A Study on Deep learning Approaches for Mood Based Music Endorsement Systems

Pendela Kanchanamala[1], Tummidi Vinay Sriram[2], Patibandla Manideep[3], Ravva Dedeepya[4], Yadla Jaswanth[5]

[1]Associate Professor, [2]19341A12B8, [3]19341A1286, [4]19341A12C3, [5]19341A1C6

Department of IT, GMRIT, AP

ABSTRACT

In this present competitive world people are so busy and engaged in many works and in that process they are undergoing through various emotions like Sad ,Happy, Anger etc. In order to cope up with the varying emotions the best activity that people choose is listening to music. Being able to suggest a song or music playlist enhances the listening experience because music elicits strong emotions and feelings in people. There exist some music recommendation systems in order to recommend the music to the users we use techniques such as Collaborative filtering and models such as Context-Based Information Retrieval system. These early models have some limitations such as dynamic evolvement, playlist generation and they are not personalized systems. In this Paper it is proposed to develop a prototype to endorse dynamic music system based on human emotions. The main objectives for this paper is to propose a model for Mood derivation of the user using Haar Cascade Algorithm and endorsing the music to the user based on the derived mood using facial landmark extraction.

Keywords: *Face Detection, Open CV, Dynamic Music, Human Emotion, Landmark Extraction.*

DOI Number: 10.14704/nq.2022.20.9.NQ44768 **Neuro Quantology 2022; 20(9):6574-6580**

INTRODUCTION

Individual interaction and communication are important aspects of daily living. It communicates precise information and a vast amount of data between people, whether through words, tone, or emotion. Particularly, the expression on someone's face is one of the fundamental components of human communication and is regarded as a type of nonverbal communication. These movements are intended to convey many facial expressions, including pleased, sad, natural, and astonished. Face detection is described as a two-step procedure that starts with looking for faces in images or videos, regardless of whether they belong to a person or not. It then creates a box around each face in the image, taking into account their expression, orientation, facial position, illumination, and occlusions. The primary method for determining the emotions on a person's face is emotion detection, which is regarded as one of the facial expressions. The main objective of this work is to develop a system that can recognise user emotions, such as joy, sadness, surprise, or natural emotions. Following the identification of the emotion, the suggested system will then provide the user a music playlist that includes music samples of particular music genres that elevate their mood. There are the processes that will be taken, such as face detection using a camera, facial landmark extraction, facial emotion detection, and connecting to a music player. The Haar cascade method and computer vision for emotion identification and music recommendation are utilized to analyze the experimental data.FER2013 is the dataset that was used. CNN (Convolution Neural Network) is the algorithm used only for training facial image datasets. Every time, users must manually choose songs based on interest and mood. When play-styles change, users find it difficult to rearrange and play music. Therefore, we have employed the notion of machine learning, which includes facial scanning and feature tracking, to ascertain the user's mood and then provide a customized playlist based on it.

2. LITERATURE SURVEY

The following papers used a variety of algorithms, including SVM, YOLOV3, COSINE ALGORITHM, and others, to play music based on a person's mood, with the exception of neutral and angry expressions.

Authors Shalini, Shantha K., et al. in this research introduces a Facial Emotion Based Music Recommendation System employing computer vision and machine learning approaches. The automatic face recognition technology is explained in this paper. The three sessions—Face detection, Feature Extraction, and Expression Recognition—are explained. They suggested using the active appearance model (AAM) approach to extract facial features such as the mouth, lips, eyes, and brows. This essay analyses and recommends fitting Bezier curves. They proposed to extract the region of interest from the input facial photos and used to extract the facial features from the original facial input images. To prepare the input image for the feature extraction process, the colour is first modified. The input photos are trained for the purpose of detecting face emotions using the classification method OpenCV. The accuracy that this paper has given is 92% [1].

Authors by Vinay p, Raj Prabhu, Bhargav et al. in this research introduces a Emotion Based Music Recommendation System. In this paper the entire work has been split into two sections they are 1. The module for facial expression recognition. Module for music suggestions 2. Identification of Users Mood using the webcam by Using the SVM supervised ML technique, a model was created . In order to anticipate the user's sentiment, an image is submitted, its features are extracted using Python's OPENCV (Open Source Computer Vision), and then these features are applied to an SVM training model. This music suggestion model uses the output from the facial expression recognition module as its input. To identify the song, a user's emotional response will be used as an input. The user can choose from a list of songs that are connected to the emotion and play it.

M Athavle et al author of this research suggested a music recommendation system based on facial emotion recognition. Support vector machines (SVM), extreme learning machines (ELM), and convolutional neural networks are all used in this study. The Kaggle Facial Expression Recognition dataset for emotion identification. Songs from Bollywood and Hindi have been used to generate the music player's data set. The database, FER2013, is divided into a training dataset and a testing dataset. There are 24176 photos in the training dataset and 6043 in the testing dataset. The collection contains 48x48 pixel grayscale portraits of people. Each image in FER-2013 is assigned to one of five emotions: surprise, happy, sad, or angry. Convolutional Neural Network is used to implement face emotion detection, with an accuracy rate of about 95.14 percent. [3].

The authors of this paper, Dr. Sunil Bhutada et al, an assistant professor, described a system for recommending music based on emotions. The principle component analysis (PCA) method is employed in the suggested model to reduce the face space dimensions. Fisher's linear discriminant (FDL) or the LDA method is then applied to acquire the feature of the image characteristics. Two algorithms are used in this investigation. They are the Haar cascade algorithm and the Fisher Face algorithm. Facial capture, face identification, emotion classification, and music selection are the steps. The dataset, called FER2013 from Kaggle, is really large and contains a lot of data. This paper's accuracy is approximately 89%, hence it does not provide accurate results for anger emotion. [4].

The authors of this research, H. Immanuel James1, J. James Anto Arnold2, et al conducted a thorough investigation into the EMOTION BASED MUSIC RECOMMENDATION SYSTEM. The project's results are divided into two phases:

Create a Python programme that can identify user emotion based on face expression. Play the music based on the face expression and incorporate the Python code into the web service. SVM classifier and linear classifier are used in this study to train the music system. According to this study, a linear classifier detects face features faster than an SVM classifier. The

procedures that were used Face segmentation, classification using SVM, emotion categorization, and music suggestion. About 90–95% of classifiers are effective. such that the system can recognise the face and the emotion being displayed even if there is any change in the face caused by environmental factors. [5].

The authors Ankita Mahadik , Shambhavi Milgir , Janvi Patel et al introduces a Android App that plays appropriate music for the emotion being recognized . The FER 2013 dataset and the MMA Facial Expression Recognition dataset from Kaggle were combined to create the dataset used for training. The grayscale photos in the FER 2013 dataset had a 48x48 pixel size. Images with various characteristics could be found in the MMA Facial Expression Recognition dataset. Keras to train and test our model for seven emotional states: surprise, fear, anger, neutrality, and disgust. It was trained for 25 epochs and has a 75% accuracy rate for mood recognition. This study divided the project into four modules: integration, mood detection model, and music suggestion. The project is an application that has been developed, so it needs a user interface so they used Android Studio. Overfitting in this trained model can occasionally cause variations in accurate detection. For instance, the "disgust" emotion is frequently categorised as the "angry" mood because both have comparable facial features (cheekbones, brows). A trained model with a 75% accuracy rate that can properly identify seven moods—anger, disgust, fear, happiness, sadness, and surprise—along with an Android app that plays music appropriate for the emotion being identified [6].

Authors Dhruv Piyush Parikh1 , Shatakshi Shree , Abhay Arora et al. divides the entire work into three sections in this paper: image capture and face detection, model training, and music redirection. They employed the segmentation method, which separates an image into groups of pixels by giving each pixel a label, for image capture and face detection. The contours in any given image can be recognised by grouping together the pixels with the same labels. Here, an image detection technique called Haar Cascades is applied. The algorithm outputs the coordinates that form a box with four corners around the face. CNN is a training model that is employed The model's features are preserved in a JSON file to prevent the model from needing to be trained again in order to save time and effort and achieve a similar level of accuracy before moving on to neural network training. To redirect to YouTube playlists, the redirecting songs module employs the pywhatkit function "pywhatkit.playonyt (keyword)". The define keywork has an associated playlist that is tied to the expression to increase personal emotional anxiety. The Haar Features can accurately and successfully identify faces in real-time video. The model's accuracy in identifying the appropriate emotion is 64.52%. [7].

The authors Akansha Bisht, Deepanshu Mishra, Harshit Gupta, el at. put up a research proposal for an algorithm for automatically recommending music based on facial microexpression recognition. They suggested a Collecting and processing micro- expression photographs of human faces; performing micro-expression detection and related feature extraction; and performing micro-expression classification. SVM classifier, Content-based Recommendation algorithm, Cosine Similarity algorithm, and CNN are the algorithms that are employed in this study. Convolutional, pooling, and fully linked layers are present in CNN. The experimental data set is the FER2013 public data set, and the experiment is run using the TensorFlow deep learning framework platform. Although the total accuracy is 77%, each emotion's accuracy varies from one another. For joy and surprise, 83%; disgust, 29%; anger, 51%; and fear, 23%. [8].

In this study Mrs. Indumathi S K, Sireesha K, Kavan MC et al. explained CNN architectures to create a real-time emotion-based music player. This study employs the FER2013 dataset, however it can also be applied to the MNIST, CIFAR-10, and ICML- 2013 datasets. The primary ideas discussed here are dataset preparation, data augmentation, class weighting, and ensembling. Algorithms like ResNet50, SeNet50, VGG16, the 5-layer model, and the GAP model are used. Among these, the GAP model

operated with an accuracy of 66.54%; after fewer parameters were used, the accuracy increased to 80%. This paper's primary flaw is that it doesn't investigate alternative multi-class SVM formulations. [9].

The authors Ahlam Alrihaili, A. Alsaedi et al. suggested an Emotion Detection using Facial Features-based Music Recommender System for Users. The The dataset being used, FER2013, contains a substantial amount of photos or data. The PCA and viola- Jonze algorithms are employed in this paper. The PCA's goal in the realm of emotion detection is to transform each image in a training set into an eigenface. Support Vector Machine (SVM) for feature vectors and multi class Adaboost during dynamic temporal warping are two classifiers that were employed in this study. The article relies on two types of geometric characteristics that depend on 52 landmark location and angle. The Euclidean Distance method is used to identify the retrieved characteristics. [10].

The authors Mansi Wairkar et al. proposed a MUSIC RECOMMENDATION SYSTEM BASED ON FACIAL EXPRESSIONS in this research. FER2013 is the dataset that was utilised. Face detection, emotion detection, the Haar cascade classifier, Open-CV, machine learning, and the Fisher Face Algorithm are the main algorithms and components in this article. The Fisher face algorithm offered by the CV2 library is used to determine the user's mood. Based on the principles of LDA and PCA approaches, this algorithm operates. Anger, Neutral, and Sad were the next most accurately identified emotions for the model created with the use of the Fisher face algorithm, with accuracy rates of 95%, 92%, 86%, and 80%, respectively. Based on the confidence and prediction values during the testing, the accuracy results are determined. In this, the accuracy is valued more highly the lower the confidence value is. [11].

The authors of this research, A. Kulkarni, S. Prajwal, H. Jayanthi et al. suggested a Music Recommendation System Using Speech-Based Emotion Recognition. AffectNet is the dataset that was used in this study. There are millions of images in there. There are seven class labels in the dataset representing states including anger, happiness, sadness, disgust, neutrality, confusion, and calmness. Four parts make up the project: face capturing, face identification, music recommendation, and integration. With a 5-fold cross validation accuracy of 85.3%, a Multi-Layer Perceptron can accurately identify an individual's emotion from their speech. The recommender is then given labels by using K-means clustering. The recommender model is finally formed by Light Gradient Boosted Machine Classifier, which has a 5-fold cross validation accuracy of 99.17%. [12].

The researchers Agrawal Abhishek Pradeep et al.suggested an Emotion Based Music Player in this study. The dataset includes faces of men and women of various ages and emotional states. Face detection is performed using the Media pipe Deep learning-based face detection algorithm. Artificial Neural Networks and KNN regression are the techniques used for music recommendation. The output of facial detection was sent into a system that recommends music, which then plays music in response. This model has an accuracy rate of about 88.5%. However, this model has restrictions for neutral and aggressive facial expressions. [13].

The authors Gokul Krishnan K et al. proposed an EMOTION DETECTION AND MUSIC RECOMMENDATION SYSTEM USING MACHINE LEARNING in this

work. VGG'16 utilises 3400 flip-flops, but CNN2 uses 10,000 times less flip-flops while maintaining the same efficiency. This project consists of five modules: a splash screen demo, dynamic permissions, an emotion API, a music recommendation system, and an API for youtube. A user interface for users to grasp is provided by the splash screen demo, and dynamic permissions enable hiding of sensitive data up to a certain point. Camera detector, detector listener, set detection, and get detection are all parts of the emotion API. The Viola-Jones object detection framework algorithm has been employed. By creating the integral image first, the Viola-Jones method may function much more effectively The accuracy of emotion recognition increases with the use of CNN2. When compared to CNN2, the VGG'16 is 0.74 percent more

accurate, but CNN2 needs 3.57 MFLOPs, while VGG'16 needs 31300 [14].

The Author S. Metilda Florence and M. Uma et al. People frequently use their facial expressions to convey their feelings. A person's mood has long been known to be changed by music. One can gradually quiet their thoughts and ultimately have a nice effect by capturing and understanding the emotion being shown by a person and playing music that suits that person's mood. Users were asked to complete a survey based on three criteria: 1. What kinds of music do they enjoy listening to when they are joyful? 2. What kinds of music do they prefer to listen to when they're depressed? 3. What kinds of music do they want to hear while they're upset? For this project's implementation, two datasets have been considered: CK+ Dataset - The Cohn Kanade Extended dataset was used to train the classifier. 593 facial action coded sequences from 123 participants make up the dataset. HELEN About 200 photos make up the dataset, which is used to train the classifier. For each image included in the dataset, 164 landmark coordinates are provided in the form of a.txt file with the photographs. [15].

3.WHY MUSIC IS DIFFERENT

When developing MRS, it is important to keep in mind that music recommendation has some unique properties compared to other domains where recommender systems are used, such as items, movies, or hotels. Some of these unique aspects of the music industry have an impact on how recommender systems are used, and they also directly drive the usage of deep learning techniques. A music track's duration is significantly less than that of a movie, vacation, or use of a product, to start. Second, there are tens of millions of tracks in commercial music libraries, which is a significant quantity of musical pieces. Because of these two factors, music nowadays can be seen as being more disposable than ever. Short consumption times and a large selection of songs suggest that proposing a few songs that don't exactly match the user's taste won't usually have a significant negative impact on the user experience. This is in contrast to movie suggestions, for example, where people are more angry when they receive a terrible recommendation because it takes them a lot longer to realise they don't enjoy the movie.

4. METHODOLOGY

Deep neural networks (DNN) are frequently used in recommender systems research that incorporates DL in the music industry to extract song or artist representations (embeddings or latent factors) from audio recordings or textual metadata, such as artist biographies or user-generated tags. Then, either directly using these latent item factors in CBF systems like nearest neighbour recommendations, incorporating them into matrix factorization techniques, or utilising them to create hybrid systems, which most frequently combine CBF and CF techniques, are the three main options available.

According to Van den Oord et alstudy,'s each song is represented by 50 latent variables that are learned from audio data using a convolutional neural network with rectified linear units (ReLU) and no dropout. [16], is likely the earliest work that uses DL for content-based MRS. They use brief audio clips of music from tracks in the Million Song Dataset (MSD) that were downloaded from 7digital5 as input [17]. After that, log- compressed Mel spectrograms (128 frequency bands, 23 ms window size, 50% window overlap), which were produced from a selection of randomly chosen 3-second audio recordings, were used to train the CNN. Two algorithmic solutions are investigated: minimising the mean squared error (MSE) and the weighted prediction error (WPE). 382 K songs and 1 M users are used in the research. Play counts for user-song pairs are transformed to binary implicit feedback data (i.e., 1 if user u listened to item I regardless of the listening frequency; 0 otherwise). The outcomes show that a CNN with MSE as the target function performs best when 500 predictions per user are considered (AUC: 0.772). The effectiveness of linear regression was dramatically decreased by the use of vector-quantized MFCCs, a popular bag-of-audio-words format (AUC: 0.645).

5. COMPARISION TABLE :

Author	Approach	Advantages	Accuracy
Shalini, Shantha K., et al.	AAM approach	Easy to train the input images	92%
Vinay p, Raj Prabhu, BhargavSatish Kumar,	SVM supervised ML technique	Abstraction and Modularity.	75%
M Athavle	Extreme learning machines (ELM), and convolutional neural networks	Mood detection is easy	95.14%
Dr. SunilBhutada	Fisher FaceAlgorithm	Used to acquire the feature of the imagecharacteristics	89%
H. Immanuel James, J. JamesAnto Arnold, J.Maria Masilla Ruban	SVM classifier	Emotion being displayed even if there isany change in the face caused by environmental factors.	90-95%
Ankita Mahadik ,Shambhavi Milgir , Janvi Patel	Convolutional Neural Network(CNN)	The model can identify different emotions accurately.	75%
Dhruv Piyush Parikh1 , Shatakshi Shree, Abhay Arora	Convolutional Neural Network(CNN)	The Haar Features can accurately and successfully identify faces in real-time video.	64.52%
Akansha Bisht,Deepanshu Mishra, HarshitGupta, Nancy Srivastava, Harris Kumar	Content-based Recommendation algorithm	They suggested a Collecting and processing micro-expression photographsof human faces	77%
Mrs. Indumathi SK, Sireesha K, Kavan MC	ResNet50, SeNet50,VGG16,	multi-class SVM formulations gives moreaccuracy	66.54%
Ahlam Alrihaili, A. Alsaedi	PCA and viola-Jonze algorithms	The Euclidean Distance method is used toidentify the retrieved characteristics.	79%
Mansi Wairkar	Fisher FaceAlgorithm	Based on the principles of LDA and PCAapproaches, this algorithm operates.	92%
A. Kulkarni, S.Prajwal, H. Jayanthi, and K. S. SowmyaPradeep	Light GradientBoosted Machine Classifier	The recommender is then given labels byusing K-means clustering.	85.3%
AgrawalAbhishek	Mediapipe Deep learning-based face detectionalgorithm	After applying the PCA the accuracy of themodel has increased it is one of the advantage.	88.3%
Gokul Krishnan	Viola-Jones object detection algorithm.	The emotion of a person is determinedfrom their voice using a Multi-Layer Perceptron.	74%
S. Metilda	VGG'16	Abstraction: As the user's data is hidden with the password protected only, he/shecan be seen when it opened.	69%

REFERENCES

[1] Shalini, Shantha K., et al. "Facial Emotion Based Music Recommendation System using computer vision and machine learning techniques." Turkish Journal of Computer and Mathematics Education 12.2 (2021): 912-917.

[2] Vinay p, Raj Prabhu, Bhargav Satish Kumar, Jayanth, A. SUNEETHA, M. TECH (Ph. D) "Facial Expression Based Music Recommendation System", Vol.10, Issue 6, June2021, DOI10.17148/IJARCCE.2021.10682

[3] Athavle, M., Mudale, D., Shrivastav, U., & Gupta, M. (2021). Music Recommendation Based on Face Emotion Recognition. Journal of Informatics Electrical and Electronics Engineering (JIEEE), 2(2), 1-11.

[4] Rumiantcev, M., & Khriyenko, O. (2020). Emotion based music recommendation system. In Proceedings of Conference of Open Innovations Association FRUCT. Fruct Oy.

[5] James, H. I., Arnold, J. J. A., Ruban, J. M. M., Tamilarasan, M., & Saranya, R. (2019). Emotion based music recommendation system. Emotion, 6(3).

[6] Ankita Mahadik , Shambhavi Milgir , Janvi Patel , Vijaya Bharathi Jagan, Vaishali Kavathekar," Mood based Music Recommendation System" International Journal of Engineering Research & Technology (IJERT) ,Vol. 10 Issue 06, June-2021.

[7] Parikh, D. P., Shree, S., & Arora, A. Emotion Based Music Recommendation System.

[8] Yu, Z., Zhao, M., Wu, Y., Liu, P., & Chen, H. (2020, July). Research on automatic music recommendation algorithm based on facial micro-expression recognition. In 2020 39th Chinese Control Conference (CCC) (pp. 7257-7263). IEEE.

[9] Muhammad, S., Ahmed, S., & Naik, D. (2021, April). Real Time Emotion Based Music Player Using CNN Architectures. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.

[10] Alrihaili, A., Alsaedi, A., Albalawi, K., & Syed, L. (2019, October). Music Recommender System for Users Based on Emotion Detection through Facial Features. In 2019 12th International Conference on Developments in eSystems Engineering (DeSE) (pp. 1014-1019).

[11] Mansi Wairkar, "MUSIC RECOMMENDATION SYSTEM BASED ON FACIAL EXPRESSIONS" Volume 8, Issue 4, 2021 JETIR April 2021,.

[12] Kulkarni, A., Prajwal, S., Jayanthi, H., & Sowmya, K. S. (2022). Music Recommendation System Using Speech-Based Emotion Recognition. In Modern Approaches in Machine Learning & Cognitive Science: A Walkthrough (pp. 317-325). Springer, Cham.

[13] Alrais, R., & Elfadil, N. (2020). Support Vector Machine (SVM) for Medical Image Classification of Tumorous. Int. J. Comput. Sci. Mob. Comput, 9, 37-45.

[14] Parmar, P. H. (2020). Music Recommendation System Using Facial Expression Recognition (Doctoral dissertation, BIRLA VISHVAKARMA MAHAVIDYALAYA).

[15] Florence, S. M., & Uma, M. (2020, August). Emotional detection and music recommendation system based on user facial expression. In IOP Conference Series: Materials Science and Engineering (Vol. 912, No. 6, p. 062007). IOP Publishing.

[16] Van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. Advances in neural information processing systems, 26.

[17] Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The million song dataset in Proceedings of the 12th International Society for Music Information Retrieval Conference. Miami, October, 24, 591-596.