



EMOTION MAPPING BASED MUSIC RECOMMENDATION SYSTEM USING MACHINE LEARNING

¹Ayush Raj Singh, ²Aakash Chauhan, ³Nikita Samtrai, ⁴Tina Rajpal, ⁵Sunita Suralkar

^{1, 2, 3, 4} Student, ⁵ Assistant Professor

Department of Computer Engineering,

Vivekanand Education Society's Institute of Technology, Chembur, Mumbai-400074, India

Abstract: Emotion Mapping based Music Recommendation System is to provide users with suggestions that match their emotions and to assist them accordingly. The image is first captured and then converted to an emoji, and future analysis is done through the emoji's face. Analyzing a user's emotional facial expressions can help them understand the subject's current emotional or mental state. Music is an area that is likely to change a person's mood. It is well known that people use facial expressions to express what they want to say and the meaning of words more clearly. By developing an emotion mapping system, users can determine their mental state, and if they are unwell, they can change their mood by listening to pop-up messages containing songs. User's facial features are captured with the help of a webcam. By combining the user's photographs and emotion, the appropriate analysis is done, and the songs are displayed.

Index Terms: Sentiment, Emotion Analysis, Music Recommendation, Emoji, Human face, Neural Networks, CNN, RNN.

I. INTRODUCTION

Emotion Mapping based Music Recommendation System using Machine Learning is a device wherein a human face is scanned and concurrently their face is transformed into an emoji. Emojis are an indispensable part of regular communication for expressing emotions. People have a tendency to express their feelings, specifically through their facial expressions. By taking pictures and recognizing the emotion of someone and playing songs for their mood can increasingly calm their thoughts and universally emerge as giving a pleasant effect.

The project aims to capture the emotion expressed through facial expressions. Emoji faces, which are ubiquitous in our daily communication are designed to support emotional communication so from an emojiified image, human sentiments are detected and accordingly music is played with the notification stating the person's current mood being scanned. The song played corresponds to the perceived emotion.

The main objective of the project is to change a person's emotion as music plays an important role in changing their mood. If the emotion detected is happy, the user will be redirected to a website which will show a playlist containing happy songs. The system is designed to capture human emotions through the webcam interface available on computer systems. The software uses image segmentation to capture the user's image. For scanning the person's face, we will be using the OpenCV library. It allows the use of Machine learning algorithms to search for faces within a picture. When it comes to emotion detection, the basic task of any emotion analysis program is to isolate the polarity of the input (facial expression) to understand whether the primary emotion presented is positive, negative, or neutral. With the use of sentimental analysis and listening to songs, we can avoid many health risks, and improve our mood.

II. OBJECTIVE

The objective of our paper is to implement machine learning ideas in a system that scans human faces and transforms them into quantifiable emotions to monitor the user's mood. This will be instrumental in recommending music to them, to change the user's emotion as music plays an important role in changing a person's mood, to provide an interface between the music system and to bridge the gap between emotion analysis and music techniques.

III. LITERATURE REVIEW

Emotion mapping is an essential task for health care decision makers for rapid decision- making. An accurate emotion mapping model helps determine the type of action to perform while assisting the subject. There are different approaches to emotion mapping. In this review, we looked at what was done in the literature on emotion mapping.

S. Metilda Florence and M. Uma proposed the music recommendation system [1]. The main goal of their music recommendation system is to provide users with suggestions based on their tastes. Analyzing a user's facial expressions / emotions can lead to an understanding of the user's current emotional or mental state. It is well known that people make use of facial expressions to express what they want to say and the meaning of words more clearly. Developing a recommender system helps users decide what type of music they should listen to, and helps them reduce their stress levels.

Users don't have to waste time searching for songs or looking up songs based on their mood, as the best track would be shown to them according to their requirements. The captured images of the user will help in identifying the songs/playlist. The system is still not able to record all the emotional states correctly due to the lack of images in the image dataset. The image that is fed into the classifier should be taken in a well-lit atmosphere for the classifier to give accurate results.

H. Immanuel James, J. James Anto Arnold, J. Maria Masilla Ruban Sara [2]. In this paper, the focus is on detecting human emotions for developing emotion-based music players, which approaches have been developed in the available music players to detect emotions, which method their music player follows to detect human emotions, and how it is better to use their system for emotion detection. It also briefly describes playlist generation, and emotion classification. By using a pycharm analysis tool, they have developed a software that understands user emotion based on facial expression. They have integrated Python code into the web service, allowing them to play music based on facial expressions.

ShanthaShalini.K, Jaichandran.R, Leelavathy.S, proposed a music recommendation system based on user emotions [3]. The face is an important aspect in assessing human emotions and moods. Emotions are extracted with the help of a camera. The face is given as an input to the process of recognizing facial emotions, and the music is played automatically based on the emotions. This system develops a prototype of recommendation, a dynamic music recommendation system based on human emotions. Based on all human listening patterns, songs are trained for all emotions. Feature extraction and machine learning techniques have been integrated. Once the mood is derived from the input image, the user will play a song that suits the particular mood. This system provides a high level of accuracy on real facial images. The Pygame package is used to interpret the sound libraries in the python programming language.

Diah Anggraeni Pitalokaa, Ajeng Wulandaria [4] implemented the proposed method and compared the pre-processing methods for facial expression recognition. Based on the experimental results obtained, face detection and cropping to capture the regions of interest has been declared to achieve the best improvement in CNN performance. The global contrast normalization step contributes more accuracy than other normalization techniques to accuracy but not as good as getting the ROI. Graph Convolutional Network (GCN) tries to reduce the distribution of data so that different contrast values are not present. The proposed CNN model works better at 32x32 and 64x64 resolutions. It seems the capacity of the model satisfies the complexity task for facial expression recognition on those resolutions. Performance of CNN can be boosted using data augmentation like combining data from step cropping and adding noises.

Fang-Fei Kuo et al and Suh-Yin Lee et al.[5] With the growth of digital music, the development of music recommendations is helpful for users. The existing recommendation approaches are based on the user's preference for music. However, sometimes, recommending music according to the emotion is needed. In this paper, they have proposed emotion-based music recommendation, which was based on the association discovery from film music. They investigated the music feature extraction and modified the affinity graph for association discovery between emotions and music features. Experimental results show that the proposed approach achieves an accuracy of 85% on average.

Anukriti Dureha et al [6]. In it, he suggested manually separating playlists and annotating songs according to the user's current emotional state. This is laborious and time consuming. Numerous algorithms have been proposed to automate this process. However, because existing algorithms were slow, the use of additional hardware (such as EEG systems and sensors) increased the cost of the entire system but had less accuracy. This paper presents an algorithm that automates the process of generating audio playlists, based on the user's facial expressions, saving the time and effort spent manually running the process. The algorithm proposed in this paper aspires to reduce the total computational time and the cost of the designed system. It also aims to improve the accuracy of the designed system. The facial expression recognition module of the proposed algorithm is validated by testing the system against user-dependent and user-independent datasets.

IV. IMPLEMENTATION DETAILS

The purpose of this work is to detect emotions and select music to play based on the detected emotions. Human emotions can be expressed through music. We have developed an application that detects the user's emotions, displays the song according to their mood, and searches for songs by name.

For developing our project, we have used Tensorflow, Keras and Python libraries like tkinter, numpy, os, PIL, threading, and OpenCV.

Finally, the identified emotion will be used to select multiple songs or a suitable playlist from the playlists dataset or if the user is online songs can be fetched from an API and then they can be played. Also, one of the features of the application will be

that the user's captured image will be cartoonized and displayed along with emotion detected.

Our project's block diagram (Fig. 1) can be broken down into 2 important blocks:

- User Input
- Music File

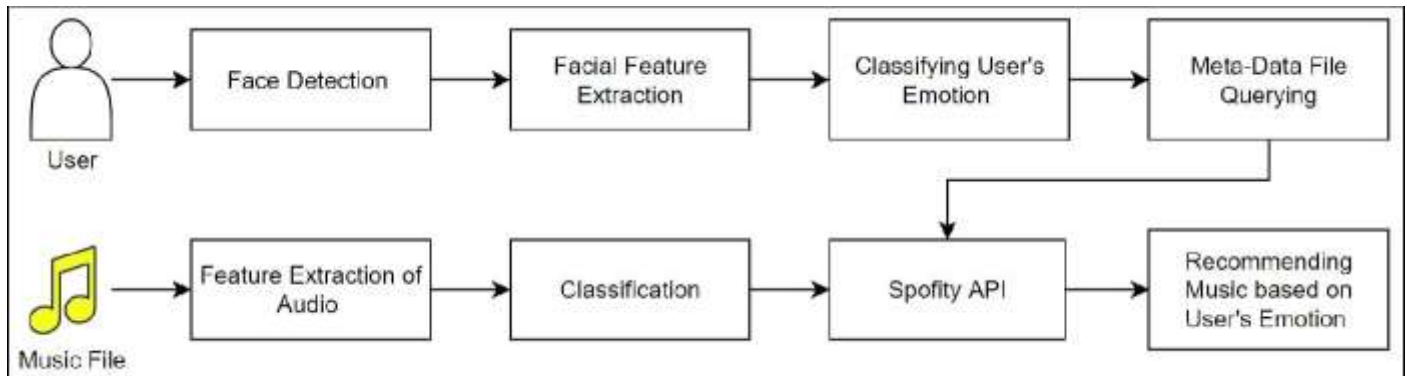


Fig. 1 Block Diagram

4.1 User Input

The human face plays an important role in determining a person's mood. The camera is used to get the required input from the human face. The input surface is then filtered, features or properties are checked, and unwanted features are removed. One of the uses of this input is to extract information to infer a person's mood.

4.2 Music File

The other half of the block diagram belongs to the Music Files. The "emotion" derived from the input provided earlier is used to get a list of songs. Here, different music files' features are extracted and classified to pass it to the database. After all this process, the appropriate music gets recommended according to the user's emotion.

It reduces the hassle of manually segregating songs and grouping them into different lists, helping to create the right playlist based on a person's emotional characteristics. Functions that are positively filtered are sent for parameter classification.

a) Dataset used

The dataset which we have used in this project is FER-2013. It contains seven facial expressions namely: Happiness, Neutral, Sadness, Anger, Surprise, Fear, Disgust



Fig. 2: Pie chart representing distribution of sample images

The dataset contains a total of 35,685 samples of grayscale images of faces and size of each image is 48x48.

For the prediction model, we have considered the crops majorly consumed in India, like, Rice and Wheat. The weather dataset from weather API. It allowed us to query the AQICN's database for the user-queried State and return the current weather data in the region which is used as the input to Gradient Boosted Regressor Model.

4.3 Machine Learning Model

Convolutional Neural Network is the model we chose for our proposed system as it works by extracting features from images. This eliminates the need for manual feature extraction i.e the features are not trained. They're learned while the network trains on a set of images.

CNNs learn feature detection through tens or hundreds of hidden layers. Each layer increases the complexity of the learned features. This made our model extremely accurate for image classification. It provided us with the highest accuracy among the models we considered (ConvLSTM, VGG-19 and ResNet).

V. METHODOLOGY EMPLOYED

5.1 Algorithm Used

Convolutional Neural Network is a Deep Learning algorithm which can take in an input image, assign importance to various aspects/objects in the image and be able to differentiate one from the other.

CNNs have an input layer, an output layer, and hidden layers. The hidden layers usually consist of convolutional layers, ReLU layers, pooling layers, and fully connected layers.

In a convolutional layer, neurons only receive input from a subarea of the previous layer. In a fully connected layer, each neuron receives input from every element of the previous layer.

Convolutional layers apply a convolution operation to the input. This passes the information on to the next layer. Pooling combines the outputs of clusters of neurons into a single neuron in the next layer. Fully connected layers connect every neuron in one layer to every neuron in the next layer.

5.2 Model Training

Emotion Extraction Module: The user's image is captured using the camera / webcam. Once the image is captured, the captured image frame is converted from the webcam feed to a grayscale image, improving the performance of the classifier. It is used to identify the face present in the image. When the conversion is complete, the image is sent to the classification algorithm. The classification algorithm can use feature extraction technology to extract faces from the frames in webcam feeds.

Individual features are retrieved from the extracted face and sent to the trained network to recognize the emotions expressed by the user. These images are used to train the classifier. This allows us to extract the location of facial landmarks from those images based on the knowledge we had already acquired from the training set when the classifier is presented with an entirely new and unknown image set. It returns the coordinates of the newly recognized face landmarks. The network is trained using the dataset. In this way, the emotions expressed by the user are identified.

Audio Extraction Module: After the emotion of the user is extracted, the music/audio clip based on the voiced emotion is displayed to the user i.e. a list of songs based on the emotion is displayed, and the user can listen to any song.

Based on the regularity that the user would listen to the songs are displayed in that order. For example, if the emotion or facial feature is categorized under happy, then songs from the happy database will be displayed to the user.

The dataset obtained is made to undergo preprocessing to make it suitable for model training. After preprocessing, the dataset is split into 70% training and 30% test dataset. Different machine learning algorithms such as Convolutional Neural Network (CNN), Convolutional LSTM (Long Short Term Memory) Network (ConvLSTM), VGG-19, and Residual Neural Network (ResNet) are used to determine which algorithm shows the best accuracy.

Early Stopping Function: We trained our 4 neural network models for the same amount of time. When we trained our networks, there came a point during training when the models stopped generalizing and started learning the statistical noise in the training dataset.

Such overfitting of the training dataset results in an increase in generalization error, making the model less useful at making predictions on new data.

Training the models came with the challenge to train the network long enough to learn the mapping, but not so long that it overfits the training data.

The models were evaluated on a holdout validation dataset after each epoch. If the performance of any model on the validation dataset starts degrading (e.g. loss begins to increase or accuracy begins to decrease), then the training process stops. The model at the time that training is stopped is then used and is known to have good generalization performance.

After choosing the appropriate algorithm the dataset is trained and later the data modelling is performed. Among our 4 models, CNN performed the best. The image recognition has led to an accuracy of 88.40% with 55 epochs.

V. RESULT AND DISCUSSION

This paper demonstrates the use of various neural network algorithms: Convolutional Neural Network (CNN), Convolutional LSTM (Long Short Term Memory) Network (ConvLSTM), VGG-19, and Residual Neural Network (ResNet), for emotion mapping of users.

A comparative study was performed using four different neural network algorithms to enhance accuracy. After training all models, the accuracies of the different models were compared. Convolutional Neural Network tops the list with an accuracy score of 0.884 and 55 epochs, followed by Residual Neural Network at 0.7804 and 26 epochs.

As mentioned earlier, because of the Early Stopping Function, all of the algorithms stopped at different numbers of epochs to avoid the over-fitting of the training dataset.

Table 5.1: Performance Comparison of Different Algorithms

Model	Accuracy Score	Epochs
Convolutional Neural Network	88.40%	55
Residual Neural Network	78.04%	26
Convolutional LSTM Network	53.64%	12
VGG-19	34.29%	5

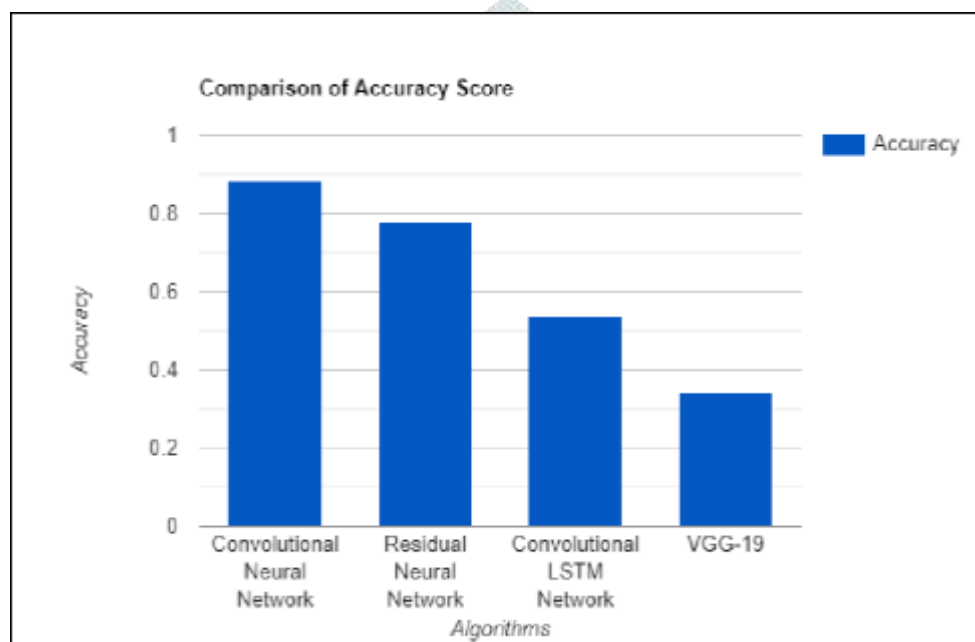


Fig. 3: Comparison of Accuracy Score

Figure 3 shows the plotting of the models we considered V/S the accuracy achieved for each of them as stated in Table 5.1.

After creating the website and setting up the machine learning model using Convolutional Neural Network, we have the following structure of the Emotion Mapping based Music Recommendation System using Machine Learning:



Fig. 4: Playlist for "Happy" emotion



Fig. 5: Playlist for "Fearful" emotion



Fig. 6: Playlist for “Sad” emotion



Fig. 7: Playlist for “Neutral” emotion



Fig. 8: Playlist for “Surprised” emotion

Figure 4, Fig. 5, Fig. 6, Fig. 7 and Fig. 8 show the interface for playlists appropriate for the emotions.

VI CONCLUSION

In this project, we have presented a machine learning model to recommend music based on emotion captured through facial expressions detected by the system. We have given a personal touch to the user by including an emoji face which also signifies the emotion of the user. This product is an Open Source Web Application which can be used by different types of people. In these difficult times when everyone is stressed and worried about what the future beholds, music has the power to heal any stress or any kind of upsets.

REFERENCES

- [1] Florence, S. Metilda, and M. Uma. "Emotional Detection and Music Recommendation System based on User Facial Expression." IOP Conference Series: Materials Science and Engineering. Vol. 912. No. 6. IOP Publishing, 2020.
- [2] James, H. Immanuel, et al. "EMOTION BASED MUSIC RECOMMENDATION SYSTEM." EMOTION 6.03(2019).
- [3] Facial Emotion Based Music Recommendation System using computer vision and machine learning techniques ShanthaShalini. Ka , Jaichandran. Ra , Leelavathy. S A, Raviraghul. Ra , Ranjitha. J a and Saravanakumar.
- [4] Pitaloka, Diah Anggraeni, et al. "Enhancing CNN with preprocessing stage in automatic emotion recognition." Procedia computer science 116 (2017): 523-529.
- [5] Kuo, Fang-Fei, et al. "Emotion-based music recommendation by association discovery from film music." Proceedings of the 13th annual ACM international conference on Multimedia. 2005.
- [6] Dureha, Anukriti. "An accurate algorithm for generating a music playlist based on facial expressions." International Journal of Computer Applications 100.9 (2014): 33-39.
- [7] Badve, Ameya, et al. "MUSIC RECOMMENDATION USING FACIAL EMOTION DETECTION AND CLASSIFICATION." Journal of Critical Reviews 7.19 (2020): 1082-1089.
- [8] Alshaabi, Thayer, et al. "How the world's collective attention is being paid to a pandemic: COVID-19 related n-gram time series for 24 languages on Twitter." Plos one 16.1 (2021): e0244476.
- [9] Samuvel, Deny John, B. Perumal, and Muthukumaran Elangovan. "Music recommendation system based on facial emotion recognition." (2020).
- [10] Chin, Yu Hao, et al. "Emotion profile-based music recommendation." 2014 7th International Conference on Ubi-Media Computing and Workshops. IEEE, 2014.
- [11] Kumar, Manoj, et al. "Expression X: Emotion Based Music Recommendation System." Journal of Computational and Theoretical Nanoscience 17.9-10 (2020): 3958-3963.
- [12] Shiha, Mohammed, and Serkan Ayyaz. "The effects of emoji in sentiment analysis." Int. J. Comput. Electr. Eng.(IJCEE.) 9.1 (2017): 360-369.
- [13] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. "Emotional state classification from EEG data using machine learning approach." Neurocomputing, Volume 129. ISSN 0925-2312. Elsevier, 2014.

- [14]Hasib, Hasib. "Natural Substitution of Emoji on Our Daily Life: Emoji makes us more specific to express our motion." Available at SSRN 3664474 (2020).
- [15]Hao, Zhu. "The development of emoji in the intelligent era." 2020 International Conference on Intelligent Design (ICID). IEEE, 2020.

