JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MUSIC RECOMMENDATION SYSTEM BASED ON EMOTION

[1]SINGAMSETTY BHARGAVI, [2]AKULA YAMUNA BHARGAVI,[3]DOKI LIKITHA, [4]MOGALI MADHU BABU, [5]SEEPANA RATNA KUMARI

[1]B.TECH(STUDENT),[2]B.TECH(STUDENT),[3]B.TECH(STUDENT),[4]ASSOCIATE PROFESSOR,[5]ASSISTANT PROFESSOR

[1]Department of Computer Science and Engineering,

[1]Satya Institute of Technology and Management, Vizianagaram, India

Abstract: A user's emotion can be detected by his/her facial expressions. These expressions can be derived from the live feed via the system's camera. A lot of research is being conducted in the field of Computer Vision and Deep Learning (DL), where machines are trained to identify various human emotions. Deep Learning provides various techniques through which human emotions can be detected. One such technique is to use CNN model with Keras, which generates a small size trained model and makes Android-ML integration easier. Music is a great connector. It unites us across markets, ages, backgrounds, languages, preferences, political leanings and income levels. Music players and other streaming apps have a high demand as these apps can be used anytime, anywhere and can be combined with daily activities, travelling, sports, etc. With the rapid development of mobile networks and digital multimedia technologies, digital music has become the mainstream consumer content sought by many young people.

***Index Terms* -: Music recommendation, emotions based, facial expression recognition, Mood.**

1.INTRODUCTION

Music prompts a reasonable passionate reaction in its audience. Melodic inclinations have been exhibited to be exceptionally associated with character qualities and mind-sets. Facial emotions are the most common and natural methods of passing on feelings, temperaments and sentiments. Convolutional Neural network, as a Deep Learning Neural Network, assumes a critical part in face image recognition. Cognition technology of CNN and Music Recommendation System based on Facial Emotion Gestures is created to distinguish a model that perceives facial articulations and prescribes music as indicated by comparing mind-set of the user or client. Human beings have the innate capacity to see somebody's face and conjecture their mind-set. This capacity if learnt by an electronic gadget - computer, humanoid robot or a mobile gadget – can have important applications. Music, an instrument for stirring emotions and feelings are undeniably more remarkable than language. Music is something which takes advantage of our emotional center as human beings. Accordingly, paying attention to good music can assist us with lifting our mind-set from a negative sense to a positive sense. For example, focusing on lively tunes when the individual is feeling grim can assist him with arising his difficulty and start feeling better. This framework proposes one such application, emotion- based music recommendation. Emotion of the client can be effortlessly speculated by taking a gander at his/her face. For this reason, face detection and emotion recognition, examining the fiducial highlights from his/her face is essential. These emotions are very subtle. Facial muscle contortions are very minimal and detecting these differences can be very challenging as even a small difference results in different expressions. Also, expressions of different or even the same people might vary for the same emotion, as emotions are hugely context dependent. Capturing and recognizing the emotion being voiced by a person and displaying appropriate songs matching the one's mood and can increasingly calm the mind of a user and overall end up giving a pleasing effect. The project aims to capture the emotion expressed by a person through facial expressions. A music player is designed to capture human emotion through the web camera interface available on computing systems. Since ancient times the best form of expression analysis known to humankind is facial expression recognition. The best possible way in which people tend to analyse or conclude the emotion or the feeling or the thoughts that another person is trying to express is by facial expression. In some cases, mood alteration may also help in overcoming situations like depression and sadness.

1.1PURPOSE

The purpose of a music video recommendation system based on emotions is to suggest videos that align with the user's current emotional state, enhancing their viewing experience and potentially uplifting their mood. By analyzing factors like tempo, lyrics, and visual content, the system aims to provide tailored recommendations that resonate with the user's feelings at any given moment.

1.2SCOPE OF THE STUDY

- This system, although completely functioning, does have scope for improvement in the future. There are various aspects of the application that can be modified to produce better results and a smoother overall experience for the user.
- Some of these that an alternative method, based on additional emotions which are excluded in our system as disgust and fear. This emotion included supporting the playing of music automatically.
- The future scope within the system would style a mechanism that might be helpful in music therapy treatment and help the music therapist to treat the patients suffering from mental stress, anxiety, acute depression, and trauma.
- The current system does not perform well in extremely bad light conditions and poor camera resolution thereby provides an opportunity to add some functionality as a solution in the future.

2.RESEARCH METHODOLOGY

Several steps would be included in the process for an emotion-based music recommendation:

- **Data collection:** Collect a dataset of music tracks, along with their associated emotion labels. This can be done using crowdsourcing or by using existing emotion-labeled datasets.
- **Feature extraction:** Extract audio features from each track in the dataset. Common audio features used for music recommendation include MFCCs (Mel-Frequency Cepstral Coefficients), spectral features, and rhythmic features.
- **Emotion detection**: Train a Haar Cascade algorithm on the emotion labels to detect emotions in audio features. The Haar Cascade algorithm is a popular method for object detection in computer vision and can be adapted for audio processing.
- **Music recommendation:** Use the emotion recognition algorithm to propose music songs depending on the user's current emotional state. This may be accomplished by matching the user's feelings to the emotions of the music recordings in the dataset.
- **Evaluation:** Evaluate the performance of the emotion-based music recommendation system using common evaluation criteria such as accuracy, recall, and F1-score.
- **Analysis**: Analyse the assessment data and generate judgements regarding the efficacy of the emotion-based music recommendation system. Comparing the system against existing music recommendation systems and suggesting areas for improvement might be part of this process.
- **Conclusion:** Summarise the findings of the research study and make recommendations for future work on emotion-based music recommendation systems utilising the Convolutional neural network.

3.Implementation:

3.1Database Description:

We built the Convolutional Neural Network model using the Kaggle dataset. The database is FER2013 which is split into two parts training and testing dataset. The training dataset consists of 24176 and the testing dataset contains 6043 images. There are 48x48 pixel grayscale images of faces in the dataset. Each image in FER-2013 is labelled as one of five emotions: happy, sad, angry, surprise, and neutral. The faces are automatically registered so that they are more or less centered in each image and take up about the same amount of space. The images in FER-2013 contain both posed and unposed headshots, which are in grayscale and 48x48 pixels. The FER-2013 dataset was created by gathering the results of a Google image search of every emotion and synonyms of the emotions. FER systems being trained on an imbalanced dataset may perform well on dominant emotions such as happy, sad, angry, neutral, and surprised but they perform poorly on the under-represented ones like disgust and fear. Usually, the weighted-SoftMax loss approach is used to handle this problem by weighting the loss term for each emotion class supported by its relative proportion within the training set. However, this weighted-loss approach is predicated on the SoftMax loss function, which is reported to easily force features of various classes to stay apart without listening to intra-class compactness. One effective strategy to deal with the matter of SoftMax loss is to use an auxiliary loss to coach the neural network. To treating missing and Outlier values we have used a loss function named categorical crossentropy. For each iteration, a selected loss function is employed to gauge the error value. So, to treating missing and Outlier values, we have used a loss function named categorical crossentropy.

3.2Emotion Detection Module:

Face Detection:

Face detection is one of the applications which is considered under computer vision technology. This is the process in which algorithms are developed and trained to properly locate faces or objects in object detection or related system in images. This detection can be real-time from a video frame or images. Face detection uses such classifiers, which are algorithms that detect what's either a face (1) or not a face (0) in an image.

Classifiers are trained to detect faces using numbers of images to get more accuracy. OpenCV uses two sorts of classifiers, LBP (Local Binary Pattern) and Haar Cascades. A Haar classifier is used for face detection where the classifier is trained with pre-defined varying face data which enables it to detect different faces accurately. The main aim of face detection is to spot the face within the frame by reducing external noises and other factors. It is a machine learning-based approach where the cascade function is trained with a group of input files. It is supported the Haar Wavelet technique to research pixels inside the image into squares by function. This uses machine learning techniques to urge a high degree of accuracy from what's called "training data".

Feature Extraction:

While performing feature extraction, we treat the pre-trained network that is a sequential model as an arbitrary feature extractor. Allowing the input image to pass on it forward, stopping at the pre-specified layer, and taking the outputs of that layer as our features. Starting layers of a convolutional network extract high-level features from the taken image, so use only a few filters.

As we make further deeper layers, we increase the number of the filters to twice or thrice the dimension of the filter of the previous layer. Filters of the deeper layers gain more features but are computationally very intensive.

Doing this we utilized the robust, discriminative features learned by the Convolution neural network. The outputs of the model are going to be feature maps, which are an intermediate representation for all layers after the very first layer. Load the input image for which we want to view the Feature map to know which features were prominent to classify the image.

Feature maps are obtained by applying Filters or Feature detectors to the input image or the feature map output of the prior layers. Feature map visualization will provide insight into the interior representations for specific input for each of the Convolutional layers within the model.

Emotion Detection:

Convolution neural network architecture applies filters or feature detectors to the input image to get the feature maps or activation maps using the ReLu activation function. Feature detectors or filters help in identifying various features present in the image such as edges, vertical lines, horizontal lines, bends, etc. After that pooling is applied over the feature maps for invariance to translation.

Pooling is predicted on the concept that once we change the input by a touch amount, the pooled outputs don't change. We can use any of the pooling from min, average, or max. But max pooling provides better performance than min or average pooling. Flatten all the input and giving these flattened inputs to a deep neural network which are outputs to the class of the object. The class of the image will be binary, or it will be a multi-class classification for identifying digits or separating various apparel items. Neural networks are as a black box and learned features in a Neural Network are not interpretable. So basically, we give an input image then the CNN model returns the results. Emotion detection is performed by loading the model which is trained by weights using CNN. When we take the real-time image by a user then that image was sent to the pre-trained CNN model, then predict the emotion and adds the label to the image.

3.3Music Recommendation Module:

Songs Database:

We created a database for Bollywood Hindi songs. It consists of 100 to 150 songs per emotion. As we all know music is undoubtedly involved in enhancing our mood. So, suppose a user is sad then the system will recommend such a music playlist which motivates him or her and by this automatic mood will be delighted.

Music Playlist Recommendation:

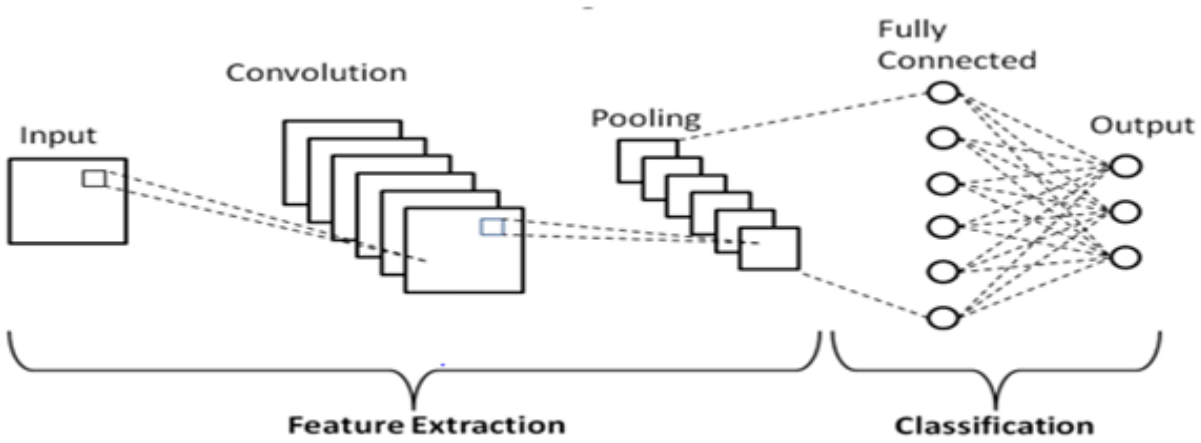
By using the emotion module real-time emotion of the user is detected. This will give the labels like Happy, Sad, Angry, Surprise, and Neutral. It will recommend playlist based on your emotions.

4.Algorithm:

Convolutional Neural Network is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolutional neural networks are widely used.

CNN takes an image as input, which is classified and process under a certain category such as dog, cat, lion, tiger, etc. The computer sees an image as an array of pixels and depends on the resolution of the image.

In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, filters (Also known as kernels). After that, we will apply the Soft-max function to classify an object with probabilistic values 0 and 1.



The three components that go into the convolution activity are:
- Input image
- Feature detector
- Feature map

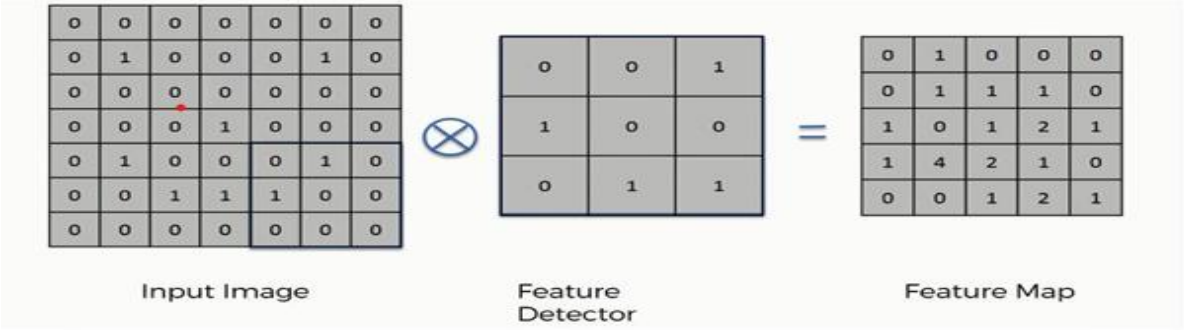


Fig: Feature Map generation through convolution operation

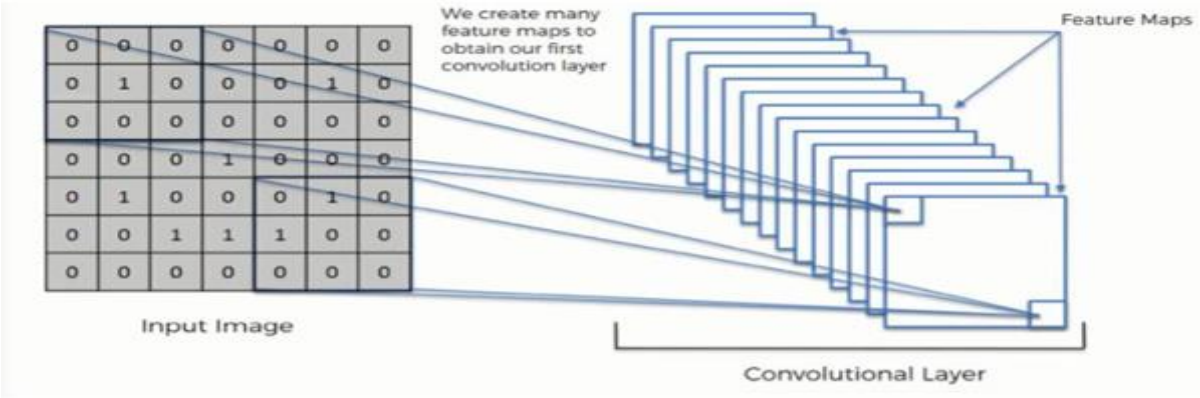


Fig: Creation of Convolution Layer

ReLu Layer:
After every convolution activity, CNN applies a Rectified Linear Unit (ReLu) function to the yield of the convolved picture. If the convolved image has negative values, it replaces them with '0'. It also introduces nonlinearity into the model.

Pooling Layer:
Pooling is the interaction where measurement of the convolved picture is decreased It does as such to diminish handling time and the registering power required. During this cycle, it ensures the fundamental component information. There are a couple of procedures that can be used for pooling. The most generally perceived ones are Max pooling and Typical pooling. In our application, we will use max pooling as it is the best an enormous segment of the events. Max pooling is fundamentally equivalent to the convolution cycle. A window slides over the component guide and thinks tiles of a predefined size. For each tile, max pooling picks the greatest worth and adds it to another component map. In this manner, the face highlights are separated utilizing convolution and pooling layers.

Output Layer:
The last fully connected layer is the yield layer which applies a SoftMax capacity to the yield of the past fully connected layer and returns a likelihood for each class.

Libraries Used

Pandas:
Pandas is a Python computer language library for data analysis and manipulation. It offers a specific operation and data format for handling time series and numerical tables. It differs significantly from the release3-clause of the BSD license. It is a well-liked open-source of opinion that is utilized in machine learning and data analysis.
NumPy:
The NumPy Python library for multi-dimensional, big-scale matrices adds a huge number of high-level mathematical functions. It is possible to modify NumPy by utilizing a Python library. Along with line, algebra, and the Fourier transform operations, it also contains several matrices-related functions.
Matplotlib:
It is a multi-platform, array-based data visualization framework built to interact with the whole SciPy stack. MATLAB is proposed as an open-source alternative. Matplotlib is a Python extension and a cross-platform toolkit for graphical plotting and visualization.
Scikit-learn:
The most stable and practical machine learning library for Python is scikit-learn. Regression, dimensionality reduction, classification, and clustering are just a few of the helpful tools it provides through the Python interface for statistical modeling and machine learning. It is an essential part of the Python machine learning toolbox used by JP Morgan. It is frequently used in various machine learning applications, including classification and predictive analysis.
Keras:
Google's Keras is a cutting-edge deep learning API for creating neural networks. It is created in Python and is designed to simplify the development of neural networks. Additionally, it enables the use of various neural networks for computation. Deep learning models are developed and tested using the free and open-source Python software known as Keras.
h5py:
The h5py Python module offers an interface for the binary HDF5 data format. Thanks to p5py, the top can quickly halt the vast amount of numerical data and alter it using the NumPy library. It employs common syntax for Python, NumPy, and dictionary arrays.

5.Results and Discussion:

- Upload Image with Face
- Pre-process & Detect Face in Image
- Detect Emotion
- Predicted Song
- Play song

6.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 right music tracks to improve the user's moods. The implemented system can detect the user's emotions. The emotions that the system can detect were happy, sad, angry, neutral, or surprised. After determining the user's emotion, the proposed system provided the user with a playlist that contains music matches that detected the mood. Processing a huge dataset is memory as well as CPU intensive. This will make development more challenging and attractive. The motive is to create this application in the cheapest possible way and also to create it under a standardized device. Our music recommendation system based on facial emotion recognition will reduce the efforts of users in creating and managing playlists.

7. Reference:

[1] Y. Kodama et al (2005)., "A music recommendation system, “Digest of Technical Papers. International Conference on Consumer Electronics, 2005. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1429796&isnumber=30841 .

[2] A. Niyazov, E. Mikhailova, and O. Egorova, "Content-based Music Recommendation System," 2021 29th Conference of https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9435533&isnumber=9435420.

[3] V. P. Sharma, A. S. Graded, D. Chaudhary, S. Kumar and S. Sharma, "Emotion-Based Music Recommendation System," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2021 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9596276&isnumber=9596065 .

[4] J. Lee, S. Shin, D. Jang, S. Jang, and K. Yoon, "Music recommendation system based on usage history and automatic genre classification," 2015 , https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7066352&isnumber- =7066289 .

[5] P. Ulleri, S. H. Prakash, K. B. Zenith, G. S. Nair and J. M. Kannimoola, "Music Recommendation System Based on Emotion," 2021, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9579689&isnumber=957 9470.

[6] Sheema Patro, P.N.V.S. Siva Dhanush, G. Sai Mahesh Music Recommendation System with Plagiarism Detection”2017-2021.

[7] In 2014 International Conference on Electronics & Communication Systems (ICECS -2014), Anaghia S. Dhavalikar and Dr. R. K. Kulkarni presented a paper titled "Face detection and facial expression recognition system".

[8] Krupa K S, Ambara G, Kartikey Rai, Sahil Choudhury, “Emotion aware Smart Music Recommender System using Two Level CNN”, 2020 Third International Conference on Smart Systems & Inventive Technology (ICSSIT).

[9] A study of face recognition based on LBPH & regression of Local Binary Features by Gao Xiang, Zhu Qiuyu, Wang Hui, Chen Yan, presented at the 2016 International Conference on Audio, Language and Image Processing (ICALIP).

[10] "Collaborative filtering for music recommender systems", 2017 IEEE Conference of Russian Young Researchers in Electrical & Electronic Engineering (EIConRus).

[11] In a paper entitled “Emotion based mood enhancing music recommendation”, presented at the IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017, Aurobind V. Iyer, Viral Pasad,Smita R. Sankhe, Karan Prajapati.

[12] IEEE International Conference on Robotics, Intelligent Systems & Signal Processing Proceedings. 2003. Shaoyan Zhang, Hong Qiao. "Face recognition with support vector machines.".