MOOD-BASED MUSIC RECOMMENDATION SYSTEM USING FACIAL EXPRESSION RECOGNITION AND TEXT SENTIMENT ANALYSIS

MADIPALLY SAI KRISHNA SASHANK[1], VIJAY SOURI MADDILA[2], PALETI KRISHNASAI[3] VIKAS BODDU[4], G KARTHIKA[5]

[1, 2, 4, 5] Department of Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed To Be University), Visakhapatnam- 530045, Andhra Pradesh, India

[3] Department of Computer Science and Engineering, Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram, Chennai, Tamil Nadu, India

E-mail: [1]madipallysasank@gmail.com, [2]vijaysouri4@gmail.com

ABSTRACT

In today's fast-paced world, everyone is under a ton of stress for different reasons. Listening to music to reduce stress and detox has become a regular activity among people of all ages. However, if the music doesn't suit the user's mood, it can have the reverse effect of aggravating the stress in the user's mind. Moreover, there are no music applications available to the users that recommend songs based on the user's mood or emotion. Hence, in this work, we propose a mood-based music player application that suggests songs based on the user's emotion. The application can detect three emotions: angry, happy, and sad. To detect the emotion, the user has the choice of taking a selfie/providing an old image of their face or write a text stating how or what they are feeling. The application uses Deep Learning models (Facial Expression Recognition and Text Sentiment Analysis) to predict the user's emotion and populates a playlist of songs based on the emotion of the user.

Keywords: *Stress, Text Sentiment Analysis, Facial Expression Recognition, Mood-Based Music*

1. INTRODUCTION

The modern world is full of stress and tensions and everyone from a teenager to a ninety-year-old is carrying tons of stress. A few top causes of stress in one's life are money, work, family responsibilities, the economy, personal relationships, health issues, personal safety, financial issues among other things [1, 2]. Recent statistics suggest that about 91% of Australians feel stressed out about at least one aspect of their life and about 86% of China's workforce reports regular stress issues. About 4.5 Lakh workers in the United Kingdom believe that stress is the reason for their poor health.

With growing emphasis on mental health in today's world, there have come up a plethora of ways to deal with stress. A few such ways are working out, yoga, meditation, watching movies, hanging out with friends, and listening to music. Many researchers across the world have suggested that music can help people lose their stress and concentrate better on their work. On the flip side, listening to music might annoy the user if the music does not align well with the user's mood. Hence, for music to be an effective stress buster, music aligning with the user's mood must be recommended to the user [3].

Currently, there are many music applications the users can choose from. The two most widely used music applications across the world are:

- Spotify: Spotify is the world's leading music application in terms of user satisfaction. Their recommendation system is the best in the world and this is the reason Spotify is so widely used.
- Apple Music: Apple Music is yet another music application used across the world. Apple Music is preferred by iPhone users over any other music application and Apple Music's recommendations are excellent.

The comparison of these applications with the new application being proposed in this work is given in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1. The table shows a few basic features of a music player and whether or not these applications have that feature.

Table 1. Comparison of Spotify, Apple Music, and Mood-Based Music Player

	Online Music Streaming	Playlist Creation	Search	Recommendations	Mood Based Recommendations
Spotify	Yes	Yes	Yes	Yes	No
Apple Music	Yes	Yes	Yes	Yes	No
Mood Based Music Player	Yes	Yes	No	Yes	Yes

As it can be seen from Table 1, there are not many music applications that can recommend songs to the user based on their mood. Hence, in this work, we propose a music player application that will recommend songs to the users based on their emotion. To classify the user's emotion, the application will use either the face image of the user or analyze the text the user enters. When the user clicks a selfie/provides an old face image or enters a text in the app stating anything they feel, the Deep Learning models in the application analyze the emotion in the image/text. Based on the emotion, the application suggests songs to the user. The application can detect the following three emotions from the input provided by the user: angry, happy, and sad. Testing the models has shown that detecting happy mood is the easiest whereas detecting angry mood is most difficult.

2. LITERATURE REVIEW

Today, music is present everywhere be it at a wedding, students studying for an exam, a festival, cricket matches, or any other social event. Music has become an integral part of our quotidian lives. Researchers across the world have suggested that one of the most essential functions of music is that it is an excellent mood enhancer [4, 5]. Hence, there has been huge emphasis on developing systems that can create playlists and suggest songs to users helping them relax and concentrate.

This idea of having a robust cum effective music player framework is not novel [6, 7] and the idea of suggesting music based on the user's mood is also not new. Robert E. Thayer [8], in his research, utilized different attributes of music such as tempo, pitch, intensity, and rhythm to classify emotion in music and defined emotions based on the stress and energy in the songs. Y. Song et. al. [9] used a Support Vector Machine (SVM) based approach to classify the different emotions in a piece of music. The four emotions used by them were happy, angry, sad, and relaxed. Mood Cloud [10] is a visualization tool that can visualize and then classify music into five classes in real-time. The five classes are happy, party, sad, aggressive, and relax.

Music players have been coupled with the human body to suggest music based on the body's response to the music. The Body Rest system [11] works by changing the speed (or tempo) of the music based on the heart rate of the user. Chung et al.'s Affective Remixer [12] utilizes foot tapping patterns and the conductance level of the user's skin (SCL) to deduce the response of the user to the music being played and suggests music accordingly. The PAPA music system [13] actively measures the heart rate of the user and also stores this data at the backend to suggest songs in the future.

An important aspect of music recommendations is the personal preference of the user [14, 15]. Different individuals coming from different backgrounds have a variety of responses to similar music. This personal response of a user may depend upon their culture, age, and gender [14, 16]. Hence, it is of high importance that music recommendations must be personalized to each user.

Mood-based music recommendation has become a growing area of research in the past few years. Stress can be handled in a plethora of ways including workout, yoga/meditation, muscle exercises, sleep, and music. Since music works as an excellent mood enhancer, it is essential to suggest music with the proper tempo and mood to satisfy the user [17]. Although there are a huge number of music applications, there are not many which allow users to select music based on their mood [18].

3. SYSTEM ARCHITECTURE

The flow of the application is shown in Figure 1. The user is given the option of choosing between providing a facial image or writing a text to analyze their sentiment. Then, the application will call the Flask API to analyze the user's emotion. Based on the user's emotion, the

Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

application will suggest songs to the user. The songs will be picked randomly from a music dataset manually categorized and classified by the developers of the application. Content-Based Song Recommendations will also be given to the user based on the song the user is currently listening to.

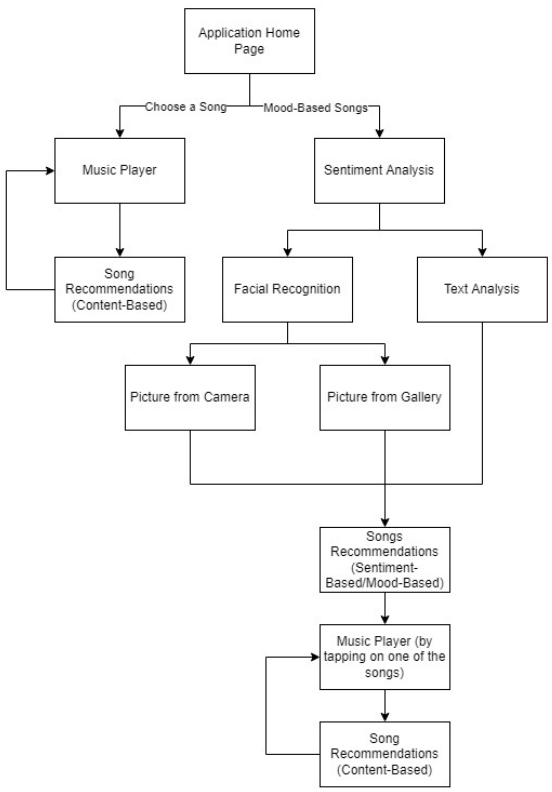


Figure. 1 Flow of the Music Application

4. PROCEDURE

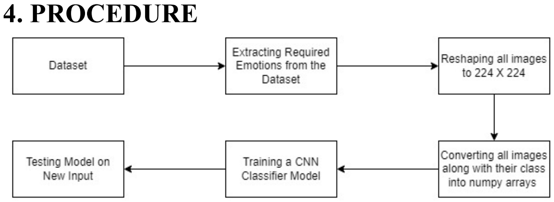


Figure. 2 Facial Expression Recognition Algorithm

The algorithm for the Facial Expression Recognition is shown in Figure 2. The dataset is first preprocessed to extract the required emotions (happy, sad, and angry) from all the classes in the dataset. Then, all these images are reshaped to 224 X 224 pixels. Once all these images are reshaped, the images are converted to numpy arrays and the output class of each numpy array is appended to it. Finally, a Convolutional Neural Network is trained on these numpy arrays and then the model is tested on new unseen input images for testing.

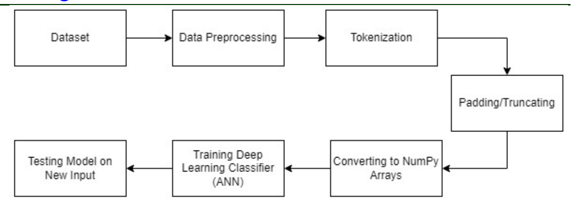


Figure. 3 Text Sentiment Analysis Algorithm

The Text Sentiment Analysis Algorithm is shown in Figure 3. The dataset is preprocessed to remove stop words and punctuations. Along with this, the required emotions (happy, sad, and angry) from all the classes are also extracted. Then, all these sentences are tokenized using Python's Tokenizer class. At this point, all the sentences are represented in the form of a sequence of numbers. Once the sentences are tokenized, all the sentences are padded (or truncated) to 100 words for uniformity. After padding/truncating, the sentences are converted to numpy arrays. Finally, these numpy arrays are fed to an Artificial neural Network for training. After training, the trained model is queried with new unseen input to test its functionality and dependability.

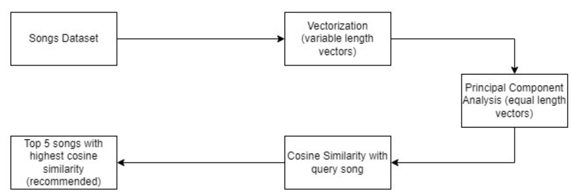


Figure. 4 Content-Based Recommendation System Algorithm

The Content-Based Recommendation is based purely on Cosine Similarity. The algorithm for the Content-Based Recommendation system is given in Figure 4. First, the songs are vectorized using Python's audio2numpy library. Since the length of each song vector 9 depends on the length of the song, the lengths of all the vectors are different. To make all the vectors of equal length, Principal Component Analysis is performed on all song vectors to convert them to equal sized vectors of length 30. Finally, when a song is queried, the cosine similarity of that song's vector is compared to the all the other vectors and 5 songs with the highest cosine similarity are returned as the result.

The application would need access to the camera of the mobile phone if the user wants to click a new selfie. Similarly, it would require access to the user's files/images if the user decides to upload an old picture. However, if the user decides to enter a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

text, then the user need not provide any permissions to the application. The Facial Emotion is analyzed through a Flask API; the image is sent to the API where a pretrained model predicts the user's emotion. Similarly, the text entered by the user is also sent to a Flask API where the text goes through a few preprocessing steps and then a pretrained model predicts the user's emotion.

After receiving the user's emotion, the application would randomly pick songs from the music dataset created for that emotion. There are a total of 150 songs in the dataset (50 songs for each emotion). A random playlist of 20 songs is generated for Mood-Based Song Recommendation. Mood-Based recommendation happens as follows: if the user is angry, then the user is suggested with slow melodious songs to bring their mind to peace. Similarly, if the user is sad, then the user is suggested songs which would cheer them up. On the other hand, if the user is happy, then they'll be suggested fast upbeat songs which the user would enjoy when they're happy. The 150 songs combined are used for Content-Based Recommendation.

4.1 User Emotion Classification Methods

The music player application uses the facial image of the user or the text entered by the user to classify the user's emotion.

4.1.1 Facial Expression Recognition

The application can classify the user emotion into three categories: angry, happy, and sad. For Facial Emotion Recognition, a model was trained on the "Natural Human Faces Images" [18] dataset taken from Kaggle. The dataset contains about 5500 grayscale images of shape 224 X 224. These images are divided into 8 emotions, namely angry, surprised, happy, contempt, fear, neutral, sad, and disgusted. From these 8 emotions, the three emotions required for the application (angry, happy, and sad) were extracted.

These extracted images were used to train a Convolutional Neural Network consisting of 16 layers for 50 epochs.

4.1.2 Text Sentiment Analysis

For Facial Emotion Recognition, a model was trained on the "Emotion Dataset for Emotion Recognition Tasks" [21] dataset taken from Kaggle. The dataset contains preprocessed English Twitter messages. These messages are divided into 6 emotions, namely angry, surprised, happy, disgusted sad, and trust. From these 6 emotions, the three emotions required for the application (angry, happy, and sad) were extracted.

These extracted texts were then stored into Python lists and split into train and test and each of these sentences was tokenized and then truncated/padded to 100 words. Then, these preprocessed sentences were used to train an Artificial Neural Network consisting of 4 layers for 30 epochs.

4.2 Song Matching

After the emotion of the user is determined, the application suggests mood-based songs to the user. A total of 150 songs are currently in the dataset. These 150 songs are divided into three classes (angry songs, happy songs, and sad songs); each class consists of 50 songs.

Once the user's emotion is retrieved from the API, the app selects 20 songs at random from the dataset of that emotion and populates a playlist which is shown at the frontend to the user. The songs played through the application have been uploaded to a Google Drive folder and the application accesses these songs through a json file which is stored in within the application with other application files.

4.3 Content-Based Recommendations

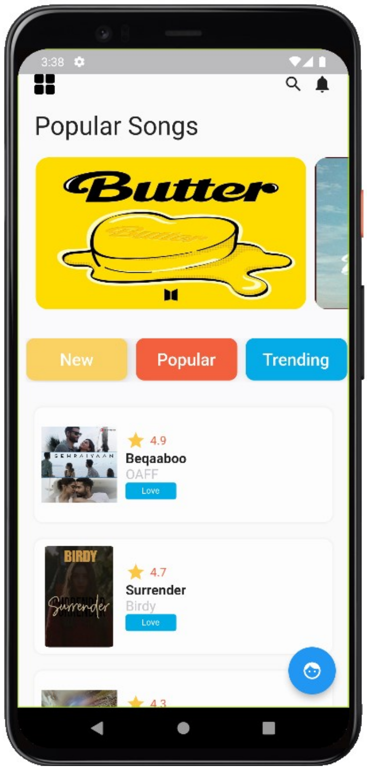
The Content-Based Recommendation System of the application makes use of all 150 songs while recommending songs. The application uses cosine similarity to analyze the similarity between songs. The songs are first converted to a vector and each song vector passes through Principal Component Analysis (PCA).

When a query id is sent to the Content-Based Recommendation System, the cosine similarity of the remaining 149 vectors is computed with the vector of the queried song and then the 5 songs with top cosine similarity are recommended to the user at the frontend.

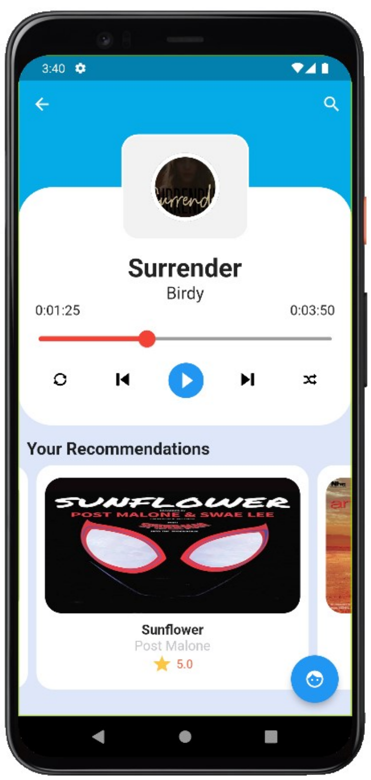
Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

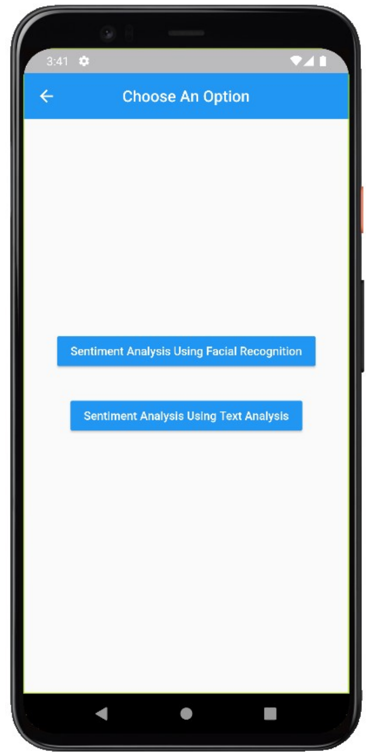
5. RESULTS



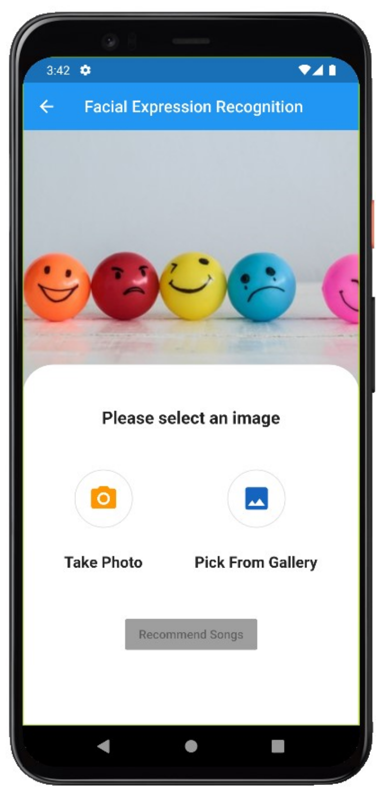
(a)



(b)

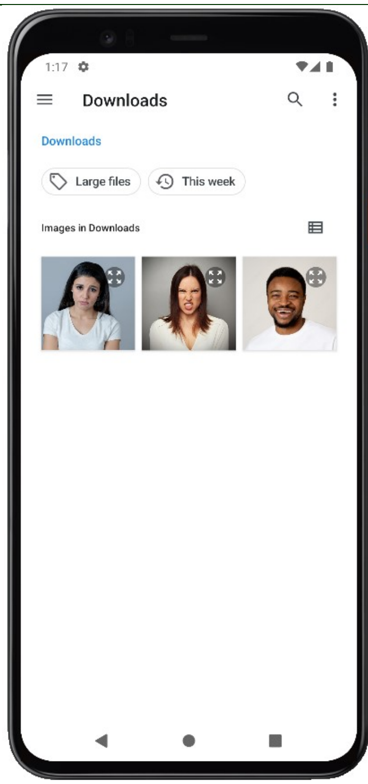


(c)

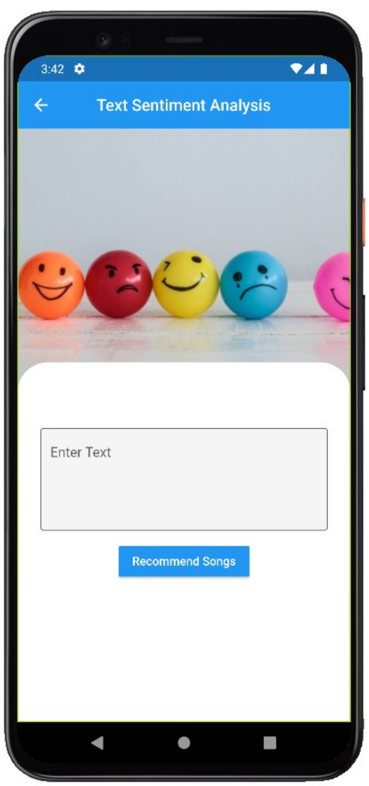


(d)

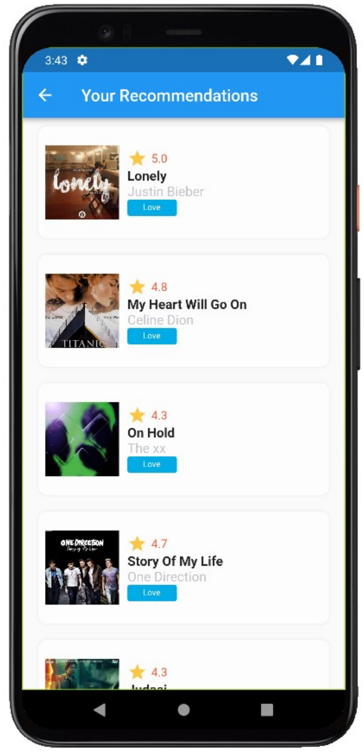
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(e)



(f)



(g)

Figure. 5 Application UI: (a) Home Page (b) Music Player Page (with Content-Based Recommendations) (c) Sentiment Analysis Choice Page (d) Facial Expression Detection Page (e) Choosing an Image from Gallery for Facial Expression Detection (f) Text Sentiment Analysis Page (g) Mood-Based Song Recommendations Page

The UI of the complete application is shown in Figure 5. When the user first opens the application, they land on the home page (Figure. 5a) where the user can choose to play a song from the New, Popular, and Trending sections. As soon as the user plays a song, they are redirected to the music player (Figure. 5b) where they can listen to the song of their choice. Along with the music player, the user is given song recommendation based on the song they are currently listening to (Content-Based Recommendations). If the user chooses to listen to music based on their emotion, they tap on the floating face button that redirects them to the Sentiment Analysis Choice Page (Figure. 5c). Depending upon their choice, their sentiment is analysed either using the Facial Expression Detector (Figure. 5d and 5e) or the Text Sentiment Analyzer (Figure. 5f) which finally leads them to the Mood-Based Song Recommendations Page (Figure. 5g).

The Facial Expression Recognition Model achieved a high training accuracy of 98.13% and an even better validation accuracy of 99.29%. Table 2 shows the training and validation metrics of the Facial Expression Recognition Model at every 10 epochs and Figure 6 shows the training and validation accuracy progression of the Facial Expression Recognition Model.

Table 2. Training and Validation Metrics of the Facial Expression Recognition Model

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	5.23	43.97%	0.99	53.19%
11	0.50	79.62%	0.42	83.69%
21	0.17	94.40%	0.0951	96.45%
31	0.08	97.61%	0.03	98.58%
41	0.08	97.28%	0.02	99.29%
50	0.05	98.13%	0.01	100%

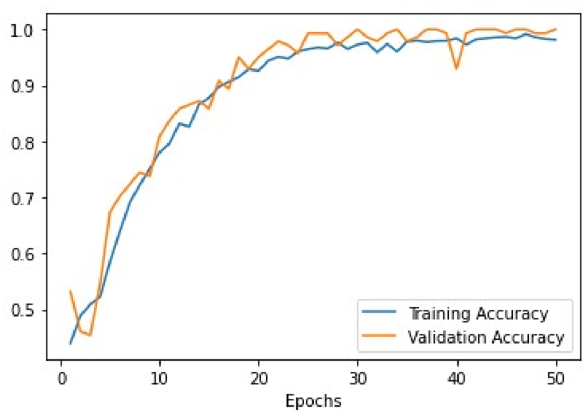


Figure. 6 Accuracy Graph of the Facial Expression Recognition Model

The Text Sentiment Analysis Model achieved a high training accuracy of 99.93% and a validation accuracy of 92.71%. Table 3 shows the training and validation metrics of the Text Sentiment Analysis Model at every 6 epochs and Figure 7 shows the training and validation accuracy progression of the Text Sentiment Analysis Model.

Table 3. Training and Validation Metrics of the Text Sentiment Analysis Model

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	1.04	44.15 %	1.02	44.81%
7	0.16	96.97%	0.22	93.42%
13	0.03	99.31%	0.16	94.58%
19	0.01	99.81%	0.18	94.13%
25	0.01	99.85%	0.23	93.42%
30	0.01	99.93%	0.28	92.71%

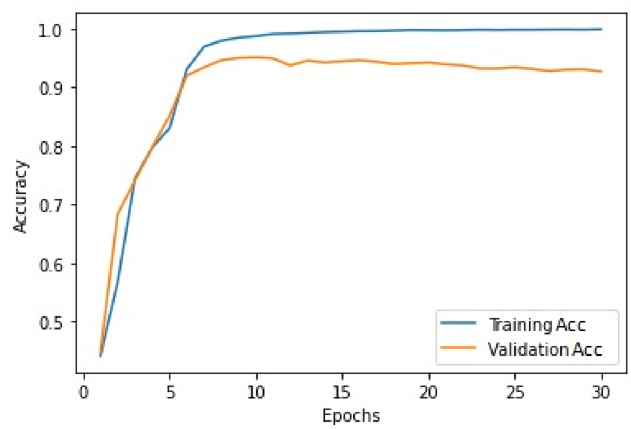


Figure. 7 Accuracy Graph of the Text Sentiment Analysis Model

6. CONCLUSION AND FUTURE SCOPE

This work proposes a music application with mood-based song recommendations. The application recommends songs to users based on their current emotion or mood. To determine the user's emotion, we use Deep Learning techniques, namely Facial Emotion Recognition and Text Sentiment Analysis. Once the user provides a new selfie, an old picture of their face, or a text, we use an Application Program Interface built using Flask web framework to extract the user's emotion and return it back to the application. Then, based on the user's emotion, the application chooses songs appropriate for the user's emotion from the music dataset and populates a playlist for the user.

The application is still elementary and it is missing many user-friendly features such as forward and rewinding a song, creating and adding songs to a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

custom playlist, and queueing songs to play next. Incorporating these features into the application is a work for the future. There are many better and much more efficient techniques for Content-Based Recommendation which can be incorporated into the application since Cosine Similarity, the similarity metric used in this application is not an appropriate option for long-term and large-scale use. Similarly, better Facial Expression Recognition and Text Sentiment Analysis techniques must be explored to use in the application. Apart from this, better methods to enhance the user's music listening experience must be explored.

REFERENCES:

[1] Healthcare Department, "Healthcare Department news," 2 February 2018. [Online].

[2] M. S. K. Sashank, Vijay S. M., Vikas B., and Y. Radhika, "Efficient deep learning based data augmentation techniques for enhanced learning on inadequate medical imaging data", ACTA IMEKO, April 2022

[3] Wendy E. J. Knight Nikki S. Rickard, Ph.D., "Relaxing Music Prevents Stress-Induced Increases in Subjective Anxiety, Systolic Blood Pressure, and Heart Rate in Healthy Males and Females," Journal of Music Therapy, December 2001.

[4] A. C. H. North and J. David. Musical preferences during and after relaxation and exercise. American Journal of Psychology, 113:43–67, 2000.

[5] A. C. H. North and D. J. Hargreaves. Responses to music in aerobic exercise and yogic relaxation classes. British Journal of Psychology, 87:535–547, 1996.

[6] J. Healey, R. W. Picard, and F. Dabek. A new affect perceiving interface and its application to personalized music selection.

[7] R. W. Picard. Affective Computing. MIT Press, Cambridge, MA, 1997.

[8] R. Thayer, "Music Mood Classification," 2014.

[9] Yading Song, Simon Dixon, Marcus Pearce, "Evaluation Of Musical Features For Emotion Classification", University of London, ISMR, 2012

[10] Cyril Laurier and Perfecto Herrera, Mood Cloud : A Real-Time Music Mood Visualization Tool, Music Technology Group Universitat Pompeu Fabra Ocata 1, 08003 Barcelona, Spain.

[11] M. Liljedahl, C. Sjomark, and N. Lefford. Using music to ¨ promote physical well-being via computer-mediated interaction.

[12] J. Chung and G. Scott Vercoe. The affective remixer: Personalized music arranging. In CHI 2006: Conference on Human Factors in Computing Systems, pages 393–398.

[13] N. Oliver and L. Kregor-Stickles. PAPA: Physiology and purpose-aware automatic playlist generation. In ISMIR 2006: 7th International Conference on Music Information Retrieval, pages 250–253, Victoria, Canada, October 2006.

[14] A.C. North and D.J. Hargreaves, The Social and Applied Psychology of Music. Oxford Univ. Press, 2008.

[15] G.G. Cassidy and R.A.R. MacDonald, "The Effects of Music Choice on Task Performance: A Study of the Impact of Self Selected and Experimenter-Selected Music on Driving Game Performance and Experience,", 2009.

[16] P.N. Juslin and J. Sloboda, Handbook of Music and Emotion: Theory, Research, Applications. Oxford Univ. Press, 2010.

[17] Wendy E. J. Knight Nikki S. Rickard, Ph.D., "Relaxing Music Prevents Stress-Induced Increases in Subjective Anxiety, Systolic Blood Pressure, and Heart Rate in Healthy Males and Females," 2001.

[18] Chankuptarat, Krittrin, Raphatsak Sriwatanaworachai, and Supannada Chotipant. "Emotion-Based Music Player." In *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, IEEE, 2019.

[19] Vijay S. M., K.S. Shirish, and M. V. S. Ramprasad, "A machine learning based sensing and measurement framework for timing of volcanic eruption and categorization of seismic data", ACTA IMEKO, April 2022

[20] Vulli, Adarsh, Parvathaneni N. Srinivasu, Madipally S.K. Sashank, Jana Shafi, Jaeyoung Choi, and Muhammad F. Ijaz. 2022. "Fine-Tuned DenseNet-169 for Breast Cancer Metastasis Prediction Using FastAI and 1-Cycle Policy" *Sensors* 22, no. 8: 2988. https://doi.org/10.3390/s22082988

[21] https://www.kaggle.com/datasets/sudarshanvaidya/random-images-for-face-emotion-recognition

[22] https://www.kaggle.com/datasets/parulpandey/emotion-dataset