Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Emotion Based Music Recommendation System

Arti Bang, Shantanu Udavant, Omkar Malpure, Anuj Nagwekar, Atharva Pande
Vishwakarma Institute of Information Technology, Pune, Maharashtra, India

Abstract: - Today, music platforms provide easy access to many types of music. They continue to strive to improve music organization and research to solve the problem of selection and make discovering new music easier. Recommendations have become popular to help people choose the right music for every situation. But there is still a difference when it comes to personal and emotion-focused recommendations. Music is beneficial to humans and is widely used to relax, regulate emotions, eliminate stress and illness, and regulate mental and physical activity. Music therapy has many therapeutic areas and applications to improve health. This article will introduce the design of music recommendations that are guided by the user’s thoughts, feelings, and content of activities. We've crafted a Convolutional Neural Network (CNN) model tailored to recommend music based on the user's facial expressions. which help people choose music according to their different mood. It has Validation Accuracy of 97% and Testing Accuracy of 76%.

Keywords: Emotion, Indian Music, Machine Learning.

I. Introduction

Significant research has been conducted on the impacts of music on the human body and intellect. People have experienced many emotions from different types of music, and it has been believed since ancient times that music has an impact on personal development and the ability to heal disease [1]. Listening to music has an impact on people's thoughts and emotions, which in turn creates an impact on the body and mind, the health-promoting meaning of music is about to become popular. Many tests and studies have been conducted to understand the effects of music on brain function, that explains the effect of music on memory. Music therapy is considered an effective standard of care in treating depression. In addition to rational decision making, emotional factors also affect driving decisions [2]. Emotional intelligence will be able to better understand people's needs and emotions and choose appropriate music according to the background of emotions. Emotions associated with music are generally thought to come from two main concepts: emotions that can be found in music (emotional awareness) and emotions that can be seen through music: Music (thought emotions). All processes in the human body are interrelated, so emotions and psychology can influence each other. The main goals of the recommender are solving problems, finding new music, improving physical and mental health. This article shows one way to use the emotion-focused personality in a beautiful way. The following sections describe the application form and explain the problem areas. Section 3 contains a description of data sets. Section 4 describes system architecture. Whereas Section 5 explains the methodology used in the experimental model of the proposal. And in the last part of the article, the results and further scope are presented.

II. Literature Survey

Mammadli, R., Bilgin, H., & Karaca, A. C [3] describes a web application using machine learning to analyze users' emotions from camera images. It employs CNN models to make predictions, comparing different layer combinations for accuracy. Data preprocessing involves cleaning datasets and discarding irrelevant classes. Additionally, the system suggests music playlists based on the determined attributes. Overall, it outlines the implementation of deep learning models for a web-based music recommendation system.

The article [4] describes music recommendation systems, with a focus on emotion recognition in music. It addresses challenges in defining emotions and proposes a model categorizing them based on valence and arousal. It suggests using social tags, lyrics, and music reviews for data collection instead of costly human annotation.

Additionally, it mentions context-based information retrieval models that utilize public opinions from social networks like Facebook, YouTube, and Twitter to recommend music based on user interactions.

Article [5] Research on emotion recognition through facial expressions is gaining traction, with image processing algorithms being key. These algorithms have broad applications in fields like medicine and human sciences. The aim is to create robust algorithms for accurate emotion classification, aiding advancements across industries. While some systems show promise in real-time testing, challenges like requiring well-lit images and minimum resolution persist. Nonetheless, progress continues in leveraging image processing for extracting and using user emotions in diverse applications.

Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay [6] document explores mood and polarity classification in Hindi song lyrics for music information retrieval. It utilizes sentiment lexicons, text stylistic features, and N-grams for classification, achieving a maximum F measure of 68.30% for polarities and 38.49% for moods. It underscores the importance of lyrics in predicting music moods, especially for Hindi songs, which dominate music sales in India. Insights are provided into developing mood and polarity classification systems based on semantic and textual features extracted from lyrics. whereas our system delivers 97% of validation accuracy and 76% of testing accuracy in music recommendation.

III. Data Sets

We use the FER-2013 dataset to develop a model that classifies users' emotions. The Facial Emotion Recognition 2013 dataset (FER-2013),[7] is used to analyze user emotions. FER-2013 contains 32,289 gray face images with a size of 48x48 pixels. Inputting such low-resolution grayscale images provides a small number of parameters in the training model. The music classification system utilizes four distinct emotions – happiness, sadness, anger, and neutrality – to recommend Hindi songs to users. The provided documents, sourced from various channels, are in .png format. It's important to note that our focus in this article is solely on recommending Hindi songs to the user, aligning with their emotional states.

IV. Proposed System Overview

The planning process helps us express the interaction between user and musician. The reason for this machine is to capture the images with the advanced camera. The captured pictures are encouraged right into a neural arrangement that predicts sentiments [8]. Then use the requirements within the captured pictures to get a list of songs. The principal purpose of our device is to provide a music playlist so that it will change the user's mood, which may be neutral, sad, surprised, or happy. The proposed system recognizes special attitudes and if a person has a positive feeling, it's going to present a list of songs that incorporate the most appropriate track that makes the person feel fantastic. There are four modules of facial recognition-based music recognition [9].

I. Real-Time Capture: In this model, the machine needs to seize the user's face.

II. Face Recognition: Right here, the person's face might be taken as an idea. Convolutional neural networks are programmed to extract features of user snapshots.

III. Emotional recognition: In this phase, unique behaviors are identified by analyzing user interactions, allowing the system to generate text that matches the user's mood. Additionally, the music recommendation feature draws from the user's genre preferences to suggest songs through an advice module. Figure No: 1. is the block diagram of the proposed system that we are using in our model.

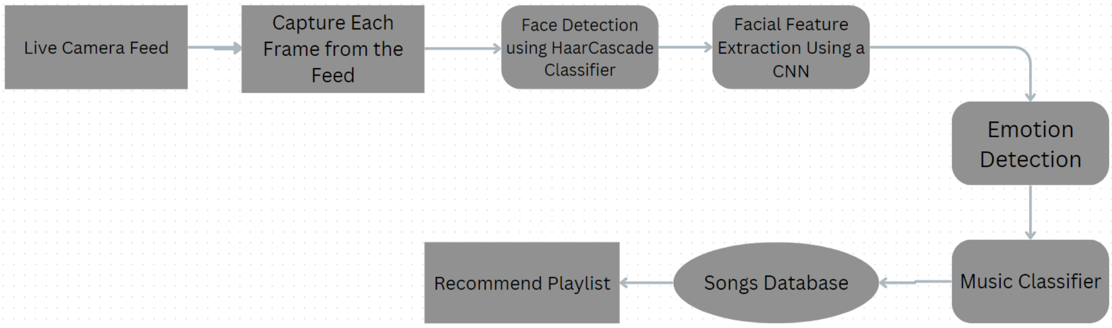

Figure No: 1. Block diagram of the proposed system.

V. Methodology

5.1 Database Description

We utilized the Kaggle dataset to create a convolutional neural community [8]. The tutoring data and testing metrics are separated into two sections of the FER2013 database. The preparation file has 24176 images, while the test file has 6043 images. This document incorporates a 48x48 pixel grayscale picture of a human face. Each picture in FER-2013 got to be labeled as one in all four temperaments: Happy, sad, angry, and neutral. Faces are recorded so that they're more or much less turned around and occupy approximately the same space in each photograph. FER-2013 includes both grayscale and 48x48 pixel exposed and unexposed avatars. Below is Figure No: 2, showcasing grayscale images depicting various user moods, all accurately labeled.

Figure No: 2. Samples from FER2013 dataset.

5.2 Emotion Detection Module

5.2.1 Face Detection

Computer technology has several uses, one of which is facial recognition. Detection of objects or related techniques in images is the manner of developing and training algorithms to appropriately locate faces or items [1]. This detection may be accomplished right away on a photo. Face detection uses clustering, which is the process of detecting the presence (1) or absence (0) of a face in an image. Several photos are used to train the classifier so that it can identify faces more accurately. Haar cascades and LBP (Local Binary Instances) are two learning methods used by Open CV [10]. The Haar classifier is used to recognize faces; to identify a specific face, the classifier is first trained using face transformation data. The primary objective of face detection is to identify human faces amidst external disturbances or distractions. Utilizing Haar wavelet technology, the system transforms pixel data into square representations, enhancing its ability to recognize input methods. These algorithms help in reducing computational load and improving the speed of detection, which efficiently detect faces by progressively applying a series of classifiers to different regions. We can observe the output in Figure No: 3 shows the images generated by the model that identifies the user's mood.

Figure No: 3. Emotion Detection

5.2.2 Feature Extraction

We use the pre-trained ensemble as a specific selection sample chain during the extraction procedure. Let the image progress, stop at the first layer, and use the output of this layer according to our specifications. The first process of convolutional ensemble extracts the features of the captured image and thus uses a series of filters. As the system gets deeper, we make the filter 2-3 times larger than the previous filter [11]. Deep filters have greater capabilities but are high-priced to add. To avoid this, we use robust, discriminative features discovered from a neural community. The output of this version can be a custom map that is the average of all layers after the first layer. Load the input image where we want to display the map to see which functions are important for the distribution of the image. The feature map is obtained by applying a filter or parameter to the input image or the feature map output from the previous layer. A map view will provide an internal representation of a unique input for each convolution layer in the model. Figure No 4 illustrates the visualization of the feature map after applying a series of filters.

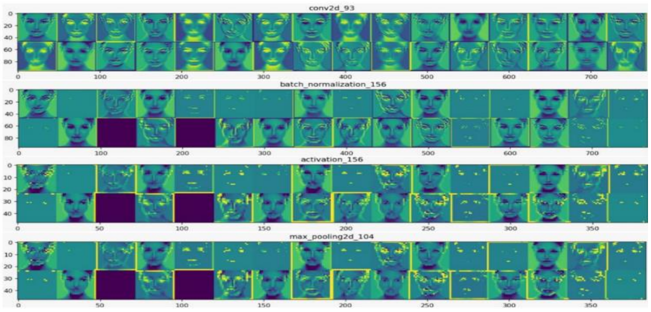

Figure No: 4. Visualization of The Feature Map

5.2.3 Emotion Detection

The Convolutional Neural Network (CNN) architecture employs filters to extract features from the input image, which are then processed using the Rectified Linear Unit (ReLU) activation function [3]. These filters act as function detectors, identifying edges, vertical lines, horizontal lines, folds, and other patterns. Subsequently, pooling is performed to downsample the feature maps. Pooling helps reduce computational complexity and control overfitting. The concept of pooling involves dividing the feature maps into smaller sections and applying operations like maximum, average, or minimum to each section. In this architecture, maximum pooling is preferred due to its superior performance compared to average or minimum pooling. Below is Figure No 5, depicting the addition of convolution layers and filters facilitated by the architecture.

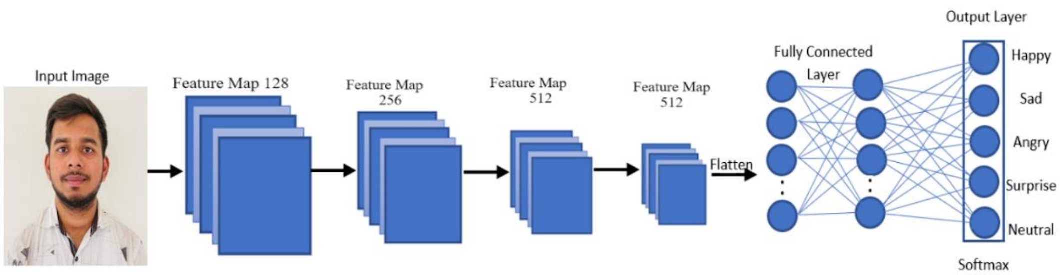

Figure No: 5. Architecture of Convolutional Neural Network

The figure below, labeled as Figure No 6, illustrates the feature extraction process of each layer within a Convolutional Neural Network.The neural community is sort of a black field and the features discovered inside the neural network can't be defined. So essentially, we took the entire photo, and the CNN model lowered the result [9]. Emotion popularity is accomplished by means of loading fashions skilled in the use of CNN weights. When we capture the person's photo, first the image is fed into the CNN version, which predicts the feelings and collects the photo.

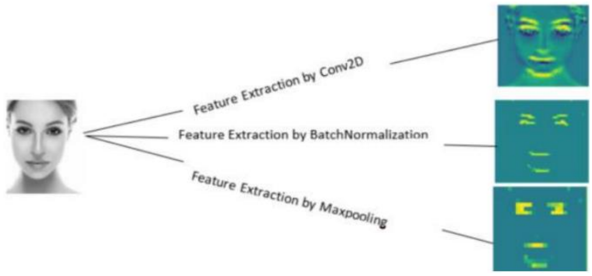

Figure No 6. Feature extraction of each layer in a Convolutional Neural Network

5.3 The Module for Recommending Music

5.3.1 Songs Database

We assembled a collection of Bollywood Hindi song lyrics. There are between 100 and 150 tracks for each feeling. If a user is happy then, the system will suggest a playlist of music that will cheer them up and make them feel joyful automatically [6]. Below is Table 1, presenting some of the songs from the database.

Table 1. Database of songs

Emotion	Songs
Happy	Track 1 “Dil Jhoom Jhoom”
	Track 2 “Ajab Si”
	Track 3 “Yeh Jawani Hai Deewani”
Sad	Track 1 “Tu Jaane Na”
	Track 2 “Aye Dil Hai Mushkil”
	Track 3 "O Bedardeya”
Angry	Track 1 “Arambh”
	Track 2 “Sadda Haq”
	Track 3 “Ravan Ravan Hoon Mai”
Neutral	Track 1 “Aap Ki Nazron Ne Samjza”
	Track 2 “Kabhi Kabhi Mere Dil Main”
	Track 3 “Kabira”

5.3.2 Music Playlist Recommendation

The mood module can identify the user's mood. This results in notes such as happy, sad, angry, surprised, and neutral. We add these tags to the folders in the music library we created using Python's os.listdir() function. Table 1 shows some of the song titles. You can use os.listdir() to get a list of all files in a directory. As a consequence, the GUI of the music player will display the user's recommended playlist together based on the emotions that are identified. To play the audio, we utilized the Pygame library, which can play a variety of multimedia types, including audio and video. The play, pause, resume, and stop songs in this collection are utilized in conjunction with the music player. The names of all songs and the state of songs that are presently playing are stored in variables called playlist, song status, and root.

VI. Result & Analysis

We evaluate various tasks using Support vector machines (SVM), Extreme learning machines (ELM), and convolutional neural networks [8]. Table 2 presents a comparison of the relevant algorithms. Algorithms and accuracy are provided for all searches. Using convolutional neural networks to improve cognitive performance.

Table 2. Validation and Testing accuracy for the three algorithms on the Fer2013 Dataset

Algorithm	SVM	ELM	CNN
Validation Accuracy	0.65	0.63	0.97
Testing Accuracy	0.67	0.62	0.76

The CNN network's training hyperparameters are displayed in Table 3. Learning keeps its weight current at the conclusion of every batch. The batch size is the number of samples that enter the network prior to the weight being altered. During training, all training data is repeatedly iterated into the network. The model can learn nonlinear predictions through dynamic activation. A common technique for determining deep learning model faults in single-label, multi-category classification problems is categorical cross-entropy loss.

Table 3. Hyperparameter for trained CNN network

Hyperparameters	Values
Batch size	128
No. of classes	5
Optimizer	Adam
Learning rate	0.001
Epoch	48
Number. of Layers	28
Activation function	Relu, SoftMax
Loss function	Categorical-cross entropy

VII. Future Scope

The future of emotion-based music recommendation systems holds promise for advancements in accuracy and personalization. Key areas of development include fine-tuning emotion recognition algorithms, incorporating multi-modal data, and integrating real-time adaptation. Ethical considerations such as privacy and consent will be paramount. Integration with wearable devices, collaborative filtering techniques, and therapy applications could further enhance user experiences. Cultural nuances and exploration of new music genres will also play a significant role in future developments.

References

[1] Alrihaili, A. Alsaedi, K. Albalawi, et al., “Music recommender system for users based on emotion detection through facial features,” in 12th International Conference on Developments in eSystems Engineering (DeSE), pp.1014-1019,2019.

[2] R. Ramanathan, R. Kumaran, R. R. Rohan, et al., “An intelligent music player based on emotion recognition,” in 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), pp.1-5,2017.

[3] “Music Recommendation System based on Emotion” Ramiz Mammadli 1 , Huma Bilgin 2 , and Ali Can Karaca 3; Cornell University, December 2022.

[4] “A Survey of Music Recommendation Systems and Future Perspectives”, Yading Song, Simon Dixon, and Marcus Pearce.

[5] “Emotional Detection and Music Recommendation System based on User Facial Expression To cite this article”: S Metilda Florence and M Uma 2020 IOP Conf. Ser.: Mater. Sci. Eng. 912 062007.

[6] “Mood Classification of Hindi Songs based on Lyrics”; Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay; ALC Anthology December 2015.

[7] lhasan Ali Alharbi(B), Mukta Dhopeshwarkar(B), and Shubhashree Savant(B "Detection of Emotion Intensity Using Face Recognition" Springer Link, February 2021.

[8] . I. J. Goodfellow, D. Erhan, Y. Bengiou, et al., "Challenges in representation learning: A report on three machine learning contests," in Neural Information Processing, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 117–124 2013.

[9] . Z. Qin, F. Yu, C. Liu,et al, "How convolutional neural networks see the world - A survey of convolutional neural network visualization methods," 2018.

[10] . V. Tabora," Face detection using OpenCV with Haar Cascade Classifiers," Becominghuman.ai,2019.

[11] . F. Norden and F.V.R. Marlevi, "A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition, "TRITA-EECS-EX:143 pp.9,2019.