***Song Recommendation System Using Facial Recognition***

Akshat Pareta Ribhav Sharma

Chandigarh University Chandigarh University

[20BCS6567@cuchd.in](mailto:20BCS6567@cuchd.in)  [20BCS6558@cuchd.in](mailto:20BCS6558@cuchd.in)

Rohan Mishra Simranjeet Singh

Chandigarh University Chandigarh University

[20BCS4456@cuchd.in](mailto:20BCS6558@cuchd.in)  [20BCS6575@cuchd.in](mailto:20BCS6558@cuchd.in)

# Abstract:

*Nowadays, music services enable instant access to vast amounts of music. They are trying consistently to boost music arrangement and search management consequently resolving the subject of choice and simplifying exploring fresh pieces. Systems that offer recommendations have gained tremendous popularity and assist users to discover acceptable music for all times. Yet, there's still a specific sector in personalisation and emotions based suggestions. Music includes a big influence on people and is commonly employed for relaxing, mood control, elimination from stress and illnesses, to take charge of both physical and mental duties. There is an outstanding range of therapeutic venues and techniques in music therapy providing well-being support. The article will demonstrate the development of the customised sound recommending system, driven by audience emotions, feelings, and activity settings. With a combination of AI technology and comprehensive musical therapy approaches, the system for recommendations is designed to aid people with song selection for diverse situations in life and maintain both their physical and mental states.*

# Introduction:

Many of the study in recent years indicate that individuals remark and react to music and therefore

music has a substantial impact on the activity of the cerebral cortex of a person. In one analysis of the factors why humans interpret tunes, researchers determined that music played a crucial purpose in connecting arousal and mood. Two of the most crucial functions of song are it is ability is consumers estimated to assist them acquire a pleasurable mood and become more self-aware. Musical preferences have been discovered to be significantly related to personality features and emotions [1].

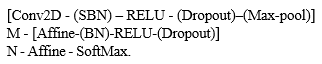
The metres, timbre, cadence, and tone of music are governed in portions of the brain that influences mood and emotions [2]. Interaction between persons may be a vital component of lifestyle. It exhibits faultless details and plenty of data between human beings, whether it comes in the form of gestures, voice, expressions on the face, or emotions [3].Today, emotion detection is regarded as the most important technology used in a variety of applications, including smart card applications, surveillance, picture databases inquiry, illegal video the indexing, and citizen applications, security, and adaptive interactions between humans and computers with multimedia environments.Automated emotion recognition in multimedia attributes like movies or music is developing quickly thanks to an increase in technology for processing digital signals and other operational methods for feature extraction, and this structure can play a significant role in many potential uses like interaction between humans and computers systems and songs entertainment. We utilise facial expressions to outline a proposed technique for emotion detection that may identify user moods and offer a range of appropriate tunes [13–24]. The proposed system recognises an individual's feelings; if the individual has an adverse feeling, a specific playlist will be produced that includes the most appropriate genres of music that would uplift his spirits. And if these feelings are positive, a personalised playlist that includes a variety of musical genres that might amplify positive feelings will be provided [4]. We used datasets from the Kaggle dataset of facial expression Identification [5] for our emotion recognition work. Indian Hindi music has been used to produce data sets to an audio player. Convolutional models of neural networks are used to implement facial emotion detection.which yields roughly 95.13% of accuracy [2].

1. **Related Work:**

There have been significant advances in the development of automatic expression classifiers. in recent years [7, 8, 9]. Some systems for recognizing facial expressions classify the expression into a range of fundamental emotionss., including joy, sad and rage.. In an effort to provide an unbiased description of the face, others have made an effort to identify the specific muscle movements that the face is capable of doing[11]. The most widely used psychological framework is the Face Action Coding System (FACS)[12]. for summarizing practically all facial movements. Using Action Units, the FACS system classifies human facial movements according to how they appear on the face (AU). A facial expression usually originates from one of the 46 atomic units (AUs) of face movement or related deformation that can be detected. Several AUs are usually added together to form an expression [7, 8]. Additionally, Multilevel Hidden Markov Model, Neural Networks, and Bayesian Networks (HMM) have all seen development in the approaches employed for face emotion identification [13],[14]. Several of them have problems with timing or detection rates. combining two or more techniques to accomplish precise recognition allows for the extraction of features as necessary. Because of illumination and feature extraction, Each technique's effectiveness is reliant on image pre-processing.

# Methods:

To assess how well these models performed at identifying facial expressions, we created CNNs with varying depths. For our analysis, we took into account the following network design.



All of these layers have the convolution layer and ReLU non - linear., are referred to as the first component of the network. These layers also include dropout, max-pooling, and spatial batch normalization (SBN), which can be present. After M convolutional layers, which are always Affine-operating & ReLU non-linear and may additionally the network is when batches normalizing (BN) as well as dropouts led to N fully linked layers. In the show's processing order, the affine layer follows the network and computes the score and softmax loss function.. The user has control over the created model's convolution and fully - connected layers layer counts along with the presence of batch normalisation, dropout, and max-pooling layers. We used L2 regularisation in addition to dropouts and batch normalisation methods. Additionally, the user can specify the quantity of filters, strides, and zero-padding; otherwise, basic settings are used.

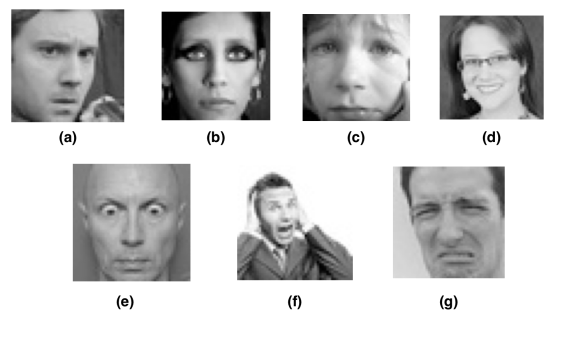
We introduced the idea of merging HOG characteristics with those recovered by convolutional layers using raw pixel data, as we shall explain in the following section. To do this, we used this framework as before, but we now added the HOG characteristics to the features exiting  final convolutional layer. The hybrid feature set is then passed to the fully connected layers, who use it to compute scoring and losses. 

Figure 1: Examples of the 7 face expressions we are taking into account for this classification issue. Aside from being furious[a], you can also feel neutral[b], sadness [c], delighted[d], surprised [e], fearful[f], or disgusted[g].

Thus order to speed up the network train program, we developed that above models in Torch and used GPU acceleration deep learning characteristics.

1. **Literature-Review:**

The analysis is carried out to learn about the processes' shortcomings so that we may address them. A literature review, also known as a review of the research, is a document included in an academic publication that includes the current understanding, notable findings, methodological advancements, and theoretical developments related to a certain topic. Numerous students, engineers, scientists, and other professionals from all over the world have been interested in the hidden human characteristics that can provide inputs to a system in a variety of ways. The person's current mental state can be inferred from their facial expressions. In interpersonal communication, we frequently use nonverbal cues like gestures with our hands, facial expressions, and voice tonality to convey our feelings. Preema and others [6] remarked that creating and maintaining a large playlist takes a lot of time and effort. According to the claim, the "song player" allegedly chooses a tune depending on the user's current mood. To produce playlists depending on mood, the application examines and categorises recordings of sound using audio characteristics. The Viola-Jonas approach, which is used to extract facial expressions, is employed by the software. In order to categorise traits into the five main universal emotions of anger, joy, amazement, melancholy, and disgust, the Support Vector Machine, also known as or SVM, was used.t.

# Proposed-System:

The recommended technique enables us to give communication between the individual utilising it and the music player. The purpose of the technique is to acquire photographs of a person's face adequately utilising the camera. Captured photographs are sent into a CNN, which forecasts the atmosphere. Then emotion collected from the recorded image is applied to create a playlist of songs. The fundamental purpose of our recommended system is to supply a playlist of songs dynamically to vary the the consumer's emotions, which might be pleased, regretful, natural, or startled. The proposed approach recognises the feelings, if the issue contains a bad sensation , then a selected playlist will have to be supplied that combines the main appropriate types of tracks that can enhance the mood of the individual positively. Songs suggestion based on face expression detection has four sections.

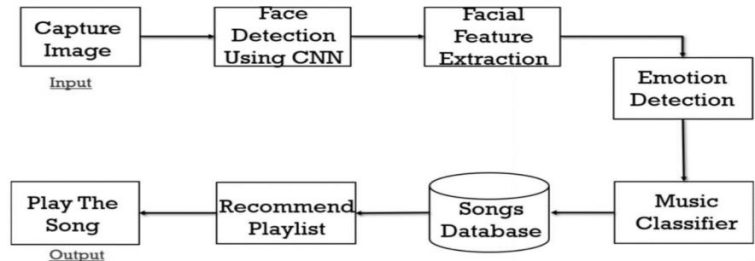
• Real-Time Scan: This component, the objective is to capture’s the facial features of the person using it accurately

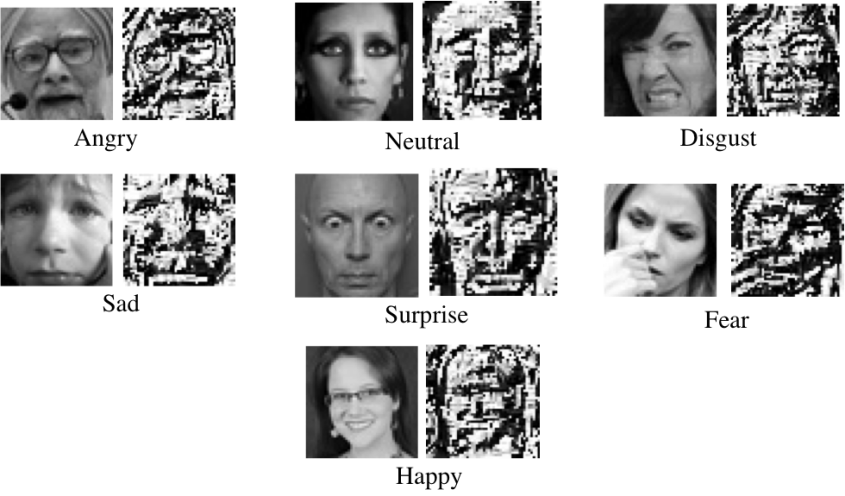
• Facial Identification: In this it will accept an individual's visage as input. The Convolutional neural network, or CNN, is programmed to analyse the characteristics of the user picture.

• Sentiment Analysis: This phase extraction of the characteristics of the person’s picture is done to identify the mood and based on the user's emotions, the system will create captions.

• Music Recommendations: Song is suggested by the recommendation module to the individual utilising it by matching their feelings to the emotive type of the song

.





## Methodology

**6.1 Database Description**

We constructed Convolutional Neural Networks (CNN) models using the Kaggle data set. The dataset is FER2013 that is segregated into two divisions train and test dataset. The initial data consists of 24176 while the test data set has 6043 photos. A total of 48x48 pixel grayscale photographs of faces in the data compilation. Every picture in FER-2013 is labelled as any of the five emotions: joyful, sorrowful, enraged, surprise, and neutral.

The human features automatically register so that they're approximately centered in each picture and take up roughly the same amount of area. The photographs in FER-2013 comprise both composed & unposed portraits, and they're in grayscale & 48x48 pixels.



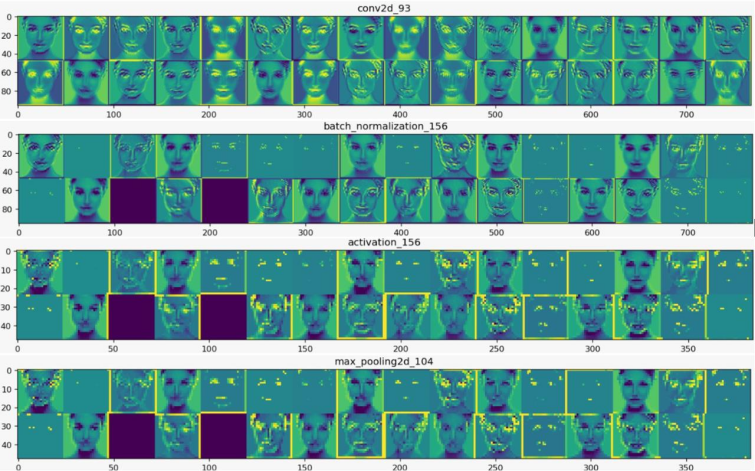
**6.2 Emotion Identification module**

**6.2.1 Facial detection**

The detection of features is one of the applications which is regarded under artificial intelligence technology. This is a method by which algorithms are devised and taught to precisely identify faces or things in object identification or associated system in photographs. This detection might be real-time using the video frame or photos. Face detection employs such classification algorithms, which are techniques that determine what is either a human face (1) or not a face (0) in a picture. Classifiers are taught to recognise features utilising numbers of photographs to obtain higher accuracy. OpenCV employs two forms of classification algorithms, LBP (Local Binary Pattern) & Haar Cascades. A Haar classification is used for face identification where a classifier is trained using specified variable facial information which enables it to recognise diverse faces appropriately. The fundamental purpose of facial recognition is to locate the human visage within the picture by eliminating exterior audio and other elements. It's a machine learning-based strategy where a cascade function is taught with a collection of input files. It supports the Haar Wavelet approach to examine pixel within the picture into square by function . This employs neural network algorithms to promote an elevated level of efficacy from what's termed "training data".

**6.2.2 Feature Identification**

While executing the extraction of features, we employ the network that has been trained which is a sequential models as an unrestricted feature extractor. Letting the input image to advance on it ahead, ceasing at the preset layer, and employing the outcomes of the specified layer as our features. Starting with each layer of convolutional neural networks to retrieve high-level data from the recorded image, thus requires just a few filters. As we create successively more complex levels, we raise the amount of filtering to double or treble the width of the filter itself of the layer that came before it. Filters of the higher levels gather more features but are computationally fairly expensive.

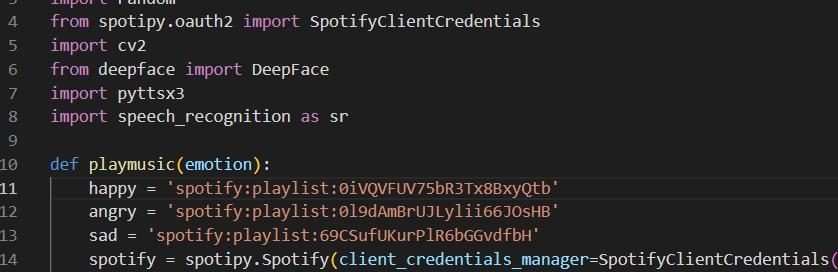


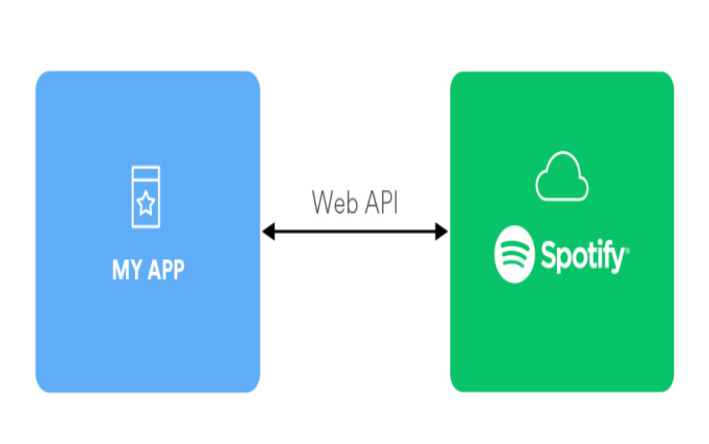
Doing this we leveraged the strong, discriminative qualities acquired through the process of convolution neural network. The results from the simulation will continue to be converted into maps of features, which serve as an intermediate format for all levels after the initial layer. Load the input picture for which we want to evaluate the Feature map to learn which features were prominent to classify the image. Feature maps are formed through the use of Filters and detectors of features on an input image or the characteristic map outputs of the previous layers. Feature map visualizations will give insight into the inner representations of specific inputs for each one of the layers of convolution within the model..

**6.3 Music Recommendation Module**

For Recommending the songs instead of creating a separate database for songs we have used spotify API for this. The advantage of using spotify API is that we don’t need to add songs seprately we just have to connect our face detection data set with the API and it is going to select songs on its own based on the emotion detected.

This will result in the proposed playlist for the individual in the graphical user interface of the music device by displaying captions according to assessed emotions. We employed a library called Pygame for streaming music since this package supports playing numerous multimedia formats like video, audio, etc.





## Future Work

This system, while wholly operational, does offer space for development in the future. There are several components of the programme that may be altered to generate enhanced outcomes and a better complete experience to the user. Some of them provide use different fashion, based on additional emotions which are eliminated in our system like revulsion and dread. This emotion includes supporting the rendition of music reflexively. The future scope inside the platform would design a mechanism that would be advantageous in music therapy treatments and enable the musical therapists to cure patients contending from mental tension, anxiety, acute depressive disorders, and trauma. The extant system does not operate well in really dreadful light situations and poor image resolution hence providing a chance to add more features for an alternative in the future.

1. **Result and Output:**

Mood being detected and music being played

Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generated

1. **Conclusion**

A thorough review of the literature tells that there are many approaches to implement Music Recommender System. A study of methods proposed by previous scientists and developers was done. Based on the findings, the objectives of our system were fixed. As the power and advantages of AI-powered applications are trending, our project will be a state-of-the-art

trending technology utilization. In this system, we provide an overview of how music can affect the user's mood and how to choose the ideal music recordings to enhance the user's moods. The constructed system can discern the user's emotions. The emotions that the computer can distinguish were joyous, sorrowful, enraged, neutral, or startled. After measuring the user’s emotion, the suggested method supplied the user with a compilation that encompasses music matching that acknowledged the mood. Processing an enormous data collection is storage as well as CPU costly. This will make development more challenging and appealing. The purpose is to design this kind of application in the lowest practicable technique and also to produce it under a standardized device. Our music recommendation system based on facial expression identification will simplify the efforts of users in creating and administering playlists.

# References

1. [https://www.kaggle.com/c/challenges-in-](http://www.kaggle.com/c/challenges-in-) representation-learning-facial-expression- recognition-challenge/data
2. <https://github.com/torch>
3. Dalal, Navneet, and Bill Triggs (2005). Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference on. Vol. 1
4. Bettadapura, Vinay (2012). Face expression recognition and analysis: the state of the art. arXiv preprint arXiv:1203.6722
5. Lonare, Ashish, and Shweta V. Jain (2013). A Survey on Facial Expression Analysis for Emotion Recognition. International Journal of Advanced Research in Computer and Communication Engineering 2.12
6. Nicu Sebe, Michael S. Lew, Ira Cohen, Yafei Sun, Theo Gevers, Thomas S. Huang (2007) Authentic Facial Expression Analysis. Image and Vision Computing 25.12: 1856-1863
7. 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.
8. M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. In Proceedings of the IEEE ConFEDence on Automatic Facial and Gesture Recognition, 2006.
9. 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
10. G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. Image and Vision Computing, 24(6), 2006.
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. P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, 1978
13. Cohen, Ira, et al. ”Evaluation of expression recognition techniques.” Image and Video Retrieval. Springer Berlin Heidelberg, 2003. 184- 195
14. Padgett, C., Cottrell, G.: Representing face images for emotion classification. In: Conf. Advances in Neural Information Processing Systems. (1996) 894900.
15. <http://scikit-learn.org/stable/>
16. <http://deepdreamgenerator.com/>
17. <https://github.com/google/deepdream>
18. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. 2012.
19. Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).
20. P. S Aleksic and A. K. Katsaggelos, “Automatic facial expression recognition using facial animation parameters and multistream hmms,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 3–11, 2006. View at: [Publisher Site](https://doi.org/10.1109/tifs.2005.863510" \t "_blank) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Automatic%20facial%20expression%20recognition%20using%20facial%20animation%20parameters%20and%20multistream%20hmms&author=P.%20S%20Aleksic%20and%20A.%20K.%20Katsaggelos&publication_year=2006" \t "_blank)
21. H. Chouhayebi, J. Riffi, M. A. Mahraz, Y. Ali, and T. Hamid, “Facial expression recognition using machine learning,” in *Proceedings of the 2021 Fifth International ConFEDence On Intelligent Computing in Data Sciences (ICDS)*, pp. 1–6, IEEE, Seoul, Korea, November 2021. View at: [Publisher Site](https://doi.org/10.1109/icds53782.2021.9626709" \t "_blank) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Facial%20expression%20recognition%20using%20machine%20learning&author=H.%20Chouhayebi,%20J.%20Riffi,%20M.%20A.%20Mahraz,%20Y.%20Ali,%20and%20T.%20Hamid" \t "_blank)