Mood Based Music Recommendation System

Ayushi Chaudhary, Anshu Sharma, Anubhav Kumar, Aditya Rajpoot, Abhishek, Aditya Singh
Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, India

Abstract- **Psychological research has demonstrated that music effectively relieves stress, elevates mood, and triggers the release of neurochemicals such as oxytocin, serotonin, and dopamine. Consequently, music is widely utilized in therapeutic settings and rehabilitation centers for various disorders. As global mental health issues escalate, it is increasingly critical to address these concerns with innovative solutions. Existing music recommendation systems are inadequate in holistically addressing user needs, with few offerings mood-based recommendations and even fewer integrating therapeutic benefits tailored to specific situations. Given the widespread reliance on music for enhancing cognition, memory, and sleep quality, while reducing anxiety, pain, and blood pressure, there is a pressing need for a comprehensive product that maximizes these benefits. Our proposed system introduces a mood-based music player that performs real-time mood detection and recommends songs accordingly, enhancing traditional music player apps with this additional feature. This system aims to analyze user images, predict emotional expressions, and suggest suitable music, thereby improving user satisfaction and overall mental health.**

Keywords: Face Recognition, Image Processing, Computer Vision, Emotion Detection, Music, Mood Detection, Content-Based Filtering, Mental Health, Machine Learning.

I. INTRODUCTION

I.A BACKGROUND AND SIGNIFICANCE

Emotions are the bodily feelings that reflect our mood, temperament, personality, and character. In 1972, Paul Ekman developed a classification system for basic emotions, identifying anger, disgust, fear, happiness, sadness, and surprise as the core expressions.

Facial expressions, created by the movements or positions of facial muscles, play a crucial role in conveying these emotions. These expressions can be both voluntary and involuntary. Voluntary facial expressions occur when individuals consciously manipulate their facial muscles to convey a particular emotion. For example, a person can intentionally bring their eyebrows together and frown to show anger. Conversely, someone might deliberately relax their facial muscles to appear calm and unaffected.

However, facial expressions are predominantly involuntary, closely tied to an individual's emotional state. It is almost impossible for a person to completely suppress their facial expressions as emotions often leak out through micro-expressions, which are brief, involuntary facial expressions that occur as a quick reaction before a person can consciously control their facial muscles.

Facial expression analysis encompasses the detection and interpretation of facial movements and the recognition of emotions. This process involves three main approaches:

Face Acquisition: Capturing the facial image using a camera or sensor.

Facial Data Extraction and Representation: Identifying and representing key facial data points such as muscle movements.

Facial Expression Recognition: Comparing the extracted facial data to a database of known expressions to classify the emotion.

The "Emotion-Based Music Player" is an innovative device designed to detect a user's emotion through facial expressions and play music that aligns with their emotional state. The system works as follows:

Emotion Reflection: The user displays their emotion through their facial expression.

Emotion Detection: The device captures the facial expression and analyzes it.

Emotion Interpretation: The system interprets the detected emotion using a pre-trained model.

Music Playback: Based on the interpreted emotion, the device selects and plays a playlist of songs tailored to the user's current emotional state.

This system focuses exclusively on the analysis of facial expressions, without considering head or broader face movements. By accurately detecting and responding to users' emotions, the Emotion-Based Music Player aims to enhance the listening experience and provide a personalized music selection that resonates with the user's emotional needs.

I.B PROBLEM STATEMENT

Music's well-known power to influence our emotions is undeniable. People throughout history have sought solace and relaxation in music's melodies, regardless of era. Research even shows that rhythm itself can act as a powerful stress reliever. However, choosing the right song can be difficult, especially when it needs to match our current mood. Staring down a disorganized list of music can be discouraging, leading most users to simply shuffle their entire library or pick random songs. This often results in a mismatch between the music and the listener's emotions. For instance, someone feeling sad might crave the emotional release of heavy rock, but searching through a large playlist is impractical. This traditional search method, used for years, often leads to random selection or simply playing everything at once, which can become monotonous.

I.C OBJECTIVE AND SCOPE

This document dives into the concept of a real-time music recommender system powered by machine learning that analyzes facial expressions. We will explore the project from various angles, including user needs, the product's overall vision, and a high-level view of requirements and limitations. Additionally, we will delve deeper into the specific functionalities required, such as the user interface, core features, and performance expectations.

This SRS document serves as the foundation for software requirements throughout the project's lifespan. It captures the final, mutually agreed-upon requirements between customers and designers. By the project's completion, all functionalities should be traceable back to this document, ensuring the final product aligns with the initial specifications. The document outlines the system's features, performance expectations, limitations, user interface, and reliability considerations for the entire project duration.

II. METHODOLOGY

A. Data Collection

Data for music recommendation systems can be sourced from various channels, with the specific approach tailored to the system's design and objectives. Common data sources include listening history, likes and dislikes, playlist creation habits, user profiles, explicit ratings and feedback, rating systems, user reviews, and contextual data. Each source provides unique insights that contribute to understanding user preferences and behaviors.

Making Accurate Recommendations: The Core of the Project

As mentioned earlier, providing users with reliable recommendations is our top priority. To achieve this, evaluating the system's ability to deliver accurate suggestions is paramount. This section explores how we will assess the recommendation system's accuracy.

Splitting Data for Evaluation:

User data will be divided into two sets. The first 10% will serve as training data, where the accuracy of recommendations is already known. The remaining 90% will be used for evaluation. Based on the established accuracy from the training data, the system's performance on the larger dataset will be assessed.

Interconnected Modules: Database and Recommendation

The Database Module acts as the central storage unit for all the software's data, utilizing a highly interconnected graph database (neo4j). This module is primarily accessed and utilized by the Recommendation Module. Queries are sent from the Recommendation Module to the Database Module, which retrieves and returns results.

Focused User Interface (GUI)

The GUI is not designed for end users but rather for stakeholders to view the results generated by the recommendation system. Stakeholders will be able to see both the expected and actual outcomes through the GUI, allowing them to assess the system's performance.

B. Data Description

Song Data:

Title: The name of the song.

Artist: The artist or group that created the song.

Album: The album to which the song belongs.

Genre: The genre(s) associated with the song.

Duration: The length of the song in minutes and seconds.

Mood Tags:

Tags or labels indicating the mood or emotions associated with each song. Examples could include "happy," "sad," "rock-rock," "neutral," etc.

Intensity/Valence:
A numerical value representing the intensity or valence of the mood. For example, a scale from 1 to 10 where 1 is extremely sad, and 10 is extremely happy.

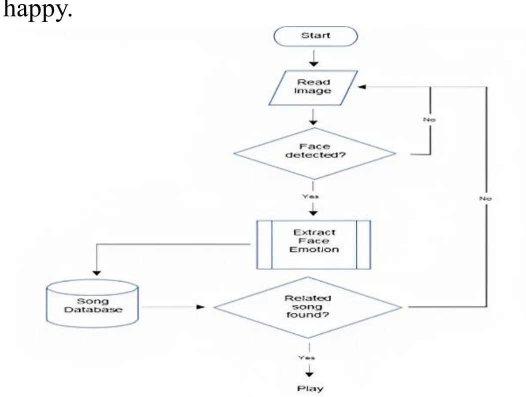


Fig: Data Flow Diagram

C. Data Processing
Processing data is a vital step to prepare it for analysis and to build an effective recommendation system. Key steps include data inspection to identify and understand the data structure, handling missing values to ensure completeness, formatting data for consistency, cleaning mood labels for accuracy, feature engineering to create meaningful attributes, processing user interaction data, encoding categorical data for machine learning models, scaling data for uniformity, and splitting the data for training and testing purposes.

Cleaning and Transformation
Cleaning and changing a mood-based music suggestion dataset is a pivotal step in planning the information for examination and building a proposal framework. Here is a common direct on the cleaning and change process:

Data Inspection:
Begin by stacking the dataset and reviewing the structure, columns, and information types.
Check for lost values, exceptions, and irregularities in the data.
Handling Lost Values:
Address any lost values in the dataset by either ascribing values or evacuating columns with lost data.
For disposition names, guarantee that each tune has a temperament tag assigned.

Data Formatting:
Standardize groups, such as guaranteeing reliable date groups, dealing with numerical values, and changing over information sorts where necessary.
Normalize content information, such as tune titles or craftsman names, to guarantee consistency.

Cleaning Disposition Labels:
Ensure that disposition names are standardized and steady over the dataset.
Correct any incorrectly spelled or conflicting temperament tags.
Consider gathering comparable temperaments together for superior examination and recommendations.

Handling Outliers:
Examine numerical highlights, such as tune term or escalated, for outliers.
Decide whether to evacuate exceptions or change them based on the nature of your data.

User Interaction Data:
Aggregate or summarize client interaction information to make important highlights. For case, calculate the normal play tally for each song.
Normalize or scale numerical highlights to guarantee they are on a comparable scale.

Sentiment Analysis:
If you have estimation examination scores for tune verses or surveys, join them into the dataset.
Transform assumption scores to adjust with the temperament scale utilized in the dataset.

Collaborative Shifting Data:
Preprocess collaborative sifting information, guaranteeing that client similitude lattices or other collaborative highlights are legitimately computed and integrated.
Feature Extraction:
Feature extraction is a basic step in building a mood-based music proposal framework. Highlights are the characteristics or properties of tunes and client

intuitive that the proposal calculation employments to make forecasts. Here are a few key highlights you might consider extricating for a mood-based proposal system:

Intensity/Valence:
If not as of now given in the dataset, calculate or infer a numerical esteem speaking to the concentrated or valence of the mood.

Feature Extraction
Feature extraction involves identifying the characteristics or properties of songs and user interactions that the recommendation algorithm will use to make predictions. Key features to consider for a mood-based recommendation system include song features (such as tempo, key, and genre), user interaction features (such as frequency of listening and skip rates), playlist features (such as thematic consistency and diversity), and collaborative filtering features (such as user and item similarities).

D. Supervised Learning for Mood Detection
Supervised learning for mood detection entails training a model on a labeled dataset where the labels indicate the mood associated with songs or user interactions. The aim is to develop a predictive model that can accurately classify or predict the mood of a song based on its extracted features. This involves selecting appropriate algorithms, training the model, and evaluating its performance to ensure reliable mood detection.

Labeled Dataset:
Create a dataset with labeled illustrations where each occurrence compares to a melody, and the name speaks to its temperament. Guarantee that you have a assorted and agent dataset for diverse moods.
Feature Extraction:
Extract significant highlights from the dataset, such as sound highlights, verses estimation scores, and client interaction features.
i) Information Splitting:
Split the dataset into preparing and testing sets. The preparing set is utilized to prepare the demonstrate, and the testing set is utilized to assess its performance.

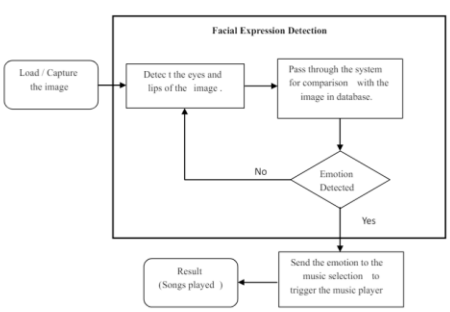
ii) Information Preprocessing:
Standardize or normalize numerical highlights to guarantee that they are on a comparable scale.
Encode categorical factors, such as disposition names, utilizing methods like one-hot encoding.
iii) Show Selection:
Choose a reasonable administered learning calculation for disposition classification. Common choices include:
Logistic Regression
Decision Trees
Random Forest
Support Vector Machines (SVM)
iv) Show Training:
Train the chosen demonstrate utilizing the preparing dataset. The demonstrate learns the designs and connections between highlights and temperament names amid this phase.
v) Demonstrate Evaluation:
Evaluate the model's execution on the testing set utilizing measurements such as exactness, exactness, review, and F1 score. Alter the demonstrate parameters if necessary.

E. Collaborative Filtering for Music Recommendation
Collaborative filtering is a technique widely used in recommendation systems to predict user preferences based on the preferences of similar users. In a mood-based music recommendation system, collaborative filtering can identify patterns and similarities between users or songs based on their mood preferences, thereby enhancing the personalization and relevance of recommendations.
i) Item-Based Collaborative Filtering:
Data Preparation:
Like user-based collaborative sifting, make a user-item framework with temperament preferences.
Normalize the framework to handle rating scales or biases.
Similarity Calculation:
Calculate the similitude between melodies based on client disposition inclinations. Likeness measurements may incorporate cosine similitude or Pearson correlation.
Prediction:
Predict the disposition inclinations for a client on unrated melodies by combining the inclinations for comparable songs.
Implementation Considerations:

Deal with the sparsity of the user-item lattice by considering as it were the most comparable items.
Update the user-item network as modern disposition inclinations are collected.
ii)Crossover Approaches:
Combine Client and Item-Based:
Combine both user-based and item-based collaborative sifting to advantage from the qualities of both approaches.
Incorporate Content-Based Filtering:
Include content-based highlights, such as tune qualities or verses estimation, to improve recommendations.
iii) Assessment Metrics:
Use fitting assessment measurements, such as Cruel Squared Mistake (MSE) or Root Cruel Squared Blunder (RMSE), to survey the execution of the collaborative sifting model.
Consider utilizing exactness, review, and F1-score for top-N recommendations.
Adaptability and Efficiency:
Optimize the calculation for versatility, particularly as the measure of the user-item framework grows.
Implement procedures like lattice factorization or parallel handling to progress efficiency.

F. Result
The proposed model saves sets of images representing various emotions (normal, sad, surprise, and happy) for comparison purposes. New images are compared with this saved dataset to detect the user's emotion accurately. Additionally, analyzing song lyrics can further improve the performance of recommendation models by providing deeper insights into mood-specific content. This feature represents a potential future enhancement for the project, offering a more comprehensive understanding of user preferences and emotional states.

III. FLOW DIAGRAM



EMOTION ACCURACY TESTING RESULTS
Set of images for each emotion (normal, sad, surprise and happy) are saved in the proposed model for the comparison purposes. The newly load images will be compared with the saved dataset to detect the emotion of the users. Table below showed the set of images that saved in the proposed model.

Table: Dataset of images saved in model

Images	Emotion
	Normal
	Sad
	Happy
	Surprise
	Angry

(Source from http://www.dreamstime.com)

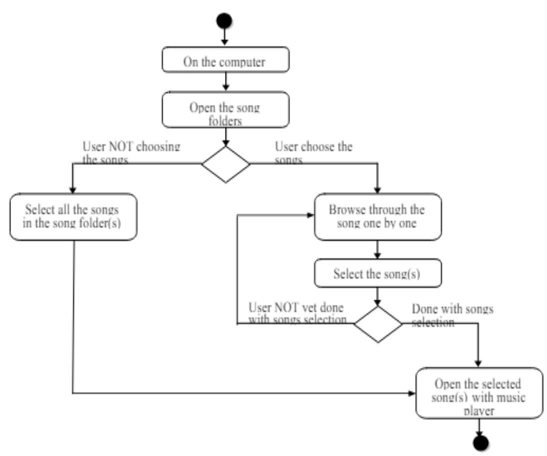


Fig: The As-Is user process Flowchart

IV. LITERATURE SURVEY

R. L. Rosa, D. Z. Rodríguez, and G. Bressan investigate how sentiment metrics derived from social networks can improve music recommendation systems. They emphasize that factors such as age, educational level, and gender can significantly influence sentiment intensity. Their research resulted in a correction factor to adjust sentiment values for more accurate user profiling and better music recommendations.[1]

V. Moscato, A. Picariello, and G. Speri developed "Viby," an emotion-based music recommender that aims to enhance user mood through music therapy. Viby uses emotion detection algorithms that continuously improve via content-based filtering. The authors suggest future improvements could include integrating physiological data such as heart rate and body temperature to enhance emotion detection accuracy.[2]

Klos, Maria Carolina et al. explored the use of an AI chatbot, Tess, for anxiety and depression support among university students. Their pilot study in Argentina showed that students who engaged more with Tess provided positive feedback, indicating the potential for AI chatbots in mental health interventions.[3]

Ivana Andjelkovic, Denis Parra, and John O'Donovan presented "Mood Play," a hybrid music recommender system that uses mood tags for interactive music discovery. Their study found that users enjoyed exploring moods through an interactive interface, though there was a noted cognitive strain from using recommendation trails. Future work will explore user interaction levels and cognitive load.[4]

Rahman, Jessica Sharmin et al. examined the effectiveness of using physiological signals such as EDA and BVP to classify music genres and emotional responses. They found neural networks to be highly accurate in classifying music based on both genre and emotion. Future research will focus on enhancing classification methods and integrating more complex features.[5]

Tris Atmaja and Masato Akagi investigated the use of acoustic features and word embeddings for speech emotion recognition (SER). Their findings showed that combining these features improved the prediction of emotional dimensions like valence, arousal, and dominance. They suggest that future research should refine these models using psychological theories [6]

Manas Sambare described a recommendation system that uses facial emotion recognition to suggest music, books, and movies. Utilizing a convolutional neural network, the system achieved good accuracy. Future work will explore incorporating additional emotions and user preferences to enhance recommendation accuracy [7]

Nitisha Raut explored machine learning techniques for facial emotion recognition, which is essential for developing intelligent systems that respond to human emotions. The paper discusses the challenges and potential algorithms for improving emotion detection from images or video [8]

Hemanth P et al. introduced "EMOPLAYER," a music player that selects tracks based on real-time emotional analysis of users' facial expressions. This system aims to improve user interaction and mood regulation through music. The authors note the potential for further advancements in emotion-based music recommendation systems [9]

V. CONCLUSION

In this paper, we employed Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN) for image recognition, and utilized several machine learning algorithms for song recommendation, including content-based filtering, collaborative filtering, and Natural Language Processing (NLP). Our project showcases a comprehensive and user-focused approach to music recommendation, harnessing advanced technologies to create a personalized and emotionally attuned music discovery experience.

The mood-based music recommendation system aims to transform how users find and enjoy music. By integrating facial expression recognition, the system adds a unique layer to understanding user preferences, enhancing the intuitiveness and personalization of recommendations. User interaction features such as like/dislike feedback and history tracking enhance the system's adaptability and continuous improvement.

The success of the project is measured not only by the accuracy of song recommendations but also by the positive impact on user satisfaction and engagement. Future improvements could include refining facial expression recognition models, incorporating more sophisticated machine learning techniques, and expanding integration with popular music streaming services. Given the array of music streaming options that lack current-emotion-specific recommendations, this project addresses a significant need by combining emotion-based music recommendations with accessible music therapy, making it available to a broad audience.

REFERENCES

[1] R. L. Rosa, D. Z. Rodríguez, and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," in Proceedings of the 2022 IEEE International Conference on Consumer Electronics (ICCE), 2015, pp. 383-384. DOI: 10.1109/ICCE.2015.7066455.

[2] V. Moscato, A. Picariello, and G.36, no. 05, pp. 57-68, 2021. DOI: 10.1109/MIS.2020.3026000.

[3] M. C. Klos, 5, no. 8, e20678, 12 Aug. 2021. DOI: 10.2196/20678

[4] I. Andjelkovic, D. Parra, and J. O'Donovan, "Mood Play: Interactive Mood-based Music Discovery and Recommendation," in Proceedings of the 2020 ACM Conference on Recommender Systems, 2020. DOI: 10.1145/2930238.2930280.

[5] J. S. Rahman, T. Gedeon, S. Caldwell, R. Jones, and Z. 11, no. 1, pp. 5-20, 2021. DOI: 10.2478/jaiscr-2021-0001.

[6] B. Tris Atmaja and M. 9, 2020. DOI: 10.1017/ATSIP.2020.14.

[7] M. Sambare, "FER2013 Dataset," Kaggle, 19 July 2020. Accessed:9 September 2020.[Online]. Available:https://www.kaggle.com/msambare/fer2013.

[8] N. Raut, "Facial Emotion Recognition Using Machine Learning," Master's Projects, San Jose State University, 2021. DOI: 10.31979/etd. w5fs-s8wd.

[9] Hemanth P, Adarsh, Aswani C. B., Ajith P., Veena A. 5, no. 4, pp. 4822-4827, April 2021.

[10] S. Kaya, D. Kabakci, I. Katircioglu, and K. Kocakaya, "Music Recommendation System: Sound Tree," presented at Dcengo Unchained, supervised by Prof. Dr. İsmail Hakkı Toroslu and Prof. Dr. Veysi İşler, sponsored by ARGEDOR, 2021.