JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Research Paper on Emotion Based Music Recommendation using Machine Learning

RIKIN ZALA[#1], HARSH THAKKAR[#2], JAYNIL ZALA[#3], PROF. BABITA PATEL[#4]

[1]B. Tech Student, Department of CSE, IITE - Indus University, Ahmedabad, Gujarat, India

[2]B. Tech Student, Department of CSE, IITE - Indus University, Ahmedabad, Gujarat, India

[3]B. Tech Student, Department of CSE, IITE - Indus University, Ahmedabad, Gujarat, India

[4]Assistant Professor, Department of CSE, IITE – Indus University, Ahmedabad, Gujarat, India

***Abstract* - An advanced Emotion-based Music Recommendation System is proposed, leveraging Speech Emotion Recognition (SER), Speech-Text Emotion Recognition (TER). The Convolutional Neural Networks (CNNs) model categorizes songs based on user's emotional states - angry, happiness, calmness and sadness, integrates with music streaming platforms (Spotify), and features a user-friendly interface for personalized music discovery. With music generation (EMOPIA) capabilities and ongoing recommendation refinement, this system aims to revolutionize music engagement by resonating with individual emotions and preferences.**

***Keywords* - Emotion Recognition, Convolutional Neural Networks, Music Recommendation, Music Generation, Valence-Arousal mapping.**

I. INTRODUCTION

In the dynamic world of music consumption, providing listeners with a personalized and engaging experience is key. Standard recommendation systems often rely solely on users' past listening behaviour, overlooking their present emotional context. The Emotion-based Music Recommendation System aims to bridge this gap by employing advanced Speech Emotion Recognition (SER) and Emotion Music Classification to deliver music selections that align with a listener's current emotional state. This innovative approach creates a more immersive music journey, fostering stronger connections between users and their preferred tunes. By seamlessly integrating with popular music streaming platforms and offering an intuitive user interface, the system transforms the way individuals discover music, enriching their listening experiences through emotionally attuned playlists.

II. PROBLEM DEFINITION

The current landscape of music recommendation systems lacks a nuanced approach that considers the emotional state of users in real-time. Traditional methods focus on user preferences and listening history, which can lead to impersonal and sometimes unsatisfactory recommendations. There is a need for a system that leverages Speech Emotion Recognition (SER) and Emotion Music Classification to analyse user's emotional cues from speech patterns and categorize music tracks by their emotional content. This system must seamlessly integrate with popular music streaming platforms to provide a diverse selection of personalized musical selections that resonate deeply with each user's current mood, enhancing their listening experience and emotional engagement with music.

III. OBJECTIVE OF THE PROJECT

The Emotion-based Music Recommendation System revolutionizes music discovery by personalizing selections based on users' emotional states, offering a unique and enriching listening experience. Utilizing advanced Speech Emotion Recognition (SER), the system analyses speech patterns to detect nuanced emotional cues and translates them into curated playlists tailored to each user's mood. Through Emotion Music Classification, deep learning algorithms categorize tracks by their emotional content, considering attributes like tempo and pitch for accurate recommendations. By partnering with major music streaming platforms, such as Spotify, the system provides access to extensive music libraries across genres and moods. A custom web application offers an immersive, intuitive user interface, enabling seamless interaction with the recommendation system via speech or manual selection. Prioritizing user experience and real-time emotion detection, the project seeks to deepen connections between users and their music, enhancing their lives through personalized melodies.

IV. LITERATURE STUDY

A. Existing Systems

Emotion-Based Music Player [1] - This project aims to effectively extract features and facial elements for emotion detection, proposing a method to generate music that aligns with the detected emotion.

Smart Music Player with Facial Emotion Recognition and Music Mood Recommendation [2] - This Android application captures the user's image to identify four emotions and uses an algorithm to create a playlist based on the detected emotions. It also allows users to add or skip songs.

Real-Time Emotion Recognition from Audio [3] - The system focuses on capturing audio signals and processing them quickly on computing devices. It extracts features to measure the degree of similarity in emotions, though this research encountered challenges with efficiency.

Smart Music Player with Facial Emotion Recognition and Music Mood Recommendation [4] - This project relies on a database using Olivetti's face dataset, which includes 400 images of various emotions. It operates with an SVM algorithm that separates data into test and training sets, learning from facial features to identify emotional states (happy or sad) and play corresponding music.

Mood Cloud [5] - It is a real-time music mood visualization tool that classifies music emotions into five categories: aggressive, happy, party, relax, and sad. It uses an SVM library to analyze an emotion database, presenting the results using Flash player.

B. *Proposed System.*

The proposed Emotion-based Music Recommendation System introduces a multifaceted approach to enrich users' musical experiences by personalizing selections according to their emotional states. The system performs acoustic analysis through valence-arousal mapping, which assesses the emotional intensity and positivity of audio signals to align with the listener's current mood. Text similarity analysis further refines recommendations by examining lyrical content and matching it with the user's emotions and preferences. Additionally, the system employs keyword-semantics analysis to capture subtle nuances in both user input and music metadata, ensuring precise emotional alignment. A unique aspect of the system is the incorporation of AI-generated instrumental songs, offering personalized compositions crafted specifically to match the listener's emotional state and preferences. Through these integrated methods, the system delivers a seamless, immersive musical journey tailored to each user's emotional landscape.

V. BACKGROUND STUDY

The process of tailoring music recommendations to users' emotional states requires a robust framework to accurately categorize and interpret emotions. In this context, the system leverages Russell's Valence-Arousal Model, a well-established psychological model that maps emotions along two dimensions: valence (ranging from positive to negative emotions) and arousal (ranging from calm to intense emotions). This model allows the system to effectively analyze and categorize the emotional content of users' audio inputs, providing a clear understanding of their emotional states.

By employing the Valence-Arousal Model, the system organizes emotions into four distinct quadrants:

1. Happiness: Representing high valence and moderate to high arousal, this quadrant encompasses feelings of joy and elation.
2. Angry: Characterized by low valence and high arousal, this quadrant includes intense and agitated emotions such as anger.
3. Sadness: Defined by low valence and low arousal, this quadrant reflects emotions of sorrow and melancholy.
4. Calmness: Representing high valence and low arousal, this quadrant encompasses tranquil and serene emotions.

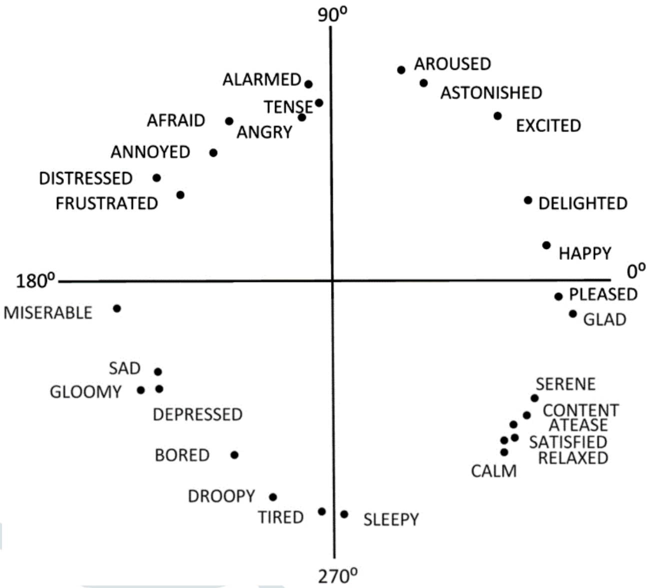

Figure 5.1: Russel's Circumplex Model [6]

Quadrant	Emotion	Valence Arousal
1	Happy	HVHA
2	Angry	LVHA
3	Sadness	LVLA
4	Calmness	HVLA

Table 5.1: Valence-Arousal Mapping based on Emotion

Emotion-based Music Recommendation System employs advanced machine learning techniques, particularly Convolutional Neural Networks (CNN), to perform speech and text emotion recognition. By analyzing the user's audio input and textual content, the system effectively discerns emotional nuances. For keyword extraction, the system utilizes the KeyBERT model to identify salient terms that capture the essence of the user's emotional state. Similarity between keywords is determined using the sentence transformer model all-MiniLM-L6-v2, which enables precise mapping between the user's emotions and available music content.

Upon recognizing the user's emotional state, the system leverages the Spotify API to retrieve a vast selection of songs, along with their corresponding valence and arousal (energy) values. These values allow the system to map the recognized emotion from the machine learning model to the emotional attributes of the retrieved music tracks. By aligning the user's emotional state with the emotional qualities of the songs, the system curates personalized playlists that resonate deeply with the user's mood, offering a truly tailored and immersive listening experience.

VI. METHODOLOGY

In an emotion-based music recommendation system that leverages Convolutional Neural Networks (CNN) for Speech Emotion Recognition (SER) and Text Emotion Recognition (TER), several critical steps ensure effective emotion detection and personalized music recommendations:

1. Data Collection and Pre-processing: Gather diverse datasets containing audio and text data, including speech samples. Pre-process the data by normalizing audio and text inputs and working on selective emotions – Angry, Calmness, Happiness, Sadness.

2. Model Development:
    a. Speech Emotion Recognition: Develop a CNN model to analyse audio features and recognize emotional states from speech patterns.
    b. Text Emotion Recognition: Build a separate CNN model to process textual inputs and identify emotional cues within the text.
    c. Valence-Arousal Mapping: Develop a Model to process each word's Valence-Arousal value and find the mean VA for whole sentence.

3. Pre-Trained Models: We utilize pre-trained models to streamline the process of recognizing user emotions and match them with appropriate music recommendations. Assembly AI transcribes audio files into text, providing valuable text data for further analysis. The KeyBERT model extracts key themes and emotional nuances from the transcribed text, while the all-MiniLM-L6-v2 sentence transformer measures semantic similarity between the extracted keywords, ensuring accurate matching with potential musical selections.

4. Feature Extraction: For SER, extract relevant features such as pitch, intensity, and spectral properties from audio inputs. For TER, employ techniques such as tokenization and embedding to prepare textual data for analysis.

5. Emotion Classification: Classify the speech and text inputs into predefined emotional categories using the trained CNN models. Combine the outputs of the SER and TER models to create a comprehensive emotional profile for the user.

6. Music Matching and Recommendation: Leverage the Spotify API to access a vast catalogue of songs and their metadata. Filter songs based on their emotional attributes (valence and arousal values) and the recognized user emotions from the SER and TER models. Match the user's emotional profile with suitable songs, ensuring the music resonates with the user's mood.

7. Playlist Creation: Curate personalized playlists based on the matching process, selecting songs that align with the user's current emotional state. Continuously update and adapt playlists as new music becomes available and as the user's emotions evolve.

8. Music Generation: integrating EMOPIA, a third-party library, to generate AI-based music. Once the model predicts the user's emotional state, we provide the VA values and emotion labels to EMOPIA, which composes personalized instrumental music tailored to the user's mood. This approach ensures a seamless and engaging listening experience aligned with the user's current emotions.

VII. UML DIAGRAMS

The Unified Modelling Language (UML) is a compact report in object-oriented computing, which serves as a useful, rule-based framework for task execution within an organization. The challenge with UML lies in its discrete nature, making it difficult to socialize. Nonetheless, UML has been recognized as the standard approach for creating object-oriented computer programs, with its key aspects being a meta-model and documentation.

UML can be integrated or applied following a planned event or occurrence. It is a common focus for achieving uniformity, considering, organizing, and documenting elements of plans without excluding plans. The structure of UML has been informed by the best approaches in composition, enabling the replication of large, disorganized systems. UML reflects current trends in design processes and regulatory measures of balanced evaluation.

The primary aims of UML are as follows:

1. By offering an organized framework for creating large-scale models, UML allows users to design and exchange comprehensive models efficiently.
2. Support fundamental, business-compatible designs and scalability.
3. Avoid assuming any specific priorities or foundations for development.
4. Facilitate resolution and standardization of the native language.
5. Promote the commercial use of object-oriented tools.
6. Encourage the implementation of high-level development techniques, including components, architectures, models, and collaborative projects.
7. Incorporate actions based on the best decisions.

1. **Activity Diagram**: An activity diagram in Unified Modeling Language (UML) is a graphical representation used to visualize the dynamic aspects of a system, focusing on the flow of control and activities performed within a process or workflow. It depicts the sequence of actions, decisions, and transitions that occur during the execution of a particular functionality or scenario within the system. Activity diagrams help to illustrate the workflow logic, including parallel and conditional behaviors, and facilitate communication and understanding among stakeholders by providing a clear and intuitive representation of system behavior.

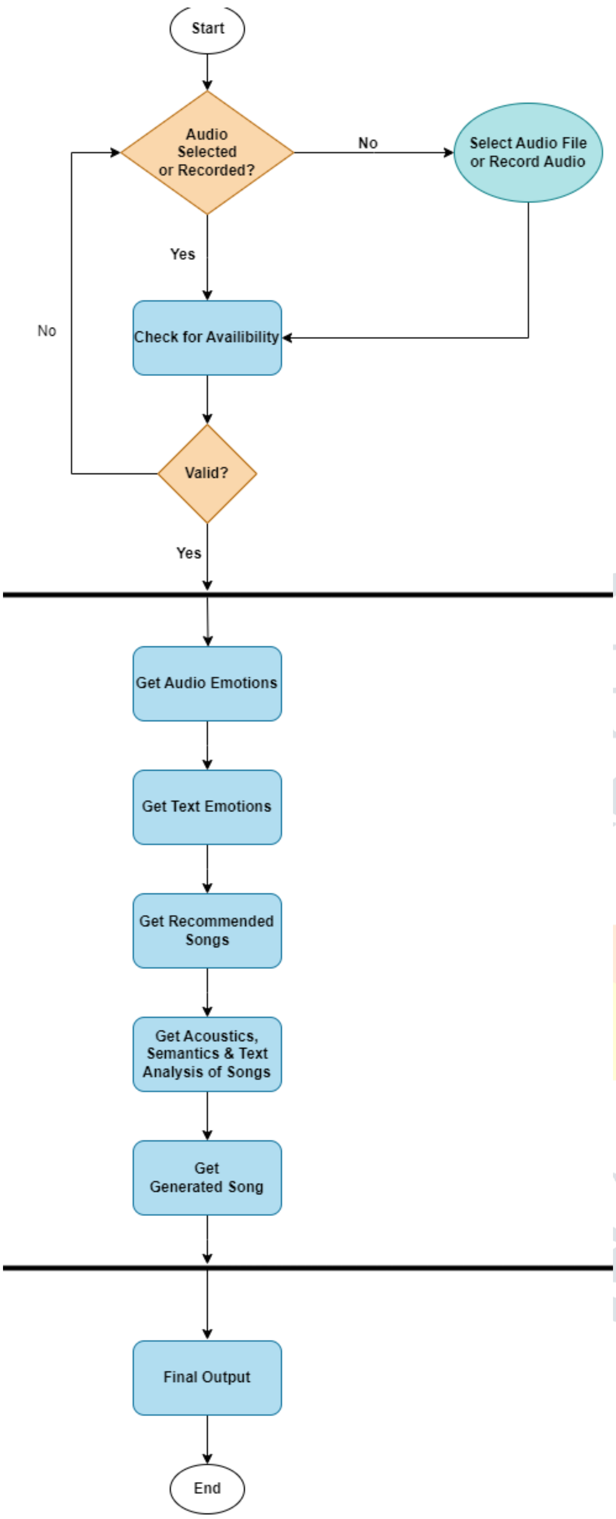

Figure 7.1: Activity Diagram

2. **Block Diagram:** A block diagram is a high-level visual representation that showcases the structure and functionality of a system by illustrating the system's components or blocks and their relationships. Blocks represent distinct system elements, such as hardware components, software modules, or subsystems, and the lines between blocks indicate connections or interactions between these elements. Block diagrams provide a simplified view of a system's architecture, helping stakeholders understand the big picture of how different parts fit and work together.

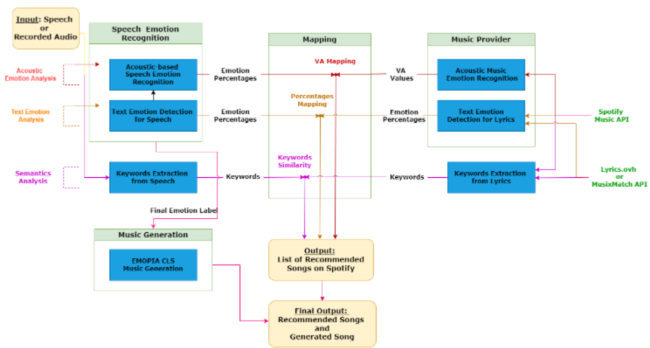

Figure 7.2: Block Diagram

3. **Data Flow Diagram:** A block diagram is a high-level visual representation that showcases the structure and functionality of a system by illustrating the system's components or blocks and their relationships. Blocks represent distinct system elements, such as hardware components, software modules, or subsystems, and the lines between blocks indicate connections or interactions between these elements. Block diagrams provide a simplified view of a system's architecture, helping stakeholders understand the big picture of how different parts fit and work together.

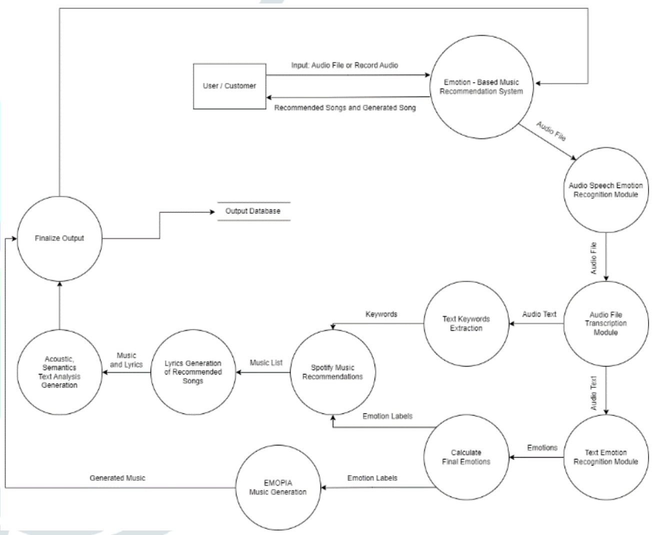

Figure 7.3: Data Flow Diagram

VIII. ALOGRITHMS

1. *Convolutional Neural Network (CNN)*

A convolutional neural network (CNN) is a type of machine learning model, specifically a deep learning network, often used for processing audio emotion recognition. CNNs consist of three main layers: convolutional, pooling, and fully connected (FC), with the convolutional layer at the beginning and the FC layer at the end.

As the network progresses from the convolutional layer to the FC layer, it becomes increasingly complex. This growing complexity allows the CNN to gradually identify more intricate parts of an audio segments until it eventually fully recognizes the pitch within it.

2. *Layers of CNN*

The convolutional layer is the foundation of a CNN, where most computations take place. This layer may include multiple convolutional layers, each containing a kernel or filter that scans the image's regions for specific features.

The pooling layer, much like the convolutional layer, uses a filter on the input image. However, unlike the convolutional layer, the pooling layer reduces the number of input parameters while sacrificing some information. This simplification helps decrease complexity and enhances the efficiency of the CNN.

The fully connected (FC) layer is where audio classification occurs based on features extracted from previous layers. "Fully connected" indicates that each activation unit or node in the next layer is linked to every input or node from the preceding layer.

a) CNN Architecture for Speech Emotion Recognition.

Input Layer: The model takes an input of shape (X_train.shape[1], 1), where X_train.shape[1] is the length of the audio input data.

First Convolutional Block: A 1D convolutional layer with 512 filters, kernel size of 5, strides of 1, and padding set to 'same' is applied with a ReLU activation function.

Followed by batch normalization. Then a 1D max pooling layer with pool size of 5 and strides of 2 is used to reduce the data size.

Second Convolutional Block: Another 1D convolutional layer with 512 filters, kernel size of 5, strides of 1, and padding set to 'same' is applied with a ReLU activation function. Followed by batch normalization. Then another 1D max pooling layer with pool size of 5 and strides of 2. A dropout layer with a dropout rate of 0.2 is added to prevent overfitting.

Third Convolutional Block: A 1D convolutional layer with 256 filters, kernel size of 5, strides of 1, and padding set to 'same' is applied with a ReLU activation function. Followed by batch normalization. Another 1D max pooling layer with pool size of 5 and strides of 2.

Fourth Convolutional Block: A 1D convolutional layer with 256 filters, kernel size of 3, strides of 1, and padding set to 'same' is applied with a ReLU activation function. Followed by batch normalization. Then another 1D max pooling layer with pool size of 5 and strides of 2. Another dropout layer with a rate of 0.2.

Fifth Convolutional Block: A 1D convolutional layer with 128 filters, kernel size of 3, strides of 1, and padding set to 'same' is applied with a ReLU activation function. Followed by batch normalization. Another 1D max pooling layer with pool size of 3 and strides of 2. A dropout layer with a rate of 0.2.

Flattening: The output of the final convolutional block is flattened to transition from convolutional layers to dense layers.

Dense Layers: A dense layer with 512 units and a ReLU activation function is applied. Followed by batch normalization. Then a final dense layer with 4 units (for the four emotion classes) and a softmax activation function is applied to output the probabilities for each class.

Compilation: The model is compiled with the Adam optimizer, categorical cross-entropy loss function, and accuracy metric.

b) CNN Architecture for Text Emotion Recognition.

Branch 1: An embedding layer takes the input text (of maximum length max_len) and converts it into an embedding space with embedding_dim dimensions. A 1D convolutional layer with 64 filters, kernel size of 3, and padding set to 'same' is applied with a ReLU activation function. Batch normalization is applied to standardize the output of the convolutional layer. Another ReLU activation function is applied. A dropout layer with a rate of 0.5 is added to prevent overfitting. A global max pooling layer aggregates the output of the convolutional layer.

Branch 2: An embedding layer converts input text into an embedding space with embedding_dim dimensions. A 1D convolutional layer with 64 filters, kernel size of 3, and padding set to 'same' is applied with a ReLU activation function. Batch normalization standardizes the output of the convolutional layer. Another ReLU activation function is applied. A dropout layer with a rate of 0.5 prevents overfitting. A global max pooling layer aggregates the output of the convolutional layer.

Concatenation: The outputs from both Branch 1 and Branch 2 are concatenated to combine the features extracted by each branch.

Hidden Layer: A dense layer with 128 units and a ReLU activation function processes the concatenated output.

Dropout: A dropout layer with a rate of 0.5 is applied to the hidden layer's output to prevent overfitting.

Output Layer: A dense layer with 4 units (for the four emotion classes) and a softmax activation function outputs the probabilities for each class.

Model Compilation: The model is compiled with the Adamax optimizer, categorical cross-entropy loss function, and metrics including accuracy, precision, and recall.

IX. IMPLEMENTATION

For the implementation of the Emotion-Based Music Recommendation System, the following steps outline the approach taken:

1. The code imports necessary packages and libraries essential for the project's functionality, including tensorflow and keras for deep learning-based models, librosa for audio analysis and processing, flask and flask-cors for creating the web application and handling cross-origin resource sharing, keybert for keyword extraction, SentenceTransformer for sentence similarity analysis, requests for handling HTTP requests, replicate for interfacing with third-party services,

spotipy for accessing the Spotify API, assemblyai for audio-to-text transcription, and pickle for serializing and deserializing objects.
2. The app.tsx file runs the main application and serves as the entry point for the program.
3. The recommend_music API is called when the "Recommend Music" button is clicked, triggering the system to provide music recommendations based on the user's input or emotional state. The generate_music API is called when the "Generate Music" button is clicked, prompting the system to generate a new music track using AI.
4. The user can upload an audio file or record audio directly within the application for emotion detection and subsequent music recommendations.
5. Upon receiving user input (audio or text), the system uses its specialized method for detecting emotion from the input data and recommends music based on the identified emotion.
6. After processing the user's input and identifying the corresponding emotion, the system uses its proprietary approach to provide music recommendations, either by retrieving songs that match the emotion from an external source (e.g., Spotify) or by generating new music.
7. The user interface returns a list of recommended songs based on the user's emotional state, along with one AI-generated track for a unique listening experience.

By leveraging the above-described methodologies and integrating the essential libraries and frameworks, the Emotion-Based Music Recommendation System has been successfully completed and is now fully functional, offering personalized music recommendations and AI-generated music based on the user's emotional state.

X. Results

The system achieves a high level of accuracy, with the implementation of an emotion-based music recommendation system utilizing convolutional neural networks (CNN). The system employs a pre-trained CNN-based deep learning model to analyze audio and text inputs, achieving 99% accuracy in speech emotion recognition (SER) and 94% accuracy in text emotion recognition (TER). Once the user's emotional state is detected, the system uses these predictions to offer music recommendations that match the user's mood.

Based on the user's predicted emotions, the system selects appropriate music files from a directory and presents a selection of songs that align with the user's mood. The user can then choose and play their preferred music. The high accuracy levels indicate the system's effectiveness, and potential improvements could be made by incorporating user feedback to further personalize the music recommendations based on individual preferences.

SER	Precision	Recall	F1-Score	Support
Anger	0.99	0.99	0.99	1583
Calmness	0.99	0.99	0.99	1450
Happiness	0.99	0.99	0.99	1520
Sadness	0.99	0.99	0.99	1579
accuracy			0.99	6132
macro avg	0.99	0.99	0.99	6132
weighted avg	0.99	0.99	0.99	6132

Table 10.1 SER Model Classification Report

TER	Precision	Recall	F1-Score	Support
Anger	0.94	0.93	0.93	275
Calmness	0.78	0.84	0.81	159
Happiness	0.95	0.94	0.94	695
Sadness	0.97	0.96	0.96	581
accuracy			0.94	1710
macro avg	0.91	0.92	0.91	1710
weighted avg	0.94	0.94	0.94	1710

Table 10.2 TER Model Classification Report

Based on these impressive results, the system demonstrates its capability to accurately recognize user emotions through speech and text analysis and provide highly relevant music recommendations, offering users a uniquely personalized and satisfying music-listening experience.

XI. Project Snapshots

The "Project Snapshots" section presents a visual overview of the Emotion-Based Music Recommendation System, offering insight into the system's features and functionality. Through a series of carefully curated images, users can explore the user interface, key components, and processes involved in providing personalized music recommendations based on emotional recognition. These visuals illustrate how the system operates, demonstrating its seamless integration of deep learning models and external music sources to enhance the user experience.

1. Initial Startup Page: The entry point of the application, where users can initiate the emotion-based music recommendation process and explore available options.

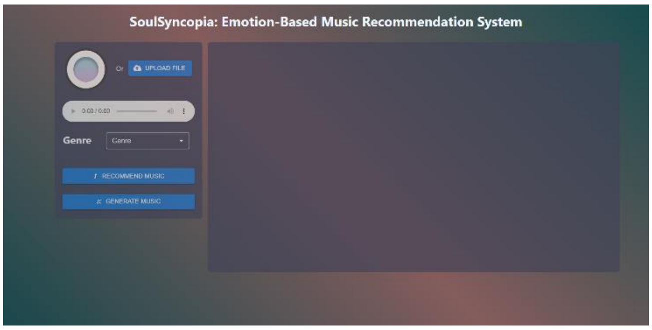

Figure 11.1: Initial Startup Page

2. Recommended and Generated Music Output: Displays the selection of music tracks recommended based on the user's emotional state and an AI-generated music piece tailored to the user's mood.

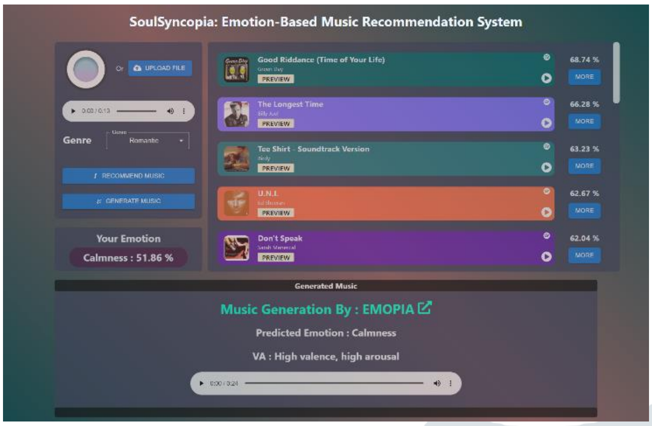

Figure 11.2: Recommended and Generated Music Output

3. Recommended Music Analysis / Full Page: Offers insights into the recommended music choices, explaining how they align with the user's detected emotions and preferences.
It also provides a comprehensive overview of the user's selections and preferences, showcasing all recommended and generated music for the current session.
Again, it displays a complete overview of the user's recommended and generated music selections, including an analysis of how each choice aligns with the user's detected emotional state and individual preferences.

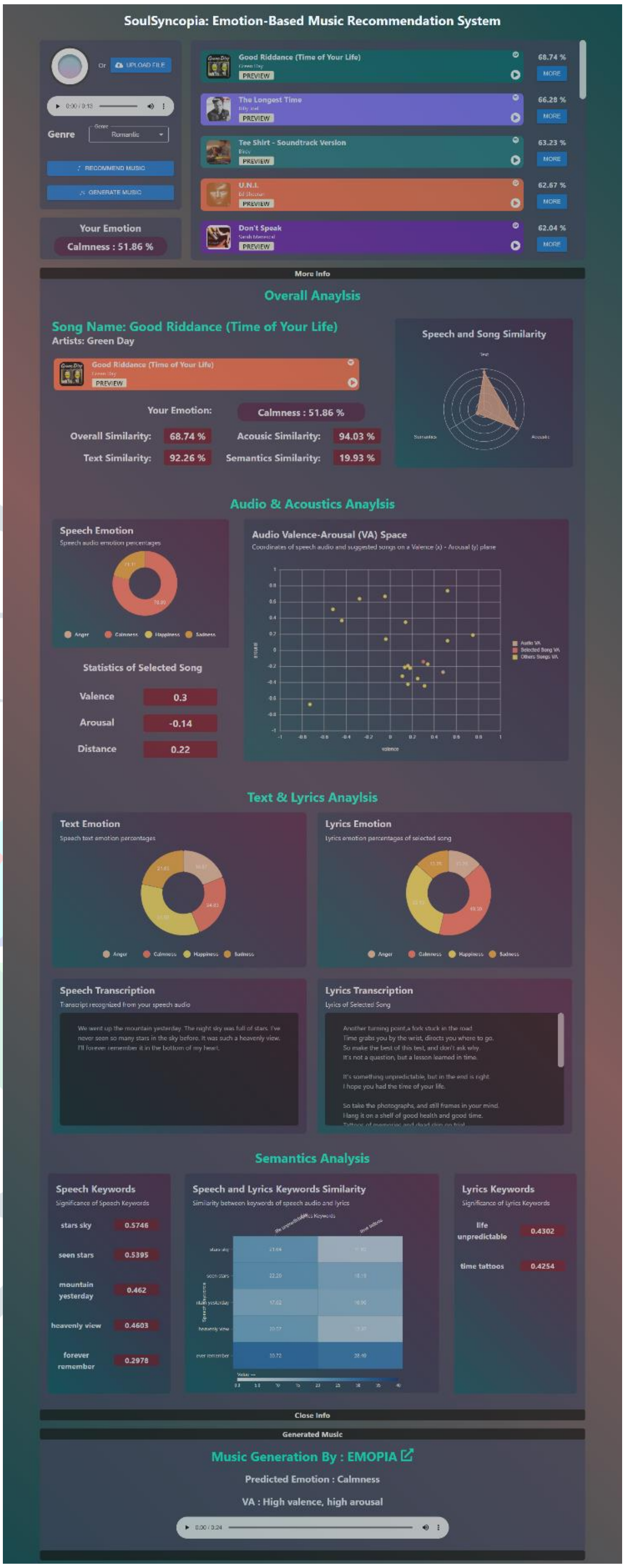

Figure 11.3: Full Page with Analysis

XII. LIMITATIONS AND FUTURE ENHANCEMENTS

Limitations: The Emotion-Based Music Recommendation System faces several limitations that could impact its accuracy and effectiveness. Variability in the accuracy of emotion recognition algorithms arises from factors such as audio quality, cultural nuances, and individual differences in emotional perception. This variability can affect the system's ability to provide relevant and effective music recommendations.

Additionally, the subjectivity of emotional responses means that emotions can be interpreted differently by individuals, which may lead to discrepancies in the perceived emotional content of music and the system's recommendations.

Another limitation is the availability and diversity of the training dataset used for emotion recognition. A limited or biased dataset can result in inadequate coverage of certain emotional states or cultural contexts, potentially leading to biased or incomplete recommendations. Cross-cultural challenges also exist, as cultural differences in emotional expression and musical preferences pose obstacles to creating a universal recommendation system. The system may struggle to accurately interpret and cater to diverse emotional and cultural preferences, including processing different language models. Technical constraints, such as computational resources and processing power, may further limit the system's scalability and performance, particularly during large-scale deployment and real-time processing.

Future Enhancements: Future plans and advanced research for the Emotion-Based Music Recommendation System aim to address these limitations and improve the system's capabilities. Refining emotion recognition algorithms to accurately detect and interpret user emotions from various data sources, such as voice analysis and physiological signals, is essential. Incorporating contextual cues like time of day, location, and activity can lead to more relevant and personalized music suggestions aligned with users' emotional and situational contexts.

Multi-modal data fusion is another potential enhancement, as it captures a more comprehensive understanding of user emotions and preferences by combining audio, visual, and textual data. Scaling up the recommendation system infrastructure is crucial for accommodating large user bases while maintaining personalized recommendations tailored to individual emotional states and preferences. Additionally, implementing user feedback mechanisms and exploring cross-cultural adaptation can further refine the system's accuracy and user satisfaction. Ethical considerations and user privacy safeguards must also be prioritized to ensure transparency, accountability, and the protection of user data in the system's ongoing development.

XIII. CONCLUSION

The Emotion-Based Music Recommendation System provides a personalized listening experience by tailoring music suggestions to users' emotional states. This approach enhances user engagement and satisfaction, contributing to higher retention rates and increased platform loyalty. The project's success highlights technological advancements in emotion recognition and its practical application in enhancing digital music platforms.

By delivering music that resonates with users' current moods, the system has the potential to positively impact emotional well-being. Future research can further refine the system by improving emotion recognition algorithms and integrating contextual cues while addressing ethical considerations. Overall, the project demonstrates the potential for technology to create more personalized and emotionally resonant music experiences, enriching users' enjoyment and engagement with digital music platforms.

References

[1] Dhurvisha Bansal, Pinkal Bhatt, Megha Dusane, Avneet Saluja, Kushal Patel“Emotion Based Music Player” Report-2020.

[2] AH. Immanuel James, James Arnold, Marta Ruban, R. Saranya“Generating Music Playlist Based On Facial Expression”: pISSN 2395-0072, IRJET 2018

[3] Hemant Yadav, AdarshSingh, AswaniIyer, Ajit Pawar “Smart MusicPlayer IntegratingFacial Emotion Recognition And Music

[4] Mood Recommendation ”Students, Department of computer science, Saintgits College Engineering Issue: 04 April2018 .

[5] Sri Charan Nimmagadda “Emotion Based Music Player ” Report-2017.

[6] Russell’s (1980) Circumplex Models.

[7] Emotional Detection and Music Recommendation System based on User Facial Expression.

[8] Study on Music Emotion Recognition Based on the Machine Learning Model Clustering Algorithm

[9] Music Recommendation Based on Face Emotion Recognition

[10] Music Emotion Recognition Based on a Neural Network with an Inception-GRU Residual Structure.