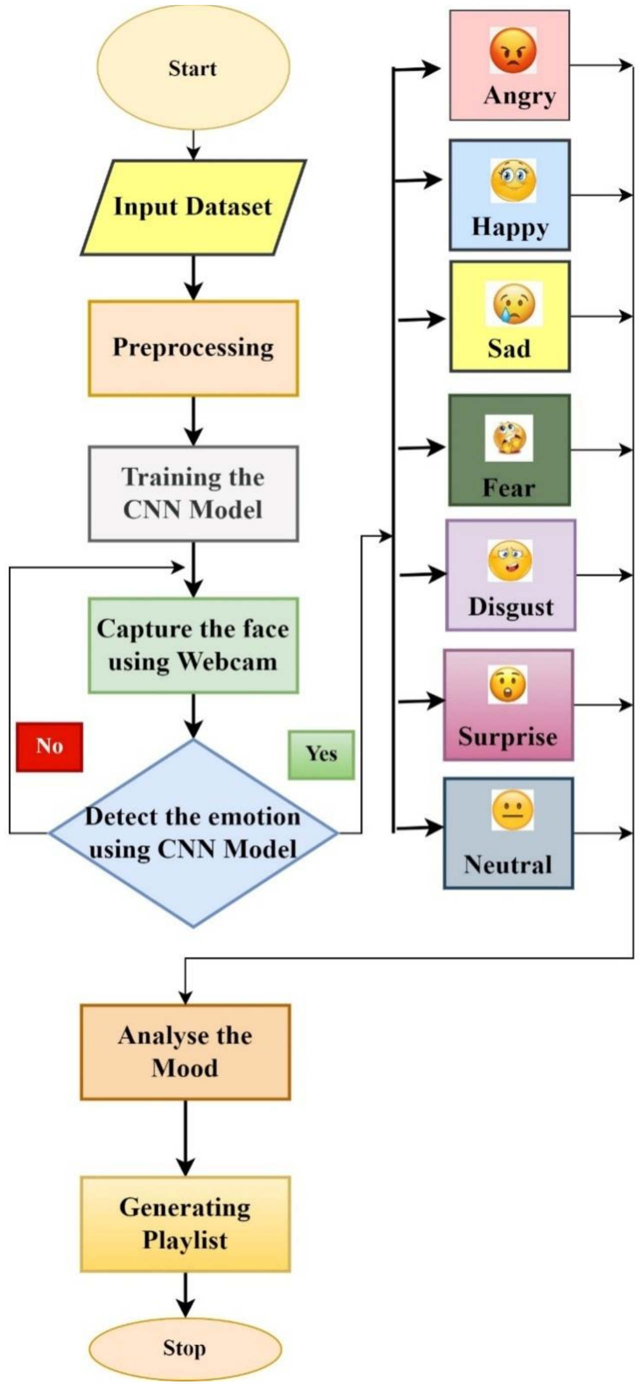

Figure 3: Proposed framework

V. IMPLEMENTATION

For the purposes of training our ANN architecture, we collected data from the FER2013 database. It was developed with the help of the Google image search API, and it was presented at the ICML 2013 Challenges. The sizes of the faces in the database have been automatically standardized to 4848 pixels. The FER2013database comprises a total of 35887 photos tagged with 7expressions (28709 training images, 3589validation images, and 3589 test images). There are a set number of pictures that are used to depict each feeling. What it all means, according to ANN (ANN): We took live frames from a web camera by utilizing the Open CV library, and then we used the HarrCascades approach to identify the faces of the pupils. Adaboost is the learning algorithm that forms the foundation of Haar Cascades.

Algorithm:

Training phase

Select the FER2013 database

Preprocess the Dataset

Read the images of N x N

Resize the images

Extract the facial features

Select training set of M number of sample images

Train the ANN model using Adaboost learning algorithm

Testing Phase

Capture the real-time image using webcam

Process the image to extract the facial characteristics

Detect the emotion and using HarrCascades

Predict the emotions

Select the songs based on user interest and predicted emotions.

VI. EXPERIMENT RESULT AND DISCUSSION

In the work that has been proposed, the Adaboost learning method has been built to select a number of significant features from a huge set, and the Image Data Generator class in the Keras library has been implemented to do image augmentation.

These parameters are utilized in the process: rotation range = 10, width shift range = 0.1, zoom range = 0.1, height shift range = 0.1, and horizontal flip = True. The ANN model that is utilized has a total of two fully connected layers, four pooling layers, and four convolutional layers. Additionally, the ReLU function is used to provide non-linearity in the ANN model, in addition to batch normalization to normalize the activation of the preceding layer at each batch and L2 regularization to apply penalties to the model's various parameters. Both of these regularization techniques are used in conjunction with the ANN model. The Convolutional Neural Network model is trained using the FER 2013 database, which includes seven emotions (happiness, anger, sadness, disgust, neutral, fear and surprise) Before being input into the ANN model, the pictures of the detected faces were rescaled to be 4848 pixels wide and converted to grayscale.

Table 1 gives the confusion matrix of the recognition of the 7 class of the facial expression.

Based on the recognition of the mood of the person, the songs will be selected and played.

CONCLUSION

An important step forward in understanding human behaviour has been developed as part of the proposed effort. 7 common human emotions are explored here. An Emotion-Based Music Player satisfies the need to classify musical selections in accordance with emotional states. Both a training and a testing phase make up the suggested system. In the training phase, the Adaboost learning algorithm is utilised to train on the FER2013 database. HarrCascades is used to analyse face expressions in real time, allowing for the identification of seven distinct human emotions and the subsequent selection of an appropriate soundtrack.

REFERENCES

[1] D.Schnaufer,B.Peterson,"Realizing5Gsub-6-GHzmassiveMIMOusing GaN," Microwave. & RF, pp. 4–8, 2018.

[2] E.O'Connelletal.,"Challengesassociatedwithimplementing5Ginmanufacturing," Telcom, Vol. 1, no. 1, pp. 48-67, 2020.

[3] C.A.Balanis,AntennaTheoryAnalysisandDesign,Fourthedition.,Hoboken, New Jersey, John Wiley & Sons, Inc., 2016.

[4] A.Tagguetal.,"Adualbandomni-directionalantennaforWAVEandWi-Fi,"20172ndInternationalConferenceonCommunicationSystems,Computing and IT Applications (CSCITA),IEEE, pp. 1-4, 2017.

[5] D. Nataraj, G. Karunakar, "Design and research of miniaturized microstrip slot with and without defected ground structure," Int. J. Recent Technol. Eng., vol. 8, no. 2, pp. 391–398, 2019.

[6] Y. Liu et al., "Some recent developments of microstrip antenna," Interna- tional Journal of Antennas and Propagation, vol. 2012, 2012.

[7] R. Garg et al., Microstrip Antenna Design Handbook, Artech house, 2001.

[8] A. Pandey, Practical Microstrip and Printed Antenna Design, Artech House, 2019.

[9] R. Mondal et al., "Compact ultra-wideband antenna: improvement of gain and FBR across the entire bandwidth using FSS," IET Microwaves, Antennas & Propagation, vol. 14, no. 1, pp. 66-74, 2019.

[10] Z. Xing, K. Wei,L. Wang, and J. Li, "Dual-band RFIDantenna for 0.92 GHz near-field and 2.45 GHz far-fieldapplications,"*InternationalJournalofAntennasandPropagation*, vol. 2017,pp. 1-9,2017.

[11] A. Roy, S. Bhunia, D. C. Sarkar, and P. P. Sarkar, "Slotloaded compact microstrip patch antenna for dual bandoperation," *Progress in Electromagnetics Research C*, vol.73,pp.145–156, 2017.

[12] M. M. Hasan, M. R. I. Faruque, and M. T. Islam, "DualbandmetamaterialantennaforLTE/Bluetooth/WiMAXsystem," *ScienceReports*,vol.8,no.1240,pp.1-17,2018.

[13] A. Salam, A. A. Khan, and M. S. Hussain, "Dual bandmicrostrip antenna for wearable applications," *Microwaveand Optical Technology Letters*, vol. 56, no. 4, pp. 916-918,2014.

[14] M.K.Khandelwal,B.K.Kanaujia,andS.Kumar,"Defected ground structure: Fundamentals, analysis, andapplicationsinmodernwirelesstrends,"*InternationalJournal of Antennas and Propagation*, vol. 2017, pp. 1-22,2017.

[15] F. Y. Zulkifli, E. T. Rahardjo, and D. Hartanto, "Mutualcouplingreductionusingdumbbelldefectedgroundstructureformultibandmicrostripantennaarray,"*Progress In Electromagnetics Research Letters*, vol. 13,pp.29-40, 2010.

[16] S. S. Kumar, S. H. Bharathi and M. Archana, "Non-negative matrix based optimization scheme for blind source separation in automatic speech recognition system," 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2016, pp. 1-6, doi: 10.1109/CESYS.2016.7889860.

[17] N. P. Yadav, "Plus shaped notch loaded rectangular patchwith D.G.S. antenna for multiband operation," in *Proc.Antenna Test & Measurement Society (ATMS India-16)*2016,pp. 1-5.

[18] D.Fistum,D.Mali,andM.Ismail,“Bandwidthenhancementofrectangul armicrostrippatchantennausing defected ground structure,” *Indonesian Journal ofElectrical Engineering and Computer Science*, vol. 3, no.2, pp.428-434,2016.

[19] A. Zaidi et al., "High gain microstrip patch antenna, with PBG substrateand PBG cover, for millimetre wave applications," 2018 4th InternationalConference on Optimization and Applications (ICOA), IEEE, pp. 1-6,2018.

[20] S. S. Kumar, B. K. Aishwarya, K. N. Bhanutheja and M. Chaitra, "Breath to speech communication with fall detection for elder/patient with take care analytics," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2016, pp. 527-531, doi: 10.1109/RTEICT.2016.7807877.

	Neutral (%)	**Sad (%)**	**Fear (%)**	**Anger (%)**	**Disgust (%)**	**Happy (%)**	**Surprise (%)**
Neutral	92.35	1.012	2.01	2.125	1.385	0	1.118
Happy	2.15	0	0.25	0.127	0	97.25	0.223
Surprise	0	1.2	0.012	0	0.005	1.2	97.583
Fear	1.2	0.825	95.023	0	1.75	0	1.202
Anger	2.53	1.3	1.385	93.432	0.123	0	1.23
Sad	3.234	92.85	1.35	0.85	0.12	1.49	0.106
Disgust	0.025	1.0123	0.356	0.61	97.7077	0	0.289

Average Recognition = 95.17081%