# Music Recommendation Using CNN

FAIZ MOHAMMAD DANGAR[1], MOHAMMED BASIM K[2], ADARSH D[3], MANJULA R CHOUGALA[4]

[1, 2, 3] *UG Student, Department of ISE, R R Institue of Technology, Bengaluru, Karnataka*

[4] *Guide*

***Abstract-*** ***A user's choice of music is influenced by their mood as well as their past musical tastes and musical substance. In this study, we introduce a powerful music recommendation system that makes music suggestions depending on the user's current mood. It primarily focuses on the Convolutional Neural Network (CNN) model, which distinguishes seven different facial expressions in humans using Mobile Net architecture. Three modules make up our system: the Recommendation Module, the Music Classification Module, and the Emotion Module. The Emotion Module uses CNN to determine the user's current mood after receiving a picture of their face as input. The Music Classification Module uses auditory attributes to classify music and achieves a phenomenal result of 98% songs into 4 categories of mood.***

***Indexed Terms-*** ***Convolutional Neural Network; Emotion Module; Music Classification Module; the Recommendation Module***

## I. INTRODUCTION

Since it is a popular kind of entertainment for music lovers and listeners and occasionally even imparts a therapeutic approach, music plays a significant part in improving a person's quality of life. Various music players have been developed with features like fast-forward, reverse, variable playback speed, local playback, streaming playback with the multicast streams, and including volume modulation, genre classification, etc. in the modern world due to the rapid advancements in multimedia and technology. One of these techniques to identify human emotion states (such as anger, fear, neutrality, happiness, disgust, sadness, and surprise) is deep learning-based facial expression identification. This technique seeks to accurately identify the emotional state by automatically detecting facial expressions. This method sends facial photos that have been labelled from a facial expression collection CNN uses these images to teach itself. The proposed CNN model determines which facial expression is used after that.

Artificial intelligence is a vast, important, and pressing field that has recently drawn many researchers and programmes. This particular domain has quickly seized control of the entire planet. Chat bots, virtual assistants like Siri, and other technology-based systems are all incorporated into daily life. Face recognition technology is one of artificial intelligence's most notable advances. The collection of Google Photos of a specific individual serves as the most basic illustration of its application.

Numerous existing technologies are capable of identifying facial expressions of emotion. On the other hand, some programmes offer music recommendations. assembling a system that will make recommendations

The overall notion presented in the paper is to play music by identifying the user's mood from their facial expressions. In the near future, emotion recognition will have more applications in industries like robotics for effective emotive analysis without the need for a second human.

## II. RELATED WORK

Instead than focusing on who is more emotional, try to derive feelings from people's statements. As a result, studies on how people express themselves with more or lesser intensity of feeling are lacking. It is crucial to collect this information in logs because people use social networks at specific times of the day, and understanding their routines and behaviours may be valuable for a variety of applications. It is feasible to create a recommendation for a specific time period by implementing logs. For instance, a person typically

reads financial news on Monday and Tuesday from 9 to 10 a.m., and they typically read entertainment news on Saturday. This paper's major objective is to show that a lexicon-based sentiment intensity A low complexity solution with benefits for consumer electronics devices can enhance the effectiveness of a music recommendation system by employing metrics linked with a correction factor. The adjustment factor makes use of a person's easily extrapolated traits from social networks. Therefore, the new metric, Esm, provides a sentiment value that is more precise. To enhance the user's quality of experience, the proposed recommendation system took ergonomic usability considerations into account. The remainder of this essay is organised as follows in light of this setting. The relevant work is presented in Section II. The proposed sentiment intensity model, which was utilised to calculate the eSM, is presented in Section III. The architecture of the suggested recommendation system based on sentiments is presented in Section IV. The experimental findings are presented in Section V and include the proposed system of recommendations coupled with an assessment of eSM effectiveness

## III. PROPOSED ALGORITHM

In the past, there was no software system to detect face extraction and recommendation, but Deep Convolution Neural Network (DCNN) was used to extract the recognition of facial emotions using the Tensor Flow Machine Learning library. It was a lengthy procedure. There was no assurance that the detection result would always be accurate.

Unknowingly, humans often reflect their emotions on their faces, which is how this happens. We can give a user-music system interaction thanks to the proposed system. The primary focus of this project is on the user's favourite music, which is suggested based on emotional awareness. We provided three options at the beginning of the proposed system, and each one has its own functionality. To this, we've provide Based on spatial perception, this list of music and feelings. Once the application is running, it uses the webcam or any other physiological device to take pictures. Our primary goal in developing this system is to create a smart music player that may improve user mood, and music is one of the best tools for doing so. Since human emotions vary from person to person and are difficult to predict, only four frequent and easily recognisable moods of the individual are taken into consideration when comparing the photos collected by the system with the data sets. Additionally, this alternative approach can be employed with the central idea. i.e, random picking of songs that could help us to brighten our moods and the other mode is queue mode with this we can make a playlist on our own and in all the modes we are not using the previous user data but we using individual user data.

### A. FACE DETECTION:

By eliminating extraneous noises and other elements, the major goal of the face identification technique is to identify the face in the frame. The FACE DETECTION PROCESS has the following steps:

1. Image pyramid
2. Oriented gradients histogram
3. Linear Classifier

Using an image pyramid with several scales, the resulting data are divided into the sample image. Simply extracting features while lowering noise and other parameters is how this technique is used.

In the realm of image processing, a feature descriptor called HOG is a method that counts the occurrences of gradient orientation in a certain area of an image and is frequently used to identify objects in images. Using this method, the primary goal is to characterise the face inside the image with a variety of variations in intensity.

### B. EMOTION CLASSIFICATION:

A bounding box will be added as an overlay to the image once the face has been properly detected in order to extract the ROI (face) for additional analysis. The 68 facial landmark points will then be recovered from the extracted ROI using the "Predictor" function, commonly known as a script, and saved in an array. The data from the features array will then be fed into a PCA reduction procedure, which will shrink the data and remove any associated coordinates, leaving only the necessary points as principle components. Each of the 68 points in the data's 68x2 array has coordinates on the x- and y-axes. The array will be transformed into a vector with 136 rows and 1 column column. A set of photos and landmark maps for each image are

used to train the facial landmark extraction code "Predictor."

Using regression trees trained with the gradient boosting approach, the code learns how to extract the facial landmark map from a given face image based on the pixel's intensity values indexed of each point. The data will be used for categorization after the CNN reduction process. In order to determine which class (feeling) the entered data belongs to, a multiclass CNN with a linear kernel is used to compare it with the stored data. If one of the three emotions—anger, fear, or surprise—is picked up, a speed-reducing command will be carried out to slow down the wheelchair in order to protect the rider.

C. MUSIC RECOMMENDATION:

Since the input is obtained in real-time, the camera is used to record the video before framing is completed. The framed photos are processed using a hidden markov model classification. For the goal of classifying emotions, the collected frames are taken into account in all frame and pixel formats. Every facial landmark's value is computed and saved for further use. The classifier's efficiency is between 90 and 95 percent, so even if the face changes as a result of environmental factors, the system can still identify the emotion being conveyed on the face. The values received from being set and from the value of the pixel are then used to identify the emotions. that is received and the values present as threshold in the code are contrasted. Values are sent from the client to the web service. The song is played in response to the sensed emotion. Each song has a set of designated emotions. The appropriate music will play when the desired feeling is conveyed. Happy, angry, sad, and surprised are the four emotions that can be employed. The songs designated for that particular emotion are played when the happy emotion is identified, and the same is true for the other emotions; in other words, the songs are played in accordance with the emotions that are detected.

## IV. SIMULATION RESULTS

This project involves an emotion-based music recommendation system. For recognising facial emotions, the system uses a two-layer convolution network model. Seven distinct facial emotions are classified by the model using the image dataset. The model's equivalent training and validation accuracy indicates that it is the most accurate and generalizable to the data. Analyzing how the system functions when other emotions are taken into account might be intriguing. Collaborative filtering can be used to gather user preferences and enhance the system as a whole. Future work will address these problems, as planned. In today's society, a music player with a facial recognition technology is absolutely necessary for everyone. This system has been improved and is advantageous features that will allow for future upgrades. The mechanism for improving music playback that occurs automatically uses facial expression recognition. The RPI camera's programming interface allows for the detection of facial expression. An alternate approach built on feelings other than revulsion and terror that are not recognised by our system. To assist the automatic playing of music, this feeling was introduced.

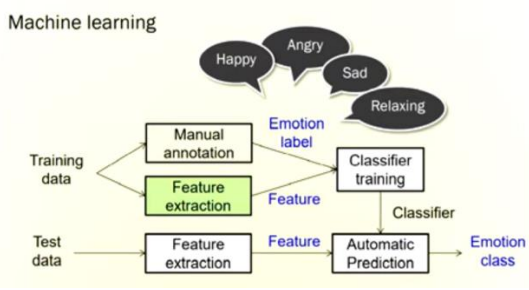


Fig.1. Diagramatic representation of the project

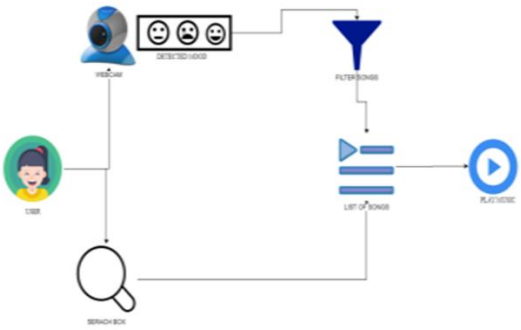

Fig. 2. Data Flow diagram of the project

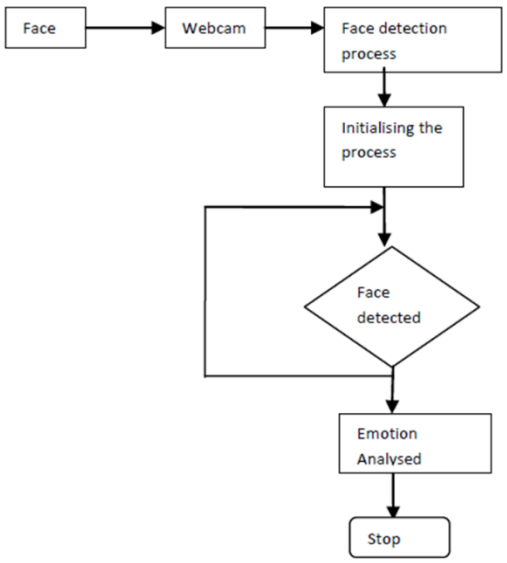


Fig. 3. Block diagram of Face detection

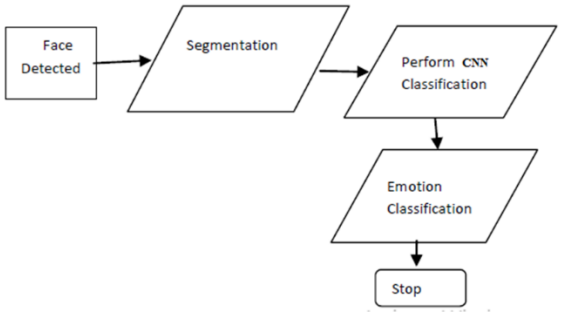


Fig 4. Block diagram of Emotion detection

## CONCLUSION AND FUTURE WORK

The simulation results demonstrated that the suggested approach outperforms the maximum number of hops metric when using the total transmission energy metre. The suggested method maximises the lifetime of the entire network and offers an energy-efficient way for data transfer. The performance of the suggested algorithm can be compared with other energy-efficient algorithms as the performance of the method is evaluated between two metrics in the future with minor alterations in design considerations. We've only employed a relatively modest network of five nodes; as the number of nodes rises, so will its complexity. We can add more nodes and assess how well they work.

## REFERENCES

[1] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632. https://doi.org/10.31979/etd.w5fs-s8wd

[2] Hemanth P,Adarsh ,Aswani C.B, Ajith P, Veena A Kumar, "EMO PLAYER: Emotion Based Music Player", International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 4, April 2018, pp. 4822-87.

[3] Music Recommendation System: *"Sound Tree"*, Dcengo Unchained: Sıla KAYA, BSc.; Duygu KABAKCI, BSc.; Işınsu KATIRCIOĞLU, BSc. and Koray KOCAKAYA BSc. Assistant : Dilek Önal Supervisors: Prof. Dr. İsmail Hakkı Toroslu, Prof. Dr. Veysi İşler Sponsor Company: ARGEDOR

[4] Tim Spittle, lucyd, GitHub, , April 16, 2020. Accessed on: [Online], Available at: https://github.com/timspit/lucyd

[5] A. Abdul, J. Chen, H.-Y. Liao, and S.-H. Chang, "An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach," *Applied Sciences*, vol. 8, no. 7, p. 1103, Jul. 2018.

[6] Manas Sambare, FER2013 Dataset, Kaggle, July 19, 2020. Accessed on: September 9, 2020. [Online], Avalilable at: https://www.kaggle.com/msambare/fer2013

[7] MahmoudiMA, MMA Facial Expression Dataset, Kaggle, June 6, 2020. Accessed on: September 15, 2020. [Online], Available at: https://www.kaggle.com/mahmoudima/mma-facial-expression

[8] Dr. Shaik Asif Hussain and Ahlam Salim Abdallah Al Balushi, "A real time face emotion classification and recognition using deep learning model", 2020 Journal. of Phys.: Conf. Ser. 1432 012087

[9] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition -

Workshops, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

[10] Puri, Raghav & Gupta, Archit & Sikri, Manas & Tiwari, Mohit & Pathak, Nitish & Goel, Shivendra. (2020). Emotion Detection using Image Processing in Python.

[11] Patra, Braja & Das, Dipankar & Bandyopadhyay, Sivaji. (2013). Automatic Music Mood Classification of Hindi Songs.

[12] Lee, J., Yoon, K., Jang, D., Jang, S., Shin, S., & Kim, J. (2018). MUSIC RECOMMENDATION SYSTEM BASED ON GENRE DISTANCE AND USER PREFERENCE CLASSIFICATION.

[13] Kaufman Jaime C., University of North Florida, "A Hybrid Approach to Music Recommendation: Exploiting Collaborative Music Tags and Acoustic Features", UNF Digital Commons, 2014.

[14] D Priya, *Face Detection, Recognition and Emotion Detection in 8 lines of code!*, towards data science, April 3, 2019. Accessed on: July 12, 2020 [Online], Available at: https://towardsdatascience.com/face-detection-recognition-and-emotion-detection-in-8-lines-of-code-b2ce32d4d5de

[15] bluepi, *"\Classifying Different Types of Recommender Systems*, November 14, 2015. Accessed on: July 7, 2020. [Online], Available on: https://www.bluepiit.com/blog/classifying-recommender-systems/#:~:text=There%20are%20majorly%20six%20types,system %20and%20Hybrid%20recommender%20system