www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 13 Issue 10 October 2024, Page No. 26560-26575
ISSN: 2319-7242 DOI: 10.18535/ijecs/v13i10.4920

A Personalized Multimodal Approach to Music Recommendation for Music Therapy

Shreyam Mandal [1], Aryan Shrivastava [2], Sidharth Nagarajan [3], Vinay Gulabani [4], Ayush Chauhan [5], Kritika Verma [6]

[1] Good Shepherd School, Bagdogra, Siliguri, India

[2] Delhi Public School, Doha, Qatar

[3] Delhi Private School, Sharjah, UAE

[4] Delhi Private School, Sharjah, UAE

[5] Department of Computer Science, Stanford University, California, United States of America

[6] Department of Computer Science, Syracuse University, New York, United States of America

Abstract
Mental health disorders and suicidality are rising among adolescents and young adults (A-YA) while rates of treatment engagement continue to remain low. Though the effectiveness of music therapy to enhance mental health is yet unknown for A-YA, the power of music to evoke emotions, facilitate communication, and promote relaxation suggests its promise. In this research study we discuss a multimodal music recommendation system which is capable of taking inputs in the form of a BFI-10 questionnaire, a series of self-view questions in the form of a survey, and an image of the user's face in order to determine music preferences based on existing correlations between personality traits and taste in music, as an effort to make advancements in the field of music therapy.

Keywords: music recommendation, convolutional neural networks, personality-based recommendation, facial expression analysis, music therapy

1. Introduction

UNICEF estimates that approximately 1 in every 7 adolescents experience mental disorders, amounting to a total of 166 million individuals [1]. Moreover, researchers have concluded that mental health illnesses in high schoolers and college students are becoming more complex and serious day by day [2, 3]. Mental illnesses have also been shown to have negative effects on adult functioning in areas such as social relationships and participation in the workforce [4]. Hence, a need has arisen for effective solutions to the rising issue of mental health among A-YA.

Recent studies have revealed that music can be used as a primary tool to diagnose mental health to observe visible improvements [5]. It helps adolescents to release inner stress and pain and express happiness, thereby serving as a means of decompression [6]. Music has also been shown to be able to mold emotions and influence actions in a positive way [7, 8]. Music Therapy can be considered as a pocket-friendly and accessible way to treat mental illnesses, especially in adolescents and young adults [9]. Music is already accessible to most youngsters on their smartphones, laptops and computers. In recent

years, recommendation systems have often been used as a means to present users with music appropriate to their tastes. These are essentially machine learning systems that use big data to suggest or recommend products to consumers. Research has also been conducted on methods to optimize such systems. The most commonly used approaches in such systems include Metadata Information Retrieval (Demographic-based Model), Signal-Based Music Information Retrieval, Emotion-Based models and Context-Based Models [10, 11].

There may be many external indicators to one's tastes in music like personality, self-views and facial expressions which can help us identify their preferences [12,13]. While all these factors have been studied extensively, little research has been conducted on promptly combining them and creating a hybrid model. Facial expressions, emotions, and a person's personality have also been proved to be interrelated. Therefore, we sought to create a music recommender system employing multimodal inputs. The study's primary hypothesis was that we can utilize existing relations between various factors and map them to their correspondence with music preferences in order to develop a robust recommendation system customized for helping adolescents and young adults who suffer from mental health issues.

2. Methods

2.1 Datasets

The Affect Net Dataset prepared by Mollhosseini, Hasani and Mahoor was used by the authors for the purpose of emotion recognition [14]. An open-access version of the database uploaded by Noam Segal on Kaggle was obtained. The dataset contains half a million images labeled on 8 basic emotions: Sad, Happy, Angry, Neutral, Fear, Surprise, Disgust and Contempt with a resolution of 96x96 pixels. It also shows the percentage of the expressed emotion and quantifies the arousal. This dataset was used to link facial expressions with emotion and arousal, which can be further used to obtain the user's music preferences [15].

The SCUT-FBP550 V2 Dataset uploaded to Kaggle by Pranav Chandane was obtained by the authors [16]. This dataset was originally prepared as the premier benchmark dataset for multi-paradigm facial beauty detection. The dataset contains 5500 350 x 350 RGB images of male and female faces with diverse properties (male/female, Asian/Caucasian, ages) and diverse labels (face landmarks, beauty scores within [1-3], beauty score distribution) [17]. This dataset was used in order to predict the user's music preferences since attractiveness and music tastes have been shown to be directly interrelated.

The "Top 10000 Songs on Spotify" Dataset prepared by Joakin Arvidsson was obtained from Kaggle as the main music database for the model [18]. This dataset was curated based on rankings from both the ARIA (Australian Recording Industry Association) and Billboard charts from 1960-2023 in order to ensure a diverse representation of songs that have gained popularity across the world. It contains the following features for 10000 songs arranged in the form of a .csv file: Track URI, Track Name, Artist URI(s), Artist Name(s), Album URI, Album Name, Album Artist URI(s), Album Artist Name(s), Album Release Date, Album Image URI, Disc Number, Track Number, Track Duration(ms), Track Preview URL, Explicit, Popularity, ISRC, Added By, Added At, Artist Genres, Danceability, Energy, Key, Loudness, Mode, Speech iness, Acousticness, Instrumentals, Liveness, Valence, Tempo, Time Signature, Album Genre(s), Label and Copyrights.

Table 1. Overview of all datasets used in this research.

Dataset	Type of Data	Description
Affect Net	Images and associated data	Images of human faces categorized under 8 emotions
SCUT-FBP5500 V2	Images and associated data	Images of people from diverse origins rated according to

		their attractiveness
Top 10000 Spotify Songs	Music metadata	Detailed information about the top songs on Spotify since 1960

2.2 Data Pre-processing

The following Python libraries were used in this procedure: (1) Pandas (For Analysis of Tables in Python), (2) NumPy (For Multi-dimensional Array Operations), (3) Matplotlib (For Visualization of Data), (4) PIL (For loading images unzipped from datasets), (5) PyTorch (For Converting Images into Tensors). Firstly, The images from the AffectNet Dataset and SCUT-FBP5500 Dataset were loaded and resized to 224 x 224 pixels in order to set a standard image resolution for the model and strike a balance between preserving the details and computational efficiency. The images in the dataset were then divided into two parts: training part (80% of the data) and testing part (20% of the data). The training part is used to train/validate the classifier, and the testing part is used to test the performance of the classifier. The 10-fold cross-validation adopted in the current model employs further splitting of the training part into subsets.[19] The images thus obtained were then converted into tensors of shape (3, 224, 224), containing pixel values scaled to [0.0, 1.0]. The main goal of using these two datasets was to train a Convolutional Neural Network (CNN) model to extract highly advanced and high dimensional features from these images. This training process involved passing the images through the CNN and extracting the feature vectors from the model's output part. This step is further elaborated later in this section. Subsequently, the 'Top 10000 Songs on Spotify' file was read using the panda's library and converted to a dataframe. The 'Album Release Date', which was mentioned in the original dataset in the form of 'MM-DD-YYYY', was deduced to the Release Year for convenience. The following features were also removed since they were not correlated with the essential features of a music track: 'Artist URI(s)', 'Artist Name(s)', 'Album URI', 'Album Name', 'Album Artist URI(s)', 'Album Artist Name(s)', 'Album Image URL', 'Track Preview URL', 'ISRC', 'Added By', 'Added At', 'Album Genres', 'Label', 'Copyrights', 'Disc Number', 'Track Number'. Following this, the authors observed that the dataframe contained 551 instances of songs which had one or more crucial values set to null. Therefore, these 551 tracks were removed from the dataframe entirely. The resulting data frame obtained was used for the recommendation model.

For the ease of further analysis, all 9449 music tracks were classified into 4 distinct genres, Reflective and complex, intense and rebellious, upbeat and conventional, and energetic and rhythmic, as per previous studies conducted by Rentfrow and Gosling on this topic [20]. They have primarily defined these classes according to the genres of music they encompass. Table 1 gives a brief outline of the various genres falling under each category of music. Therefore, the genres under 'Artist Genres' in our dataframe were used to detect the category of music which each sound track may fall under. The authors decided to initialize four variables for each soundtrack, 'Reflective', 'Intense', 'Upbeat' and 'Energetic'. Each of these variables stored a binary value indicating whether the artist produces music of that category or not. In order to ensure that all possibilities get explored, the authors allowed each soundtrack to fall under multiple categories at the same time.The resulting values were added to the dataframe.

Table 2. Genres under each category of music.

Category of Music	**Genres**
Reflective and Complex	Blues, Jazz, Classical, Folk
Intense and Rebellious	Rock, Alternative, Heavy Metal
Upbeat and Conventional	Country, Religious, Pop
Energetic and Rhythmic	Electronica, Funk, Rap, Hip Hop

2.3 Determining User's Preferences

Libraries Used:

- NumPy (For Multi-dimensional Array Operations)
- Pandas (For Analysis of Tables in Python)
- PyTorch (For Building Effective CNNs)
- Scikit Learn (For Predictive Data Analysis)
- Web browser (For Accessing Data using URLs)

2.3.1 Through Personality Analysis

Research conducted in the past years suggests that a person's personality is directly related to his/her music preferences [23, 24, 25]. Rentfrow and Gosling (2003) examined the relationship between these categories and the personalities of the listeners. They linked the categories to the listener's traits classified under three different subtopics: personality, self-views, and cognitive ability [12]. The Big Five Personalities seemed to be the major deciding personality factors, viz. extraversion, agreeableness, conscientiousness, neuroticism and openness [24]. Political views and verbal abilities were also deemed equally important. Based on these findings, the authors decided to include a Big Five Personality Questionnaire as a part of the procedure to find out the user's music preferences. It was concluded that John and Srivastava's 44-item Big Five Inventory (BFI) was too long to be a part of the model since filling it out would take 4 minutes or more [24]. Therefore, the authors considered including shorter versions of the BFI like BFI-20 [25], and BFI-10 [26]. Ultimately, BFI-10 designed by Rammstedt and John was selected since it can provide an adequate assessment of the user's personality in less than one minute [26]. A module was implemented to ask the following 10 questions one by one and take a number 1-10 as an input and store it in a list.

I see myself as someone who:

1. Is reserved
2. Is generally trusting
3. Tends to be lazy
4. Is relaxed, handles stress well
5. Has few artistic interests
6. Is outgoing, sociable
7. Tends to find faults in others
8. Does a thorough job
9. Gets nervous easily
10. Has an active imagination

The values for Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N) and Openness (O) were then calculated using the relations provided by Rammstedt and John [26].

$$E = avg(1R,6)$$
$$A = avg(2,7R)$$
$$C = avg(3R,8)$$
$$N = avg(4R,9)$$
$$O = avg(5R,10)$$

1, 2, 3, … 10 represent the inputs for questions 1-10 respectively

R represents the reverse scoring of the number

The authors also decided to consider certain self-views of the user in order to determine the user's music preferences. Rentfrow and Gosling (2003) found out that such self-perceived qualities have significant correlation with their music preferences [12]. Therefore, the authors included a feature to ask the user the following questions about their perception about themselves and take the necessary inputs on a scale of 1-10.

How much do the following adjectives suit you?

1. Politically liberal
2. Politically conservative
3. Physically attractive
4. Wealthy
5. Athletic
6. Intelligent

Consequently, the inputs for each of the six questions were taken in and labeled as PL, PC, Atr, W, Ath, I for convenience. In their paper, rent frow and Gosling provided a direct correlation between the traits (E, A, N, O, C, PL, PC, Atr., W, Athl., I) and the Categories of Music (Reflective and Complex, Intense and Rebellious, Upbeat and Conventional, and Energetic and Rhythmic). People having high scores of openness, intelligence, and political liberalism, and low scores of athleticism and political conservatism tend to prefer Reflective and Complex music. Those who are more open, athletic and consider

themselves intelligent tend to prefer Intense and Rebellious music. Upbeat and Conventional music was linked with people who are more extraverted, agreeable, athletic, conservative and attractive and were less open and liberal. Energetic and Rhythmic songs were preferred by those who were more extraverted, agreeable, liberal, attractive, athletic and less conservative. [12]

Table 3. Linking personality factors with music preferences.

Category of Music	Positively Affecting Factors	Negatively Affecting Factors
Reflective and Complex	Openness, Self-perceived Intelligence, Political liberalism	Athleticism, Political Conservatism
Intense and Rebellious	Openness, Athleticism, Self-perceived intelligence	None
Upbeat and Conventional	Extrovertedness, Agreeableness, Athleticism, Political conservatism, Attractiveness (Self-perceived)	Openness, Political Liberalism
Energetic and Rhythmic	Extrovertedness, Agreeableness, Political Liberalism, Attractiveness (Self-perceived), Athleticism	Political Conservatism

Therefore, values indicating the user's preference for each category of music were calculated as per the following equations.

$$Reflective = avg(O, I, L, 1 - A, 1 - PL)$$
$$Intense = avg(O, Athl, I)$$
$$Upbeat = avg(E, A, PC, Atr, Athl, 1 - O, 1 - PL)$$
$$Energetic = avg(E, A, PL, Atr, Athl, 1 - PC)$$

These values were then added to our dataframe for further usage during the recommendation of music.

2.3.2 Through Facial Recognition

Multiple traits of the user can be deduced using facial recognition which can guide the model towards his/her music preferences. Prior research has been conducted in this matter by Tian, Alaei and Rule [13], which suggests that the variables which have the highest weights in determining a person's music preferences are gender, attractiveness, and energeticness. To avoid biases, the authors decided that gender should be taken as an input from the user, while attractiveness and energeticness can be calculated with the help of CNNs.

Studies have shown that individuals who are more attractive tend to lean towards Energetic and Rhythmic music. People who are more energetic tend to prefer listening to music that can match their mood, like Energetic and Rhythmic music, and Upbeat and Conventional Music. Neat-looking people prefer Reflective and Complex music, while messiness is a clear indication that the person likes Intense and Rebellious, and Energetic and Rhythmic music [13]. Therefore, the values for attractiveness and energeticness obtained from the CNNs were used to calculate the user's likely preferences for music in each of the four categories. The mean of the user's preferences for each category obtained through facial recognition and personality tests were then calculated and stored as a quantitative measure of the user's perceived liking towards the category of music. Two CNN models were defined for feature extraction: energy model: For extracting features related to facial expressions indicating energy. Attractiveness Model: To extract features related to perceived attractiveness. Both models are simple feedforward neural networks with one fully connected layer.

Figure 1: Sample of Images used to train the Convolutional Neural Network.

2.3.2.1 Attractiveness Model

The SCUT-V2 Dataset contains pairs of images and their attractiveness score as rated by experts. Thus, this dataset was chosen by the authors in order to train the CNN. A 6-layer CNN was decided to be used for this task, with 3 Convolutional Layers, and 3 Linear layers. The ReLU activation function was used to introduce non-linearity in the model which would contribute in helping mitigate the vanishing gradient problem which may emerge since our model involves processing complex relationships between high-dimensional data. PyTorch's 2-Dimensional Max Pooling function was applied after every step to downsample the data. As proven by many researchers, Convolutional layers are efficient in extracting both low-level and high-level features from images, and thus can be used for extraction of facial features [27] which can help us in mapping the images to an attractiveness score. Therefore, the authors employed three Convolutional layers. The input tensor contains 3 channels (Each for red, green, and blue). As the image passes through each layer, more and more channels get added, each representing new features extracted by the respective layer. Max pooling is then applied to reduce the height and width of the tensors and increase the number of channels. Therefore, the input tensor of size [3, 128, 128] is converted into a tensor of size [128, 16, 16].

Subsequently, 3 fully-connected layers are implemented. Before feeding into these layers, the multi-dimensional feature maps from the Convolutional Layers are flattened into a one-dimensional tensor, making it suitable for the dense layers. By connecting every neuron in one layer to every neuron in the next, fully-connected layers can learn complex, non-linear combinations of the features. These layers take the high-level features extracted by the Convolutional layers and use them to deduce the final output value. Dropout was used to prevent overfitting with a drop percentage of 50% [28]. The last fully-connected layer provides us the desired regression value which is the attractiveness score for the image. The model was then trained on the SCUT-V2 Dataset using the Adam optimizer and MSE Loss function for 30 epochs with a batch size of 32.

Table 4: Architecture of the Attractiveness Model

S/No	Layer Type	Layer Parameters					Activation Function	Pooling	Dropout
		Input Channels	Output Channels	Kernel Size	Stride	Padding			
1	Convolutional	3	32	3	1	1	ReLU	Max Pooling	-
2	Convolutional	32	64	3	1	1	ReLU	Max Pooling	-
3	Convolutional	64	128	3	1	1	ReLU	Max Pooling	-

4	Linear	128	512	-	-	-	ReLU	Max Pooling	0.5
5	Linear	512	128	-	-	-	ReLU	Max Pooling	0.5
6	Linear	128	1	-	-	-	-	-	-

2.3.2.2 Energeticness CNN

The concept of Action Units (AUs) was used in order to correlate the images with an 'energy score'. In Facial Action Encoding Systems, AUs are basically the fundamental actions of individual muscles or groups of muscles. AUs have been proven to be useful in facial expression analysis [29], and therefore, the authors attempted to use certain AUs in order to obtain energy scores of the data on which the model is being trained. The OpenFace API, developed by B. Amos, B. Ludwiczuk and M. Satyanarayanan was used to extract the AUs from the images [30]. The following High Energy AUs and Low Energy AUs were considered in order to determine the energy scores.

Table 5: AUs utilized in order to map images to an 'energy score'

Low Energy AUs	**High Energy AUs**
AU 1: Inner brow raiser	AU 6: Cheek Riser
AU 4: Brow lowerer	AU 7: Lid tightener
AU 15: Lip Corner Depressor	AU 9: Nose Wrinkler
AU 17: Chin Raiser	AU 10: Upper Lip Raiser
AU 45: Blink	AU 12: Lip Corner Puller
	AU 26: Jaw drop

The energy score was then calculated by subtracting the average of the low energy AUs from the average of the High energy AUs. To make the model less dependent on the OpenFace API at runtime, the authors decided to map the images in the dataset to their energy scores using a CNN. The AffectNet dataset was chosen to train this model. The various emotions expressed in the images ensure that the model learns to analyze different types of emotions and accurately deliver an energy score. The images were first compressed to 224x224 and then converted into PyTorch tensors so they can be processed by the CNNs. A CNN architecture similar to the Attractiveness model was used for the Energeticness Model, including three convolutional layers, and three fully connected linear layers were used with identical parameters. However, at the second fully connected layer, the sigmoid activation function was used instead of the ReLU function in order to obtain a regression value between 0-1. Thus, the energy score for the input image is obtained.

Following this, the authors utilized the regression coefficients found by researchers [13] in order to map the attractiveness score and energy score to the user's preference for each category of music. The age and gender of the user were also taken in as an input since they are shown to have heavy weightage in determining music preferences as per the research mentioned. The following formulae were used for obtaining the user's preference for each category based on what has been deduced from their appearances.

$$\text{Reflective} = 0.05 \times A - 0.14 \times E + 0.27 \times AGE + 0.20 \times G$$

$$\text{Intense} = 0.06 \times A + 0.08 \times E - 0.24 \times AGE - 0.34 \times G$$

$$\text{Upbeat} = 0.00 \times A + 0.23 \times E + 0.10 \times AGE + 0.56 \times G$$

$$\text{Energetic} = 0.06 \times A + 0.27 \times E - 0.23 \times AGE - 0.16 \times G$$

A: Attractiveness Score, E: Energy Score, G: Gender (encoded as 1 for female, 0 if not mentioned, and -1 for male)

The means of these values and the values obtained from personality-based inference for each

category of music were then obtained and stored for further use.

2.3.2.3 Pinpointing Music Preferences by User-Rated Songs

In this section, the user is presented with 5 randomly selected songs and is asked to rate it based on their personal opinion. The authors decided to include this feature as a confirmatory test to ensure that the resultant output is exactly relevant to the user's choices [31]. The information thus obtained plays a crucial role in recommending the perfect songs for the user.

Firstly, the data frame was classified on the basis of its categorization under Reflective and Complex, Intense and Rebellious, Upbeat and Conventional, and Energetic and Rhythmic as done earlier. From the resulting data frame, 3 random songs for the user's most preferred category of music (as deduced earlier) and 2 random songs for the user's second preferred category of music were selected using the in-built 'random' module. Subsequently, Python's in-built 'web browser' module was used to play the randomly selected soundtracks using the Track URIs of the respective song in the data frame. For each song, the user was asked to rate his/her experience on a scale of 1-10. Along with the user's rating, the duration of the song played by the user was also noted in order to quantify how much the user likes the song [32], and was stored in a new data frame, containing all the information about the randomly selected soundtracks. In order to compare the user's preferences between all 5 randomly-selected songs, 'preference coefficients' were calculated for each song. This coefficient was essentially the average of the normalized data points for the user ratings (R) and the duration of the song played (D) which were obtained after min-max normalization. The following equation may be interpreted as the formula for calculating the preference coefficients (P).

$$P = \frac{1}{2} \times \left(\frac{D - D_{min}}{D_{max} - D_{min}} + \frac{R - R_{min}}{R_{max} - R_{min}}\right)$$

These preference coefficients were then stored in our dataframe for further use during the recommendation of music based on the user's preferences.

2.4 Music Recommendation Based on User's Preferences

Libraries Used:

- Pandas (For Data Analysis and Manipulation)
- NumPy (For Multi-dimensional Array Operations)
- Scikit Learn (For Predictive Data Analysis)

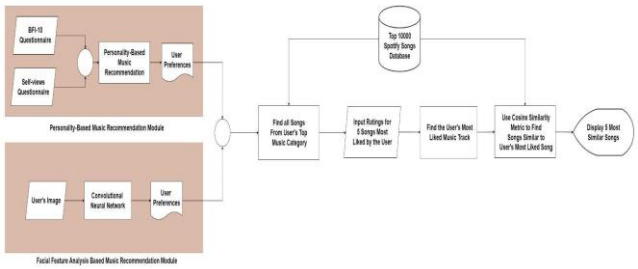
The resultant data frame obtained from the previous steps contained the metadata for 5 songs along with the preference coefficients. Thus, the authors attempted to explore various ways to recommend songs from the database that match the user's preferences. Content-based filtering, Collaborative filtering and Deep Learning Approaches were considered. In Content-based filtering, songs similar to the ones that the user has a greater preference for are recommended to the user. This method was considered viable, since the authors had enough metadata about each soundtrack (like valence, acousticness, danceability, etc.) [33, 34] In Collaborative filtering, the user is recommended songs which other users similar to him/her had a greater preference for. However, due to the lack of sufficient and reliable data about the preferences of other users based on their metadata (age, gender, location etc.), the authors rejected this approach. [35] The Deep Learning Approach involves using deep learning models such as neural networks, autoencoders, or transformers to learn complex patterns in the metadata of the user's preferred songs to come up with more accurate recommendations. However, due to the low number of user-rated songs, this approach was also considered unviable. [36]

Thus, the authors chose to use the Content-Based Filtering approach in order to recommend songs similar to the ones that the user has rated highly. Therefore, the authors started to explore which similarity metrics would be the perfect fit for the proposed model. Only 5 out of the 9449 songs in

our data frame have been rated by the user. Thus, the authors must work with very sparse data. Hence, similarity metrics which depend greatly on the magnitude of data like Euclidean Distance [37], Manhattan Distance [38] and Pearson's Correlations [39] were rejected. This leaves us with similarity metrics like Jaccard Similarity [40], Dice Coefficient [41], Hamming Distance [42], Rank Correlation [43], and Cosine Similarity [44]. Jaccard Similarity, Dice Coefficient, and Hamming Distance were rejected since these metrics do not work well with high-dimensional data. Rank correlation is also computationally quite expensive for large datasets. Overall, the cosine similarity metric seemed to be easier to interpret and computationally efficient for high-dimensional data [45]. Thus, it was chosen as the similarity metric for the recommendation model.

$$Cosine\ Similarity(A,B) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$$

The cosine similarity () function in the sklearn. metrics. pairwise module was used in order to analyze the music metadata. All non-number values are removed from the data frame containing the metadata for all rated and unrated songs and the data frame is passed into the mentioned function. The output matrix thus contained values for the pairwise cosine similarities between the values of each row in the input data frame. Subsequently, the data frame was arranged in descending order of its similarity to the user's most liked song (the song with the highest preference coefficient). Thereafter, upon the user's request, the model uses the Track URIs for the top recommended songs and plays them one by one.



Flowchart 1: The system architecture of the Music Recommender Model

2.4 Exceptions Encountered

The study conducted by Rentfrow and Gosling (2003) [12] had revealed direct connections between tempo, acousticness, and speechiness of a music with its categorization under the four music dimensions: Reflective and Complex, Intense and Rebellious, Upbeat and Conventional, and Energetic and Rhythmic. Therefore, the authors tried to develop a model to map these features with the categories of music. Since the data available was unlabelled, the authors attempted to use clustering algorithms to find required correlations. However, this attempt resulted in failure. K-Means Clustering, Agglomerative Clustering, DBSCAN were among the attempted algorithms. An autoencoder with K-Means Clustering was also tried. When the results were graphed, K-Means Clustering seemed to be forming the most effective clusters. However, when a few samples were loaded and examined, it became clear that K-Means Clustering had failed to form proper clusters. Upbeat, and energetic music was classified as rebellious, while certain rebellious songs were classified as upbeat. The only category which was correctly classified was the Reflective and Complex category. Various potential exceptions may arise in the genre preference prediction model, requiring robust error handling to ensure smooth execution and accurate debugging. When specified dataset paths or image files are missing, there is a common exception. Check for file existence before loading and provide a clear error message if files are missing. The model's forward pass can be affected by dimensions mismatches or issues with image processing. Therefore, the authors ensured that all images were resized to 224 x 224 pixels before being passed into the CNNs.

3. Results

In order to evaluate the Music Recommendation Model, a tool named Anvil was used to develop a Web Application that can act as a front-end to the model, i.e., inputs taken from the Anvil Web App would be sent to a server containing the functions and machine learning models necessary for the task. Fig. 2 illustrates the different forms

presented to the users on the web application which acted as the sole interface of interaction between the model and the user.

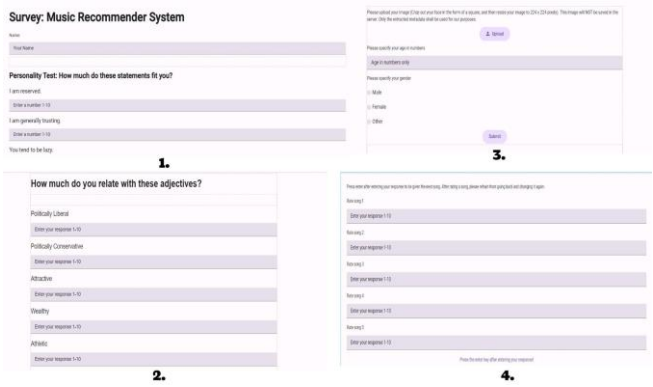
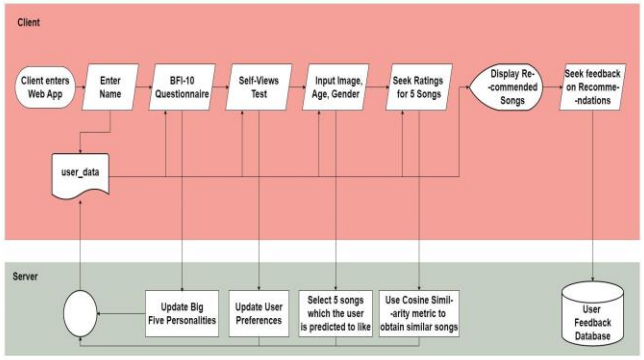


Figure 2: User Interface of the Survey Web App. 1. BFI-10 Questionnaire. 2. Self-views test. 3. Portrait Upload. 4. Rating of Songs

The user's name was used as the means of identification in the database, and therefore was sought at the very beginning. The user was presented with Rammstedt and John's BFI-10 [26], and then with a Self-views test to further aid personality-based music recommendation [12]. A 224x224 image of the user's face was then sought for facial feature-based music recommendation. After collecting data regarding the user's preferences of various types of music in the predicted category, the data was sent to the Anvil server where required computations were undertaken, and 5 songs most similar to the user's choice were returned to the client. The ratings were then sought and stored in the User Feedback Database. Flowchart 2 depicts the system architecture of the Anvil Web App.



Flowchart 2: System architecture of the Music Recommendation Web App designed for Receiving Feedback on Recommendations for the purpose of Evaluation of the Model.

Much of the computation was done on the Anvil Server in order to ensure smooth functioning of the client's device. The survey was conducted on 8 individuals. Thereafter, the User Feedback Database contained information for 40 music tracks (5 tracks recommended to each of the 8 individuals). The following table contains the precision and recall values for each of the 4 music categories from the data obtained in the Feedback Database.

Table 6: Precision and Recall Values for Different Music Categories

Category	Precision	Recall
Reflective and Complex	0.9	1.0
Intense and Rebellious	0.538	0.778
Upbeat and Conventional	0.8	0.727
Energetic and Rhythmic	0.8	0.727

A Confusion Matrix for the data was also calculated. Fig. 3 shows the obtained confusion matrix.

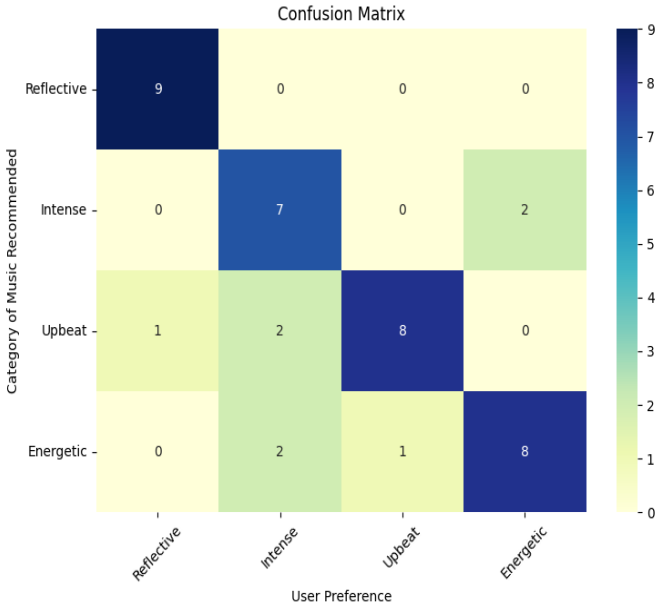


Fig. 3: Confusion Matrix plotted with Preferences of Users on the x-axis and Categories of Songs recommended to them on the y-axis

The hit rate (Accuracy) for the model was also calculated to be 0.8.

4. Discussion

The Music Recommendation Model developed had an accuracy of 0.8 which implies that it can be considered as an accurate system that can recommend music based on numerous inputs.If we take a look at the precision and recall metrics for each of the four music categories, it is evident that the authors' model was extremely effective in correctly recommending Reflective and Complex music. This can be attributed to the distinct features of this kind of music, viz. low tempo, high acousticness, and high valence scores [12]. Upbeat and Conventional, and Energetic and Rhythmic music tracks were recommended appropriately. However, Intense and Rebellious music had a particularly low precision and recall score. Few songs falling under this category were recommended to users who preferred Energetic and Rhythmic music and vice-versa. This may have been because of the high energy characteristic that both of these categories share. Some Upbeat and Conventional tracks were also recommended to those who preferred Intense and Rebellious music, which may be due to their shared characteristic of high tempo and high danceability.

Compared to existing methods of predicting music preferences, the authors' approach shows a significant change in improving accuracy and providing insights by using facial features and personality traits [46, 47]. However, it has limitations as well, such as the reliance of the model on frontal face images and potential appearance-based biases (genre, types, no. of years, names, lyrics etc.) in the datasets used. Future research could focus on expanding the dataset, incorporating 3-dimensional images, and utilizing more advanced machine learning models to enhance the model's accuracy and efficiency. A need for clearly distinguishing Intense and Rebellious music from other genres is also evident. In the future, attempts can be made to address this need by utilizing advanced clustering algorithms. Ensuring user privacy and data security is the first priority of any developer, and the authors highly regarded the importance of obtaining user consent and implementing data protection measures. User feedback from the beta version of the web application will be invaluable in refining and improving the model and user interface. Continuous improvement based on user interactions and feedback will help create a more accurate and user-friendly system

Healthcare, being on top, can use this system to deal with the therapy of the people suffering from depression, post-traumatic stress disorder (seen in army veterans), cancer patients undergoing rough treatments and individuals coping with daily stress and pressure. Sports and Athletics: The model can be used to uplift the mood and relax athletes before their games. Using the situational facial features, it can help provide with the spotify playlists that help in reducing the pre-game anxiety, boost their morale and try to bring them in the "zone". Workplace: Implementing this model in the Offices and Workplaces can help make the environment much more relaxed, calm, while boosting the energy of the employees and enhancing their productivity. This approach can help in managing work-related stress and increasing overall job satisfaction. The proposed Music Recommendation model contributes to the advancement of personalized music recommendation systems through multi-modal inputs. By continuing to work on this approach, we can improve the overall well-being, productivity and health across many domains.

5. Conclusion

This study aimed to build a robust recommender system that can utilize multimodal inputs. The inputs include answers from a BFI-10 Questionnaire, a Self-views test, and a facial image of the user. Parameters obtained from these inputs were linked with the user's music preferences based on pre-existing regression equations. The technique proved to be effective, for the model has achieved an accuracy of 0.8. More emphasis has to be laid on establishing clear-cut parameters for Intense and Rebellious music. According to the author's findings, the model misclassified this category of music the most. Future studies could explore incorporating

loudness as a factor in their models, since high loudness could be the distinction between this category and the others.

Future research could also focus on refining emotional analysis algorithms, diversifying the music database, conducting surveys on a larger scale across different age groups, nations, cultures, and traditions, and collaborating with therapists to better understand the conditions of depressed A-YA. Its effectiveness could also be improved through extensive training on advanced hardware. By leveraging advanced machine learning and music therapy techniques, the proposed system offers an innovative approach to mental health interventions, providing accessible and effective support for A-YA struggling with mental health issues. [48]

Supplementary Material

Not applicable

Author contributions

S.M. data curation, methodology, validation. V.G. methodology, conclusion, writing - original draft. S.N. writing - original draft, writing - review & editing, validation. A.S. writing - original draft, visualization, abstract. A.C., K.R. conceptualization, validation, writing - review & editing.

Competing financial interests

The authors declare no competing financial interests.

References

1. "Adolescent mental health statistics," UNICEF DATA. https://data.unicef.org/topic/child-health/mental-health/#_edn1
2. D. S. Pledge, R. T. Lapan, P. P. Heppner, D. Kivlighan, and H. J. Roehlke, "Stability and severity of presenting problems at a university counseling center: A 6-year analysis.," Professional Psychology Research and Practice, vol. 29, no. 4, pp. 386–389, Aug. 1998, doi: https://doi.org/10.1037/0735-7028.29.4.386.
3. S. A. Benton, J. M. Robertson, W.-C. Tseng, F. B. Newton, and S. L. Benton, "Changes in counseling center client problems across 13 years.," Professional Psychology Research and Practice, vol. 34, no. 1, pp. 66–72, Feb. 2003, doi: https://doi.org/10.1037/0735-7028.34.1.66.
4. D. Nopf, M. Park, and T. Mulye, "The Mental Health of Adolescents: A National Profile, 2008," 2008. Available: https://nahic.ucsf.edu/wp-content/uploads/2008/02/2008.MentalHealthBrief.pdf
5. L. Rebecchini, "Music, mental health, and immunity," Brain, Behavior, & Immunity - Health, vol. 18, no. 100374, p. 100374, Dec. 2021, doi: https://doi.org/10.1016/j.bbih.2021.100374.
6. J. Huang and X. Li, "Effects and Applications of Music Therapy on Psychological Health: A Review," Advances in Social Science, Education and Humanities Research, 2022, doi: https://doi.org/10.2991/assehr.k.220110.186.
7. S. Koelsch, "Music-evoked emotions: principles, brain correlates, and implications for therapy," Annals of the New York Academy of Sciences, vol. 1337, no. 1, pp. 193–201, Mar. 2015, doi: https://doi.org/10.1111/nyas.12684.
8. A. C. North, D. J. Hargreaves, and S. A. O'Neill, "The importance of music to adolescents," British Journal of Educational Psychology, vol. 70, no. 2, pp. 255–272, Jun. 2000, doi: https://doi.org/10.1348/000709900158083.
9. "Resource-Oriented Music Therapy in Mental Health Care," barcelonapublishers.com. https://barcelonapublishers.com/resource-oriented-music-therapy-mental-health-care(accessed Jun. 29, 2024).

10. D. Paul and S. Kundu, “A Survey of Music Recommendation Systems with a Proposed Music Recommendation System,” Advances in Intelligent Systems and Computing, pp. 279–285, Jul. 2019, doi: https://doi.org/10.1007/978-981-13-7403-6_26.
11. Y. Song, S. Dixon, and M. Pearce, “A Survey of Music Recommendation Systems and Future Perspectives.” Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e0080299afae01ad796060abcf602abff6024754
12. P. J. Rentfrow and S. D. Gosling, “The do re mis of everyday life: The structure and personality correlates of music preferences.,” Journal of Personality and Social Psychology, vol. 84, no. 6, pp. 1236–1256, 2003, doi: https://doi.org/10.1037/0022-3514.84.6.1236.
13. L. Tian, R. Alaei, and N. O. Rule, “Appearance Reveals Music Preferences,” Personality and Social Psychology Bulletin, vol. 48, no. 12, p. 014616722110482, Sep. 2021, doi: https://doi.org/10.1177/01461672211048291.
14. A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild,” IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18–31, Jan. 2019, doi: https://doi.org/10.1109/TAFFC.2017.2740923.
15. T. Schäfer and P. Sedlmeier, “What makes us like music? Determinants of music preference.,” Psychology of Aesthetics Creativity and the Arts, vol. 4, no. 4, pp. 223–234, Nov. 2010, doi: https://doi.org/10.1037/a0018374.
16. Pranav Chandane, “Facial Beauty Rating - SCUT-FBP5500 V2,” Kaggle.com, 2024. https://www.kaggle.com/datasets/pranavchandane/scut-fbp5500-v2-facial-beauty-scores (accessed Jun. 29, 2024).
17. L. Liang, L. Lin, L. Jin, D. Xie, and M. Li, “SCUT-FBP5500: A Diverse Benchmark Dataset for Multi-Paradigm Facial Beauty Prediction,” arXiv:1801.06345 [cs], Jan. 2018, Available: https://arxiv.org/abs/1801.06345
18. “Top 10000 Songs on Spotify 1960-Now,” www.kaggle.com.https://www.kaggle.com/datasets/joebeachcapital/top-10000-spotify-songs-1960-now/data
19. A. I. Siam, N. F. Soliman, A. D. Algarni, F. E. Abd El-Samie, and A. Sedik, “Deploying Machine Learning Techniques for Human Emotion Detection,” Computational Intelligence and Neuroscience, vol. 2022, p. e8032673, Feb. 2022, doi: https://doi.org/10.1155/2022/8032673.
20. A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: a database for facial expression, valence, and arousal computing in the wild,” IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18–31, Jan. 2019, doi: https://doi.org/10.1109/taffc.2017.2740923.
21. D. Rawlings and V. Ciancarelli, “Music Preference and the Five-Factor Model of the NEO Personality Inventory,” Psychology of Music, vol. 25, no. 2, pp. 120–132, Oct. 1997, doi: https://doi.org/10.1177/0305735697252003.
22. P. G. Dunn, B. de Ruyter, and D. G. Bouwhuis, “Toward a better understanding of the relation between music preference, listening behavior, and personality,” Psychology of Music, vol. 40, no. 4, pp. 411–428, Mar. 2011, doi: https://doi.org/10.1177/0305735610388897.
23. T. Schäfer, P. Sedlmeier, J. Krems, and H. Bruhn, “Determinants of Music

Preference (Bestimmungsgrößen für Musikpräferenz)" Dissertation, 2008. Accessed: Jun. 29, 2024. [Online]. Available: https://monarch.qucosa.de/api/qucosa%3A19075/attachment/ATT-0/

24. "John Srivastava 1995 big5," 1995. Available: http://jenni.uchicago.edu/econ-psych-traits/John_Srivastava_1995_big5.pdf
25. H. J. Pandya et al., "Label-free electrical sensing of bacteria in eye wash samples: A step towards point-of-care detection of pathogens in patients with infectious keratitis," Biosensors and Bioelectronics, vol. 91, pp. 32–39, May 2017, doi: https://doi.org/10.1016/j.bios.2016.12.035.
26. B. Rammstedt and O. P. John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," Journal of Research in Personality, vol. 41, no. 1, pp. 203–212, Feb. 2007, doi: https://doi.org/10.1016/j.jrp.2006.02.001.
27. GhavamiNejad P, GhavamiNejad A, Zheng H, Dhingra K, Samarikhalaj M, Poudineh M., "A Conductive Hydrogel Mi-croneedle-Based Assay Integrating PEDOT: PSS and Ag-Pt Nanoparticles for Real-Time, Enzyme-Less, and Electro-chemical Sensing of Glucose," Advanced Healthcare Materials, vol. 12, no. 1, Oct. 2022, doi: https://doi.org/10.1002/adhm.202202362.
28. M. Safavieh et al., "Paper microchip with a graphene-modified silver nano-composite electrode for electrical sensing of microbial pathogens," Nanoscale, vol. 9, no. 5, pp. 1852–1861, 2017, doi: https://doi.org/10.1039/c6nr06417e.
29. Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 97–115, 2001, doi: https://doi.org/10.1109/34.908962.
30. B. Amos, B. Ludwiczuk, M. Satyanarayanan,"Openface: A general-purpose face recognition library with mobile applications,"CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.
31. Pundlik, A.; Verma, S.; Dhingra, K. Neural Pathways Involved in Emotional Regulation and Emotional Intelligence. J. Knowl. Learn. Sci. Technol. 2024, 3 (3), 165-192. https://doi.org/10.60087/jklst.vol3.n3. p.165-192.
32. S. J. Philibotte, S. Spivack, N. H. Spilka, I. Passman, and P. Wallisch, "The Whole is Not Different From its Parts," vol. 40, no. 3, pp. 220–236, Feb. 2023, doi: https://doi.org/10.1525/mp.2023.40.3.220.
33. J. Bernard, T. Ruppert, M. Scherer, Jörn Kohlhammer, and T. Schreck, "Content-based layouts for exploratory metadata search in scientific research data," CiteSeer X (The Pennsylvania State University), Jan. 2012, doi: https://doi.org/10.1145/2232817.2232844.
34. R. Van Meteren and M. Van Someren, "Using Content-Based Filtering for Recommendation 1." Available: https://users.ics.forth.gr/~potamias/mlnia/paper_6.pdf
35. F. Cacheda, V. Carneiro, D. Fernández, and V. Formoso, "Comparison of collaborative filtering algorithms," ACM Transactions on the Web, vol. 5, no. 1, pp. 1–33, Feb. 2011, doi: https://doi.org/10.1145/1921591.1921593.
36. Mehta, A.; Alaiashy, O.; Kumar, P.; Tamilinian, V.; Besong, S.; Balpande, S.; Verma, S.; Dhingra, K. Advancing Model-Based Systems Engineering in Biomedical and Aerospace research: A

Comprehensive Review and Future Directions. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 133-147. https://doi.org/10.60087/jklst.v3.n4.p133.

37. Chilmakuri, L.; Mishra, A. K.; Shokeen, D..; Gupta, P..; Wadhwa, H. H.; Dhingra, K..; Verma, S. A Wearable EMG Sensor for Continuous Wrist Neuromuscular Activity for Monitoring. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 148-159. https://doi.org/10.60087/jklst.v3.n4.p148.
38. E. Kobayashi, Takayasu Fushimi, K. Saito, and T. Ikeda, "Similarity Search by Generating Pivots Based on Manhattan Distance," Lecture notes in computer science, pp. 435–446, Jan. 2014, doi: https://doi.org/10.1007/978-3-319-13560-1_35.
39. Chandna, R..; Bansal, A.; Kumar, A.; Hardia, S.; Daramola, O.; Sahu, A.; Verma, K.; Dhingra, K.; Verma, S. Skin Disease Classification Using Two Path Deep Transfer Learning Models. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 169-187. https://doi.org/10.60087/jklst.v3.n4.p169.
40. S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of Jaccard coefficient for keywords similarity," ResearchGate, Mar. 2013, [Online]. Available: https://www.researchgate.net/publication/317248581_Using_of_Jaccard_Coefficient_for_Keywords_Similarity
41. V. Thada and V. Jaglan, "Comparison of Jaccard, Dice, Cosine Similarity Coefficient to Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm." Available: https://dknmu.org/uploads/file/6842.pdf
42. M. Norouzi, D. J. Fleet, and R. R. Salakhutdinov, "Hamming Distance Metric Learning," Neural Information Processing Systems, 2012. https://proceedings.neurips.cc/paper/2012/hash/59b90e1005a220e2ebc542eb9d950b1e-Abstract.html (accessed Jun. 29, 2024).
43. C. Shao, P. Cui, P. Xun, Y. Peng, and X. Jiang, "Rank correlation between centrality metrics in complex networks: an empirical study," Open Physics, vol. 16, no. 1, pp. 1009–1023, Dec. 2018, doi: https://doi.org/10.1515/phys-2018-0122.
44. F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Semantic Cosine Similarity," www.semanticscholar.org, 2012. https://www.semanticscholar.org/paper/Semantic-Cosine-Similarity-Rahutomo-Kitasuka/41ff3934f40c32ac8643270822de1c763e16c71b
45. P. K. Singh, P. K. D. Pramanik, and P. Choudhury, "A Comparative Study of Different Similarity Metrics in Highly Sparse Rating Dataset," Data Management, Analytics and Innovation, pp. 45–60, Sep. 2018, doi: https://doi.org/10.1007/978-981-13-1274-8_4.
46. C.-C. Lu and V. S. Tseng, "A novel method for personalized music recommendation," Expert Systems with Applications, vol. 36, no. 6, pp. 10035–10044, Aug. 2009, doi: https://doi.org/10.1016/j.eswa.2009.01.074.
47. A. Nanopoulos, D. Rafailidis, P. Symeonidis, and Y. Manolopoulos, "MusicBox: Personalized Music Recommendation Based on Cubic Analysis of Social Tags," IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 2, pp. 407–412, Feb. 2010, doi: https://doi.org/10.1109/tasl.2009.2033973.
48. A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of

methods and applications," Psychological Medicine, vol. 49, no. 09, pp. 1426–1448, Feb. 2019, doi: https://doi.org/10.1017/s0033291719000151.