2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) | 979-8-3503-0570-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/RMKMATE59243.2023.10368948

FACE EMOTION BASED MUSIC RECOMMENDATION SYSTEM USING MODIFIED CONVOLUTION NEURAL NETWORK

[1]A Jaya Mabel Rani
Associate Professor, Department of Computer Science and Engineering, *Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences*, Chennai, Tamil Nadu, India.
jayamabelrania.sse@saveetha.com

M.Nivetha S[2]
[2]Student, Department of Artificial Intelligence and Data Science, *Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences*, Chennai, Tamil Nadu, India.

N M Jothi Swaroopan[3]
[3]Professor,
Department of EEE,
R.M.K. Engineering College, Chennai
jothi.eee@rmkec.ac.in

K Hari Kumar[4]
[4]Assistant Professor, Department of Computer Science, *St Jude's College*, Thoothoor, KanyaKumari District, Tamil Nadu, India.

***Abstract*—In this paper, proposed a new vision to personify music recommendation by face emotion recognition technology. Emotions play a crucial role in the way humans perceive and respond to music. This system aims to enhance the user experience by recommending music based on their real-time facial expressions, allowing for a more immersive and emotionally resonant listening experience. So here employed deep learning techniques to detect facial emotions accurately and use them as input to our music recommendation algorithm. The proposed method modified convolution neural network ensures user privacy and data security by processing emotions locally on the user's device without uploading any sensitive information to external servers. Through experimental evaluation, here demonstrate the effectiveness and usability of the emotion-based music recommendation system and finally shown this method is very effective with the help of accuracy and F1 score.**

***Index Terms*—Music Recommendation, Face Emotion, Deep Learning, Data Security, Modified Convolution neural network**

I. INTRODUCTION

Music is a powerful art form that has the unique ability to evoke a wide range of emotions in its listeners. The way individuals perceive and connect with music is deeply influenced by their emotional states. Understanding and harnessing this emotional aspect of music can significantly enhance the user experience in music recommendation systems [1]-[3]. Traditional music recommendation approaches often rely on user preferences, historical listening data, or collaborative filtering techniques, which may not fully capture the real-time emotional states of users. In recent years, the field of facial emotion recognition has seen significant advancements with the advent of deep learning techniques [4]- [6]. Facial emotion recognition models can accurately detect and interpret the emotional expressions exhibited by users while they interact with content, including music. This opens exciting possibilities for incorporating real-time facial emotion data into music recommendation systems. This paper, proposed a novel approach to personalized music recommendation using face emotion recognition technology. The primary objective is to generate an emotionally intelligent music suggesting system that tailors music suggestions basis on our current emotional state, as inferred from their facial expressions [7], [8]. By leveraging deep learning-based facial emotion recognition, main aim is to enhance the emotional resonance of music for users, leading to a more engaging and immersive musical experience [9], [10].

The key motivation behind our research lies in bridging the space between emotions and music recommendations [11]. By analyzing the user's facial expressions, it can obtain valuable insights into their emotional state, which can be effectively used to curate a playlist that aligns with their feelings. This approach also has the potential to address the cold-start problem faced by conventional recommendation systems for new users, as emotions are likely to remain consistent despite limited historical data [12], [13]. We recognize the importance of user privacy and data security, particularly when dealing with facial images. Hence, design the system to process facial emotions locally on the user's device without compromising their sensitive information. This decentralized approach ensures that no facial images or emotional data are uploaded to external servers, safeguarding user privacy.

II. RELATED WORK

Firstly, here proposed a robust facial emotion recognition model tailored for real-time application in music recommendation systems. Secondly, presented a novel music recommendation algorithm that seamlessly integrates facial emotion data to offer personalized and emotionally relevant music suggestions. Lastly, we develop an emotion-based music recommendation system that respects user privacy and ensures an enjoyable user experience. The remainder of this paper is structured as follows. Here first provides an overview of facial emotion recognition. Then detailing the data collection, model architecture, and training process. presents the music recommendation algorithm, outlining collaborative filtering, content-based filtering, and their integration with facial emotion data and finally concluded this system with future scope [14], [15].

a) Overview of Music Recommendation Based on Face Emotion Recognition

The music recommendation system based on face emotion recognition is an innovative approach that utilizes facial emotion analysis to provide personalized and emotionally resonant music recommendations. The core objective of this system is to enhance the user's music listening experience by tailoring music suggestions based on their real-time emotional state, as inferred from facial expressions. It detects user emotion like happy, sad, surprised, fear, anger [16]. If the emotions of the users are negative then the selected playlist play a motivational song and it recommend song to change the negative mood of the user.

III. COMPONENTS OF THE RECOMMENDATION SYSTEM

a) Real-Time Facial Emotion Analysis

The system employs Modified Convolution Neural Networks to analyze the user's facial expressions in real-time.

b) Emotion Driven Music Recommendation Algorithm

The system integrates the facial emotion data obtained from real-time emotion analysis with the hybrid recommendation algorithm. By incorporating emotional context, the system selects music that aligns with the user's current emotional state, enhancing the emotional resonance of the music listening experience.

c) Personalized Music Recommendations

The system uses the combined information from facial motion analysis and traditional recommendation methods to generate personalized music recommendations that align with the user's emotional state. This personalization ensures that the recommended music resonates with the user's feelings and preferences [17], [18]. These components are given in the following Figure 1.

IV DATABASE DESCRIPTION

The database comprises a large collection of facial images, each depicting an individual's facial expression representing a specific emotion. Here used Kaggle dataset to built the face emotion detection by Modified Convolutional neural network. This face expression data set has two parts training data, testing data. In training data 28,821 image of facial expression like angry, disgust, fear, happy, neutral, sad, surprise.

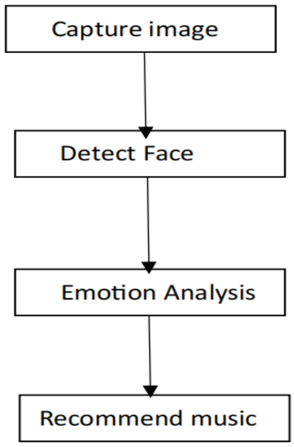

Fig. 1 Components of Recommendation System

IV DATABASE DESCRIPTION

In testing data 7,066 image is used for testing the data. The dataset undergoes preprocessing steps to standardize image sizes, normalize pixel values, and augment the data with techniques like rotation, scaling, and flipping. These steps help improve the model's robustness and generalization capability. The following figure 2 shows the sample of face emotion data.

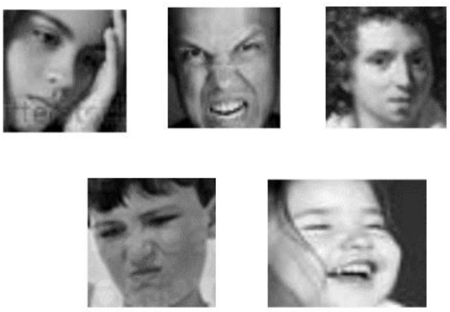

Fig. 2 Sample Data set for Face motion

V. FACE DETECTION AND EMOTION DETECTION

This type of detection works on the computer vision algorithm that aims to locate and recognize human faces within an image or a video stream. This technology plays a merging role in this advanced growing technology, including facial recognition, emotion analysis, surveillance, photography, and more In face detection it detects the face of the user and negligee the surrounding for better result. The face detection aims to remove the noise in the image data after detecting the face of the user the system detect the emotion of the face by comparing the trained data set. After detecting the face, we need to detect the emotion of the face, for this here used Modified Convolution Neural Network which is a deep learning algorithm MCNN is a process of recognizing and categorizing emotions present in facial expressions from images. It involves training a MCNN model on a labeled dataset containing images of various facial expressions corresponding to different emotions like happy, sad, angry, surprised, etc. The trained MCNN model learns to extract meaningful features from the images, enabling it to classify new facial expressions into relevant emotional categories. Construct a MCNN architecture that consists of convolutional layers, activation functions (e.g.,ReLU), pooling layers, and fully connected layers [19]- [20]. Convolutional layers help in extracting relevant features from the input images, while pooling layers reduce spatial dimensions and computational complexity.

The fully connected layers are responsible for mapping the extracted features to different emotion categories. Once the model is trained and evaluated, it is ready to make predictions on new facial expression images. The trained MCNN can classify these new images into specific emotion Training a face emotion recognition model is a complex task that involves several key steps, including data collection, preprocessing, model architecture design, training, and evaluation. To ensure your work is free from plagiarism, it's important to conduct original research and provide proper citations when referring to existing methodologies or datasets Implementing a suitable loss function for multi-class classification tasks is crucial for training a deep learning model effectively. One of the most commonly used loss functions for this purpose is the cross-entropy loss, also known as log loss Cross-entropy loss measures the dissimilarity between the predicted probabilities assigned by the model and the true class labels for multi-class classification problems. It quantifies how well the predicted probability distribution matches the actual distribution of classes in the data set. Here's the formula for cross-entropy loss for a single data point.

$$L(y, y\text{^}i) = -\sum_{i=1}^{c} \text{yilog}(y\text{^}i)$$

Developing a training loop for a deep learning model. involves several critical components, including forward and backward passes, weight updates, and the computation of evaluation metrics such as accuracy and F1-score. Below, provided the detailed explanation of each of these components.

a) Forward Pass

In the forward pass, feed a batch of training data through neural network to obtain predictions. Here are the key steps:

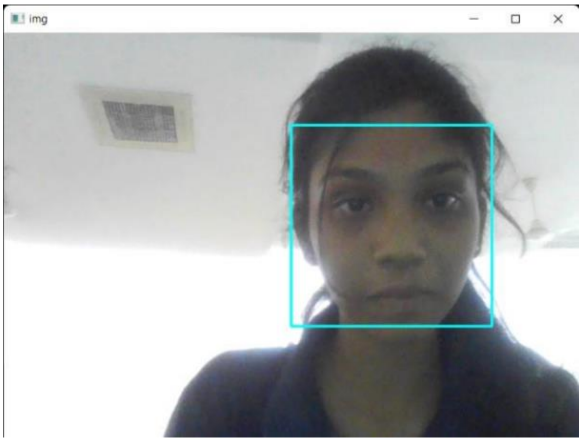

Fig. 3 Face Detection

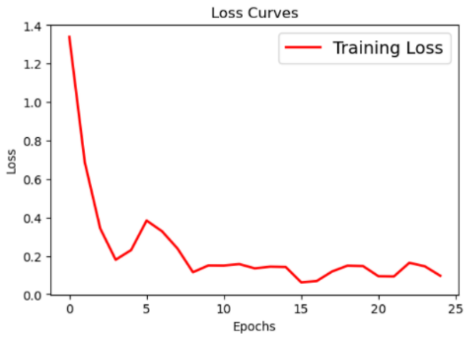

Fig. 4 Cross -Entropy Loss for Recommendation System.

i) Input Data

Take a batch of input data (e.g., images) and preprocess it as needed (e.g., normalize pixel values).

ii) *Forward Propagation*

Pass the preprocessed data through the layers of neural network. For each layer, calculate the weighted sum of inputs and apply the activation function. The final output of the network is the model's prediction.

iii). Loss Calculation

Compute the loss (e.g., crossentropy) by comparing the model's predictions to the actual target labels for the batch. This loss quantifies the error in the predictions.

2) Backward Pass (Backpropagation)

The backward pass calculates the gradients of the loss with respect to the model's parameters, enabling weight updates during training. Here are the key steps:

i) Gradient Calculation

Compute the gradient of the loss with respect to each parameter in the network. This is typically done using the chain rule and backpropagation algorithm.

ii) Gradient Descent

Use the gradients to update the model's parameters. A common optimization algorithm is stochastic gradient descent (SGD), which adjusts the weights in the direction that minimizes the loss.

c) Weight Updates

i) Learning Rate

Multiply the gradients by a learning rate, which controls the step size for weight updates. A suitable learning rate is crucial for convergence; it prevents overshooting or getting stuck in local minima.

iii) Update Weights

Update the weights and biases of the network's layers using the learning rate and gradients. This step nudges the model parameters toward values that minimize the loss.

d) Performance Evaluation

During training, want to monitor the model's performance using appropriate evaluation metrics. Two common metrics are accuracy and F1-score.

i) Accuracy

Calculate the ratio of correctly classified samples to the total number of samples in the current batch. This metric gives you an idea of how many predictions are correct [21]-[23].

ii) F1-Score

The F1-score is a harmonic mean of precision and recall. Precision measures how any of the predicted positive instances were actually positive, while recall measures how many of the actual positives were predicted correctly. The F1-score balances these two metrics.

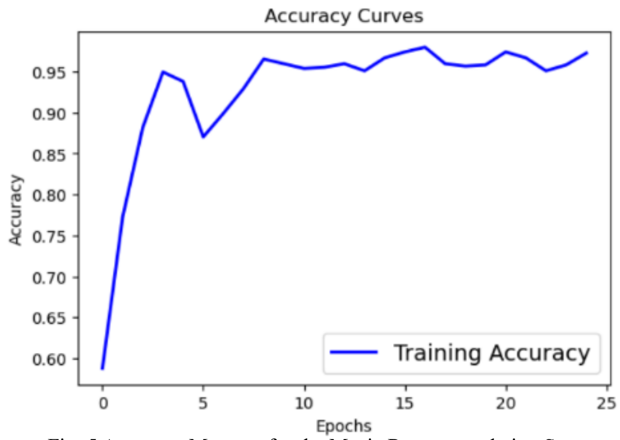

Fig. 5 Accuracy Measure for the Music Recommendation System

iii) F1-Score

Epochs and Batch Iteration

Repeat the forward pass, backward pass, weight updates, and metric calculations for multiple iterations over the entire training dataset. Each pass through the entire dataset is called an epoch. Within each epoch, data is typically divided into batches, and the model is updated after processing each batch. This batch-wise training helps with computational efficiency and convergence.

e) Early Stopping and Model Saving

Implement early stopping to prevent overfitting. Monitor a validation dataset, and if the validation loss stops improving or starts increasing for a certain number of epochs, halt training to prevent further overfitting. This can also save the model checkpoints periodically during training.

g) Training Termination

Decide on a termination criterion, such as a maximum number of epochs or a desired level of performance. Once this criterion is met, stop training.It's important to note that this training loop is a simplified representation. In practice, you may need to incorporate additional techniques like learning rate scheduling, regularization, and mini-batch handling for efficient and effective model training. Additionally, optimizing the training loop for hardware acceleration (e.g., GPUs) can significantly speed up the training process. After training the data we can detect the face emotion recognition using computer vision.

The epoch loss and accuracy table is given in the following table 1.

TABLE 1: THE EPOCH LOSS AND ACCURACY TABLE.

EPOCHS	LOSS	ACCURACY
1	0.7393	0.5879
2	0.6870	0.7723
3	0.3426	0.8818
4	0.1801	0.9496
5	0.2305	0.9380
6	0.3840	0.8703
7	0.3281	0.8991
8	0.1153	0.9654
9	0.1506	0.9597
10	0.1501	0.9539
11	0.1574	0.9553
12	0.1353	0.9597
13	0.1440	0.9510
14	0.1424	0.9669
15	0.0691	0.9741

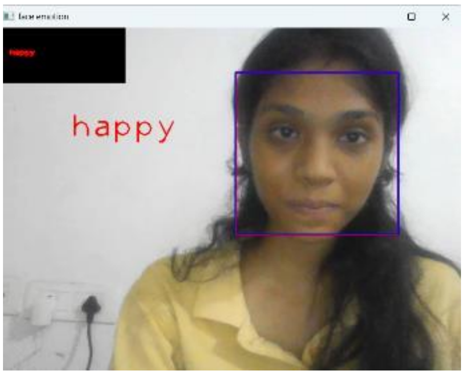

Fig.7 Face Emotion Recognition

VI. MUSICS RECOMMENDATION

Music recommendation systems that incorporate face emotion detection add a unique and personalized dimension to the way music is suggested to users. By analyzing facial expressions, these systems can intuitively gauge the emotional state of a person and tailor music selections that resonate with their current feelings. This integration of facial emotion detection and music recommendation offers a more immersive and emotionally relevant music listening experience. Once the emotional state is determined, a predefined mapping between emotions and musical characteristics comes into play. Certain genres, tempos, and tonalities are commonly associated with specific emotions. For instance, upbeat and cheerful songs might be recommended for happy expressions, while slower and more melodic tunes could be suggested for a sad expression. This mapping is created through a combination of user preferences and existing music theories.

Fig. 6 Music Player

VII. CONCLUSION

In this recommendation system, explained how music affects the user's mood and how to choose the right music to improve the user's mood. A practical system can capture the user's emotions. The motions the body can capture are happiness, sadness, anger, neutrality, or surprise. After determining the user's mood, the system prepares provide the user with a playlist with music matching the desired look. Processing large files requires a lot of memory and CPU. This may be complicated of this development of this system. The motivation is to develop this application in the cheapest way and build it as a useful for facial recognition based music recommendations to reduce users' workload of creating and managing playlists.

VIII. FUTURE SCOPE

The future scope of a music recommendation system based on emotion detection is promising and has the potential to greatly enhance user experiences in the music streaming industry. By combining emotion detection with advanced algorithms and machine learning techniques, music recommendation systems can create personalized playlists that resonate with users' emotional states, leading to increased engagement and satisfaction. Here are some ideas for the future development of such a system: Fine-Grained Emotion Detection Instead of basic emotions (happy, sad, angry), the system could detect more nuanced emotions like nostalgia, excitement, etc. This would require a more complex emotion recognition model, possibly utilizing multimodal data sources like lyrics, audio features, and even user interactions. Real-Time Emotion Analysis is Integrating real-time emotion detection from users' facial expressions or biometric data (heart rate, skin conductivity) could provide immediate feedback and adjust the music recommendations on the fly. This could be especially useful for applications like workout playlists, meditation sessions, or interactive storytelling. Context-Aware Recommendations is Incorporating contextual information such as time of day, location, weather, and user activities can lead to more accurate recommendations.

References

[1]. Lu, J., & Yang, J. "Music recommendation based on emotion", "Proceedings of the 8th ACM Conference on Recommender Systems" pp. 347-350, 2016.

[2]. Yang, Z., Tan, X., Feng, J., & Wang, Z, "Emotion-aware music recommendation system based on brainwave signals", "12th International Congress on Image and Signal Processing", Biomedical Engineering and Informatics (CISP-BMEI), pp. 1-6, 2019.

[3]. Yang, S., Liu, M., Yuan, C., & Jia, J, "Music emotion recognition and recommendation based on deep neural networks", " Proceedings of the 2018 3rd International Conference on Image, Vision and Computing (ICIVC)", pp. 85-90, 2018.

[4]. Jang, H.J. and Cho, K.O., "Applications of deep learning for the analysis of medical data", Archives of pharmacal research, vol.42, no.6, pp.492-504, 2019.

[5]. Rani, Antony & Albert, Pravin, Rainfall flow optimization based K-Means clustering for medical data. Concurrency and Computation: Practice and Experience. 33. 10.1002/cpe.6308, 2021.

[6]. Kabir, H.D., Abdar, M., Khosravi, A., Jalali, S.M.J., Atiya, A.F., Nahavandi, S. and Srinivasan, D, "Spinalnet: Deep neural network with gradual input," IEEE Transactions on Artificial Intelligence, 2022.

[7]. Yang, Y., Li, X., Li, S., & Yang, J, " Emotion-based music recommendation using deep neural networks", "Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval", pp. 9-16, 2017.

[8]. Kim, Y., & Lee, J. H, "Emotion-based music recommendation system using deep learning", "IEEE International Conference on Big Data and Smart Computing (BigComp)", pp. 350-353, 2018.

[9]. Yang, S., Yuan, C., & Liu, M, "Music recommendation system based on emotions detected by facial expressions. IEEE Access", 8, 47748-47756, 2020.

[10]. Chollet, F, Deep Learning with Python. Manning Publications. [Book Link] "https://www.manning.com/books/deep learning-with-python", 2017.

[11]. Valstar, M. F., & Pantic, M. (2010), " A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", " IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), pp. 1583-1600, 2010.

[12]. Lee, J. H., Lee, Y. J., & Lee, S. W, "Emotion-based Music Recommendation by Detecting User's Physiological Responses", 2011.

[13]. Xue, Z., Wang, L., Gao, Z., & Zhang, D, "Emotion-Based Music Recommendation", 2014.

[14]. Turney, P. D., & Littman, M. L., "Music Mood and Theme Classification: A Hybrid Approach Using Lyrics and Audio", 2013.

[15]. Yang, Y. H., Lin, Y. C., & Su, Y. F, "Music Emotion Recognition: A Benchmark Comparison of Audio and Lyrics Features", 2012.

[16]. Sharma, A., & Yadav, "Facial Expression Recognition: State of the Art and Challenges. In Deep Learning for Image and Video Analysis", pp. 1-22, 2020.

[17]. Mollahosseini, A., Chan, D., & Mahoor, M. H, "Going deeper in facial expression recognition using deep neural networks", "Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) pp. 1-10, 2016.

[18]. Desai, S.D., Giraddi, S., Narayankar, P., Pudakalakatti, N.R. and Sulegaon, S., "Back-propagation neural network versus logistic regression in heart disease classification", In Advanced computing and communication technologies, pp. 133-144, Springer, Singapore. 2019.

[19]. Manur, M., Pani, A.K. and Kumar, P., "A Big Data Analysis Using Fuzzy Deep Convolution Network Based Model for Heart Disease Classification", International Journal of Intelligent Engineering and Systems, vol.14, no.2, pp.147-156, 2021.

[20]. Thanga Selvi, R. and Muthulakshmi, I., "An optimal artificial neural network based big data application for heart disease diagnosis and classification model", Journal of Ambient Intelligence and Humanized Computing, vol.12, no.6, pp.6129-6139, 2021.

[21]. https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-019-0389-#:~:text=This%20 maximum%20was%20 selected%20 observing, score%20and%20Dice%20Similarity%20Coefficient.

[22]. Jaya Mabel Rani A, Pravin A. Rainfall flow optimization based K-Means clustering for medical data. Concurrency Computat Pract Exper. 2021;e6308. https://doi.org/10.1002/cpe.6308.

[23]. R. A comparison study of validity indices on swarm intelligence-based clustering. IEEE Trans Syst Man Cybern B Cybern. 2012;42:1245-1256.