SAVITRIBAI PHULE PUNE UNIVERSITY

2024-2025

A PROJECT REPORT ON

Speech Emotion Based Song Playlist Recommendation System Using Deep Learning

SUBMITTED TOWARDS THE
PARTIAL FULFILLMENT OF THE REQUIREMENTS OF

BACHELOR OF ENGINEERING

(Computer Engineering)

By

Mayank Adhav	B400080500
Aaditya Meher	B400080501
Jeet Harne	B400080469
Pratik Hagawane	B400080468

Under The Guidance of

Mrs. Soudamini T. Somvanshi

DEPARTMENT OF COMPUTER ENGINEERING

D. Y. PATIL COLLEGE OF ENGINEERING,

AKURDI, PUNE-44

D. Y. PATIL COLLEGE OF ENGINEERING, AKURDI, PUNE-44

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Project Entitled

Speech Emotion Based Song Playlist Recommendation System using Deep Learning

Submitted by

Mayank Adhav	B400080500
Aaditya Meher	B400080501
Jeet Harne	B400080469
Pratik Hagawane	B400080468

is a bonafide work carried out by students under the supervision of Mrs. Soudamini T. Somavanshi and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering).

Mrs. Soudamini T. Somavanshi
Internal Guide
Dept. of Computer Engg.

Dr. M. A. Potey
H.O.D
Dept. of Computer Engg.

Dr. Mrs. P Malathi
Principal
D. Y. Patil College of Engineering, Akurdi, Pune 44

Signature of Internal Examiner

Signature of External Examiner

PROJECT APPROVAL SHEET

A Project Title

<u>Speech Emotion Based Song Playlist Recommendation System using Deep Learning</u>

Is successfully completed by

Mayank Adhav	B400080500
Aaditya Meher	B400080501
Jeet Harne	B400080469
Pratik Hagawane	B400080468

at

DEPARTMENT OF COMPUTER ENGINEERING

D. Y. PATIL COLLEGE OF ENGINEERING, AKURDI-44

SAVITRIBAI PHULE PUNE UNIVERSITY,

PUNE ACADEMIC YEAR 2024-2025

Mrs. Soudamini T. Somvanshi
Internal Guide
Dept. of Computer Engg.

Dr. M. A. Potey
H.O.D
Dept. of Computer Engg.

Abstract

In the rapid era of digital life today, music is no longer simple entertainment it is a friend that reflects and affects our mood. Yet, conventional music recommendation systems are usually based on user behavior, preferences, or explicit choices, providing insufficient emotional intuition. Motivated by these growing concerns, this project proposes to bridge this gap by creating an intelligent, speech emotion-driven music playlist recommendation system that dynamically adjusts to the emotional state of the user in real time.

The fundamental concept is straightforward that let the user naturally talk into a microphone, input his/her audio and let the system identify their underlying emotion based on speech alone. Leveraging deep learning methods, specifically convolutional neural networks, the system processes vocal characteristics such as tone, pitch, and energy to precisely label emotions like happiness, sadness, anger, fear, and neutrality etc. Depending on the emotion that has been detected, the system then dynamically suggests or creates a personalized playlist that aligns with the mood of the user.

What distinguishes this system is its non-intrusive and human-oriented approach — users don't need to manually type in their mood or navigate through complicated interfaces. Unlike traditional recommenders, this system emphasizes emotional relevance over popularity or listening history, resulting in a deeply personalized and emotionally supportive music experience. Rather, their emotions are naturally sensed and interpreted by their own voice, thus rendering the experience emotionally aware and seamless. Apart from individual music experiences, this affect-based recommendation platform can be modified for numerous real-world applications. It can serve as an auxiliary tool for patients suffering from emotional distress in the health care field by providing them with music that corresponds with their state of mind. It can also improve student interest in the educational sector by giving them emotionally relevant background music while relaxing which is as important as study.

Acknowledgments

*It gives us great pleasure in presenting the preliminary project report on '**Speech Emotion Based Song Playlist Recommendation System Using Deep Learning'.***

*I would like to take this opportunity to thank my internal guide, **Mrs. Soudamini T. Somvanshi**, for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestion were very helpful*

*I am also grateful to **Prof. M. A. Potey**, Head of the Computer Engineering Department at D. Y. Patil College of Engineering, Akurdi, for their indispensable support and suggestions.*

*In the end our special thanks to **Mrs. Supriya Sathe** for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project.*

Mayank Adhav
Aaditya Meher
Jeet Harne
Pratik Hagawane
(B.E. Computer Engg.)

Contents

List of Tables

LIST OF FIGURES

1. Synopsis

1.1. Project Title

Speech Emotion Based Song Playlist Recommendation System Using Deep Learning

1.2. Project Option

1.2.1. Speech Emotion Recognition using DL

1.2.2. Real Time Emotion Detection from Users audio input

1.3. Internal Guide

Mrs. Soudamini T. Somvanshi

1.4. Sponsorship and External Guide

None

1.5. Technical Keywords (As per ACM Keywords)

1.5.1. Computing Methodologies (Artificial Intelligence, Machine Learning, Speech Recognition, Neural Networks)

1.5.2. Software Engineering (Requirements, Specifications, Design, Software Architectures, Development, Testing and Debugging, Maintenance and Enhancement)

1.5.3. Data Storage Representations/Data Encryption

1.5.4. Database Management

1.5.5. Information Storage and Retrieval (Content Analysis and Indexing)

1.5.6. Information Interfaces and Presentation (User Interfaces, Audio input section)

1.5.7. Pattern Recognition (MFCC, Spectrograms)

1.6. Problem Statement

To develop a system that analyzes a user's real-time emotional state by taking audio input and classifying it to an emotion to provide personalized music recommendations, addressing the limitations of traditional algorithms that rely solely on past behavior and genre preferences.

1.7. Abstract

In the emotionally engaged digital age of today, this project proposes a speech emotion-based music playlist recommendation system that can change in real time with a user's emotional state. In contrast to conventional systems based on user action or manual data, this system employs deep learning using Convolutional neural network to process vocal characteristics such as tone, pitch, and energy from speech in order to detect emotions like happiness, sadness, anger, fear and neutrality. According to the identified emotion, it suggests a user specific playlist based on the user's present mood, providing an emotionally aware experience. The system allows user to convey emotion in a natural way via speech. Apart from personal entertainment, it is poised to extend its application in mental health assistance, education, home automation and virtual assistants.

1.8. Goals and Objectives

1. Provide Emotion Driven Music Experience: Develop a system that recommends songs based on the user's real-time emotional state, enhancing the emotional connection to music and improving user engagement.
2. To Enhance Accuracy and Reliability: Improve the accuracy and reliability of the detection system through the usage of training data ensuring high performance across various scenarios and conditions.
3. Real-Time Emotion Detection from Speech: Utilize deep learning models to recognize emotions from speech input in real time, ensuring fast and accurate playlist generation tailored to the user's current mood.
4. Seamless Integration with Music Platform: Integrate the system with popular music streaming services (e.g., Spotify, Apple Music) to provide continuous, personalized music recommendations based on emotional changes.
5. Improve User Experience through Dynamic Playlist Updates: Allow dynamic updates to the playlist as the user's mood shifts, maintaining a constantly relevant and engaging music selection.

1.9. Relevant mathematics associated with the Project

1. Linear Algebra: Essential for matrix operations in speech processing, which are crucial for extracting features like Mel-frequency cepstral coefficients (MFCCs) from audio data. These features help the model recognize the nuances of spoken input.

2. Signal Processing: Utilized in feature extraction from audio signals (e.g., MFCCs), signal processing enables the system to analyze speech inputs accurately. This step is vital for identifying and interpreting users' emotional cues from their speech.
3. Optimization Algorithms: Gradient Descent and Backpropagation are used to train deep learning models like Convolutional Neural Networks (CNN).
4. Convolutional Neural Networks (CNNs): CNNs are commonly used for image analysis tasks such as facial recognition and liveness detection.
5. Mel-frequency Cepstral Coefficients (MFCC): MFCCs are extracted from speech data to capture important features representing the vocal tract's shape. These coefficients are critical for emotion detection, serving as inputs to machine learning models for classification.
6. Emotion Detection for Speech: Let S(t) represent the speech signal as a function of time. The system extracts feature F(S(t)) and classifies them into an emotion E using a function $f(F(S(t))) \rightarrow E$.
7. Recommendations of Playlist: Once the emotion E is detected, the system uses a function $g(E) \rightarrow P$ to map the emotion to a playlist P, where P represents a collection of songs that match the emotion.

1.10. Names of Conferences / Journals where papers can be published

a. IEEE Conference and Journal

1. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP): Focuses on speech recognition, emotion detection, and machine learning applications in audio processing.
2. IEEE International Conference on Machine Learning and Applications (ICMLA): Suitable for research on deep learning methods applied to speech emotion recognition and personalized music recommendations.
3. IEEE Transactions on Audio, Speech, and Language Processing: A high-impact journal for research on speech processing, emotion recognition, and deep learning techniques in audio analysis.
4. IEEE Transactions on Neural Networks and Learning Systems: Ideal for publishing advanced work in deep learning, neural networks, and emotion recognition in speech.

5. IEEE International Conference on Artificial Intelligence and Signal Processing (AISP): Good for AI-driven speech emotion recognition systems and personalized music recommendation algorithms.

b. ACM Conferences and Journals:

1. ACM Conference on Multimedia (ACM MM): Focuses on multimedia content analysis, including audio emotion recognition and music recommendation.
2. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD): Suitable for research on emotion-based song recommendations using deep learning and data mining techniques.
3. ACM Transactions on Intelligent Systems and Technology (TIST): Publishes work on AI, machine learning, and intelligent systems, including emotion recognition and music recommendation.
4. ACM International Conference on Human-Computer Interaction with Mobile Devices and Services: Relevant for papers that focus on interactive music systems using speech emotion recognition for personalized recommendations.
5. ACM Transactions on Audio, Speech, and Music Processing (TASMP): A specialized journal for research on audio, speech, and music technologies, including emotion-based systems for song recommendations.

1.11. Review of Conference/Journal Papers supporting Project idea

1. **Smita Bhosale et al., "Speech Emotion Based Music Recommendation System" [1], published in the *International Journal of Innovative Science and Research Technology*, Vol. 9, Issue 4, April 2024,** introduces a system that utilizes speech emotion recognition to recommend music, aiming to improve user emotional states. This "Viby" system focuses on personalizing music for therapeutic benefits like stress reduction and mood enhancement, addressing gaps in existing platforms. Its effectiveness hinges on accurately detecting subtle emotions across diverse voices and noisy environments, which presents a significant practical challenge for consistent performance.
2. **Chirag Kothawade et al., "A Survey on Speech Emotion Based Music Recommendation System" [2], published in the *International Journal of***

***Scientific Research in Engineering and Management (IJSREM)*, Vol. 08, Issue 04, April 2024, DOI: 10.55041/IJSREM30150,** presents a survey on speech emotion-based music recommendation systems, emphasizing music's therapeutic benefits for mental health. It highlights the lack of holistically curated algorithms in existing systems despite music's proven positive impacts on cognition, memory, and stress. The paper aims to advocate for a comprehensive "Viby" system that extensively extracts music's benefits by analyzing user speech. A limitation is that as a survey, it may not propose novel technical solutions, but rather summarize existing ones.

3. **Emotion Based Music Recommendation System [3], published in the *International Journal of Research Publication and Reviews*, Vol. 4, No. 12, December 2023,** proposes an emotion-based music recommendation system leveraging machine learning, specifically deep neural networks, implemented in Python. It aims to enhance user satisfaction by recommending music tailored to emotional context, utilizing robust data preprocessing and feature extraction from audio signals. The system's efficacy is evaluated through recommendation accuracy and adaptability. However, its performance heavily relies on comprehensive emotion-labeled datasets, and scaling robustly to handle diverse real-world emotional expressions and subtle user preference shifts remains a challenge not fully elaborated.
4. **Music Moody - Facial Recognition and Voice Recognition To Detect Mood And Recommend Songs [4], published in the *International Journal of Advanced Research and Publications*, Volume 6, Issue 10, October 2023,** presents a comprehensive music recommendation system integrating live voice recognition for mood detection. It employs neural networks trained on audio recordings to predict user mood from vocal features like pitch and tone. The system further enhances recommendations through collaborative filtering based on user preferences and classifies songs by mood using base and frequency features. While promising a holistic approach, a notable limitation could be the system's ability to seamlessly integrate and balance the contributions of both facial and voice recognition modalities.
5. **SONG RECOMMENDATION BASED ON VOICE TONE ANALYSIS [5], published in the *Journal of Engineering Sciences*, Vol. 15, Issue 02, 2024,**

proposes an AI and machine learning-driven song recommendation system that addresses the gap in conventional methods by considering the user's emotional state. It employs MFCC for feature extraction from voice tone and utilizes deep learning models like Artificial Neural Networks for enhanced accuracy. The system aims to provide individualized music suggestions. A major challenge highlighted is consistently and accurately detecting the user's emotional state from speech for effective data determination in the recommendation process.

6. **Song recommendation using speech emotion recognition [6], published in the *International Journal of Health Sciences*, 6(S1), pp. 10428–10434, 2022, DOI: 10.53730/ijhs.v6nS1.7498,** proposes a song recommendation system that interprets a user's mood from their speech to suggest appropriate music. The project aims to provide personalized recommendations by analyzing spoken input, drawing parallels with recommendation systems used in e-commerce and media. It highlights the relaxing and beneficial aspects of music. While the abstract emphasizes interpreting mood from speech, it does not explicitly detail the specific machine learning algorithms or the architecture used for speech emotion recognition and song recommendation, which could be a limitation in understanding the system's technical depth.
7. **Speech based Emotion Recognition using Machine Learning[7], published in the *International Journal of Research and Analytical Reviews (IJRAR)* in 2021,** details a machine learning approach to speech emotion recognition that relies on feature extraction and classifier training. The system extracts speaker-specific features like tone, pitch, energy, and Mel-frequency cepstral coefficients (MFCCs) from audio signals, utilizing the RAVDESS dataset for training and testing, and also implements feature extraction on recorded natural speech. It aims to accurately recognize emotions such as neutral, anger, fear, and sadness. However, a potential limitation is its primary reliance on an acted speech corpus (RAVDESS), which might not fully capture the complexities and nuances of spontaneous, natural human emotions, potentially affecting its performance and generalization in real-world applications where emotional expressions are less exaggerated.
8. **Speech Emotion Recognition using CNN [8], published in the *International Research Journal of Engineering and Technology (IRJET)*, Vol. 07, Issue 06,**

June 2020, presents the implementation of speech emotion recognition using a Convolutional Neural Network (CNN) deep learning model adapted from image processing. Programmed in Python with Keras and TensorFlow, the model achieved a mean accuracy of 79.33% for five emotions (happiness, fear, sadness, neutral, anger) using English language recordings. Its contributions include the adaptation of CNN for audio processing and the development of an experimental software environment. A potential limitation is that its performance might be sensitive to the quality and diversity of the audio recordings used for training, especially given the inherent variability of real-world speech.

9. **An Emotion Based Music Recommender System Using Deep Learning [9], an MSc Research Project from the *National College of Ireland*, 2020,** proposes an emotion-based music recommender system leveraging deep learning and music therapy to mitigate negative emotions, particularly those associated with dementia (Anger, Fear, Sadness, Confusion), aiming for neutral or positive states (Joy, Trust). The system identifies targeted emotions using a Convolutional Neural Network (CNN) model, built on curated emotion-labeled audio clips converted to Mel Frequency Spectrograms (MFS), and then recommends music via content-based filtering. It achieved high classification metrics (accuracy: 95.25%, recall: 72.44%, precision: 87.42%, F1 score: 0.7922). A potential limitation is that while the system aims to *influence* emotions, its effectiveness in a therapeutic context for conditions like dementia would require extensive clinical validation beyond classification accuracy.
10. **Music Player Based on Emotion Recognition of voice signals [10], presented at the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT),** outlines a smart music system that recognizes emotions from voice speech signals to provide mood-based music. The system identifies five emotions—anger, anxiety, boredom, happiness, and sadness—by processing speech from the Berlin emotional database, extracting features, and classifying emotions using GMM (76.31% accuracy) and SVM (81.57% accuracy). Once an emotion is recognized, it automatically selects music from a stored playlist. A potential limitation could be its reliance on the Berlin emotional database, which, being

an acted dataset, might not fully generalize to spontaneous real-world speech, potentially affecting the system's accuracy in diverse, unconstrained environments.

1.12. Plan of Project Execution

Task	Start Week	End Week	Duration
Idea Discussion and Topic Finalization	Week 1	Week 2	2 Weeks
Paper Publication	Week 3	Week 5	3 Weeks
Information Gain and Research	Week 6	Week 7	2 Weeks
Data Collection	Week 7	Week 11	3 Weeks
Data Curation	Week 12	Week 13	2 Weeks
Designing Phase	Week 14	Week 15	2 Weeks
Data Preprocessing and Feature Extraction	Week 16	Week 20	4 Weeks
Model Training, Algorithm Development and Implementation	Week 21	Week 22	2 Weeks
Frontend and API Development	Week 23	Week 25	3 Weeks
Deployment and Documentation	Week 26	Week 28	3 Weeks

Table 1.1 Plan of Project Execution

2. Technical Keywords

2.1. Area of Project

1. Speech Emotion Recognition
2. Audio Signal Processing
3. Machine Learning and Deep Learning
4. Security and Privacy
5. Personalized Music System
6. Music Information Retrieval
7. Emotion Classification
8. MFCC (Mel-frequency cepstral coefficients)

2.2. Technical Keywords

1. Computing Methodologies (Artificial Intelligence, Machine Learning, Speech Recognition, Neural Networks)
2. Software Engineering (Requirements, Specifications, Design, Software Architectures, Development, Testing and Debugging, Maintenance and Enhancement)
3. Data Storage Representations/Data Encryption
4. Database Management
5. Information Storage and Retrieval (Content Analysis)
6. Information Interfaces and Presentation (User Interfaces, Audio input section)
7. Pattern Recognition (MFCC, Spectrograms)

3. Introduction

3.1. Project Idea

1. Develop a speech emotion-based song playlist recommendation system, to recognize the user’s emotions by analyzing their speech input and then suggest music that corresponds to or improves their emotional state.
2. Enhance a personalized playlist by mapping the identified emotion to a database of songs, creating more engaging and emotionally supportive music experience.

3.2. Motivation of the Project

1. Current music recommendation systems often fall short in truly understanding a user's dynamic emotional state, frequently leading to generic or mismatched suggestions.
2. Integrating Speech Emotion Recognition (SER) with deep learning provides a precise and intuitive method to interpret user mood from voice, enabling dynamic music curation.

3.3. Literature Survey

S. no.	Title and Authors	Conference/ Journal Name and Publication Year	Topic Reviewed/ Algorithms or Methodology Used	Advantages and disadvantages
1	Speech Emotion Based Music Recommendation System	International Journal of Innovative Science and Research Technology, Vol. 9, Issue 4, April 2024	System develops a speech emotion-based music recommender using audio input, MFCC/LFPC feature extraction, and classification to suggest mood-aligned playlists.	**Advantages:** It offers personalized, emotion-aware recommendations for user satisfaction with 76% accuracy. **Disadvantages:** Challenges include environmental noise impacts and inherent speech recognition limitations.
2	A Survey on Speech Emotion Based Music Recommendation System	International Journal of Scientific Research in Engineering and Management (IJSREM), Vol. 08, Issue 04, April 2024	Covers acoustic feature extraction MFCCs and emotion classification using GMMs, SVMs, and deep learning.	**Advantages**: Personalized, mood-based music recommendations to enhance user mental state and overall listening experience. **Disadvantages**: Accuracy is challenged by individual speech variations, accents, and noisy environments. Significant data privacy concerns arises.
3	Emotion Based Music Recommendation	International Journal of Research	Python and machine learning for data collection, audio	**Advantages**: Enhances user experience by providing personalized,

	System	Publication and Reviews, Vol. 4, No. 12, December 2023	feature extraction and collaborative filtering for recommendations.	mood-aligned music, increasing engagement and retention. It offers valuable insights into user preferences and emotions for content strategies. **Disadvantages**: Challenge to accurately recognize emotions from diverse inputs and effectively extracting relevant emotional features from music.
4	Music Moody - Facial Recognition And Voice Recognition To Detect Mood And Recommend Songs	International Journal of Advanced Research and Publications, Volume 6, Issue 10, October 2023	Uses neural networks, collaborative filtering, and multiclassification with base/frequency features (pitch, volume, tone) for song organization and personalized recommendations	**Advantages**: Offers highly personalized and emotionally tailored music suggestions, reducing manual search efforts. **Disadvantages**: Integrating two modalities adds significant complexity and potential for discrepancies. The system's robustness could be challenged by varying environmental conditions and subtle human expressions.
5	Song Recommendation Based On Voice Tone Analysis	Journal of Engineering Sciences, Vol. 15, Issue 02, 2024	AI/ML-based song recommendations by analyzing user emotional state from voice tone. It extracts features using MFCC.	**Advantages**: Offers personalized music suggestions appropriate for the user's current mood, with improved accuracy through deep learning. It has the potential to transform music listening. **Disadvantages**: A major challenge lies in accurately and consistently detecting the user's emotional state from speech.

6	Song recommendation using speech emotion recognition	International Journal of Health Sciences, 6(S1), pp. 10428–10434, 2022	It employs a Keras/TensorFlow Sequential model for emotion classification and utilizes MongoDB for song storage	**Advantages**: Offers personalized song suggestions based on user mood via text or speech, enhancing entertainment. Achieves good accuracy (86% training, 76% testing) in emotion classification. **Disadvantages**: Its reliance on speech-to-text conversion might limit the nuance of emotion detection from voice. The system's effectiveness is constrained by the predefined emotional categories.
7	Speech based Emotion Recognition using Machine Learning	International Journal of Research and Analytical Reviews (IJRAR) in 2021	It uses preprocessing, feature extraction (MFCC, Pitch), and a multiclass Support Vector Machine (SVM) classifier	**Advantages**: Achieved an overall accuracy of 79.4872% (around 80%) using three features, showing improvement over two features. **Disadvantages**: Reliance on an acted speech corpus (RAVDESS) might limit generalization to natural speech. Sadness classification accuracy did not improve significantly and was susceptible to misclassification as happiness.
8	Speech Emotion Recognition using CNN	International Research Journal of Engineering and Technology (IRJET), Vol. 07, Issue 06, June 2020	It utilizes the RAVDESS dataset, extracts MFCC features, and processes audio to classify five emotions (happiness, fear, sadness, neutral, anger)	**Advantages**: Achieved promising accuracy (79.33%) for emotion classification, demonstrating CNN's effectiveness for audio processing. The model can identify a range of emotions relevant for personalized applications. **Disadvantages**: RAVDESS might limit

				its generalization to natural, unconstrained speech in real-world scenarios.
9	An Emotion Based Music Recommender System Using Deep Learning	MSc Research Project from the National College of Ireland, 2020	Deep Learning (CNN) to influence negative emotions in dementia patients. It converts audio to Mel Frequency Spectrograms (MFS) for classification and uses content-based filtering for music suggestions	**Advantages**: Offers a non-intrusive, deep learning approach to influence patient mood with music therapy, achieving high classification accuracy (95.25%). **Disadvantages**: Limitations include the lack of a predefined scoring metric for audio files and model overfitting due to silent frames.
10	Music Player Based on Emotion Recognition of voice signals	2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)	It involves preprocessing feature extraction (MFCC), and classification using Gaussian Mixture Models (GMM)	**Advantages**: Achieves successful emotional classification performance, with SVM providing better accuracy (81.57%) than GMM (76.31%). It automatically selects music based on recognized emotions. **Disadvantages**: Primarily relies on an acted emotional speech database (Berlin Emo-DB), which might limit its generalization to natural, spontaneous speech.

Table 3.1 Literature Survey

4. Problem Definition and scope

4.1. Problem Statement

To develop a system that analyzes a user's real-time emotional state by taking audio input and classifying it to an emotion to provide personalized music recommendations, addressing the limitations of traditional algorithms that rely solely on past behavior and genre preferences.

4.1.1. Goals and objectives

1. Provide Emotion Driven Music Experience: Develop a system that recommends songs based on the user's real-time emotional state, enhancing the emotional connection to music and improving user engagement.
2. To Enhance Accuracy and Reliability: Improve the accuracy and reliability of the detection system through the usage of training data ensuring high performance across various scenarios and conditions.
3. Real-Time Emotion Detection from Speech: Utilize deep learning models to recognize emotions from speech input in real time, ensuring fast and accurate playlist generation tailored to the user's current mood.
4. Seamless Integration with Music Platform: Integrate the system with popular music streaming services (e.g., Spotify, Apple Music) to provide continuous, personalized music recommendations based on emotional changes.
5. Improve User Experience through Dynamic Playlist Updates: Allow dynamic updates to the playlist as the user's mood shifts, maintaining a constantly relevant and engaging music selection.

4.1.2. Statement of scope

1. Scope of the Software
   a) Size Input: Input include audio recordings of speech signals, typically 1 to 5 seconds long, to detect emotions such as happiness, sadness, anger, fear, disgust.
   b) Bounds on Input: The input speech must be clear and properly segmented. The system relies on removing silence and irrelevant noise using preprocessing techniques.

2. Input Validation

   Input validation involves ensuring that the audio data has sufficient quality and removing silence sections. Validation also includes checking that the speech duration is within the expected range.

3. Input Dependency

   The performance of the system is highly dependent on the quality of the voice signals and the accuracy of feature extraction techniques like Mel-Frequency Cepstral Coefficients (MFCC).

4. Major Inputs

   a) Audio recordings from users, typically capturing voice tones and emotions, ranging from 1 to 5 seconds.

   b) Extracted features from speech inputs using techniques like Mel-Frequency Cepstral Coefficients (MFCC), energy, pitch, and spectral features.

5. Scope Limitations

   a) The software focuses on detecting emotions from speech and does not detect emotions from facial expressions or other physiological signals.

   b) Depending on system constraints, the analysis may not be entirely in real-time, and slight delays in playlist updates may occur.

4.2. Major Constraints

1. Computational Resources: Implementing CNN-based speech emotion recognition and playlist generation requires significant computational power, especially for real time emotion detection.
2. Training Data: The system's accuracy relies heavily on the availability of diverse and high-quality emotional speech data.
3. Scalability: As the user base grows, the system must scale to handle multiple simultaneous requests. This could require cloud-based infrastructure and optimizations to support large-scale, real-time speech processing and playlist generation.
4. Regulatory Compliance: The project must adhere to data protection laws and ethical guidelines, particularly regarding user consent for captured audio.

4.3. Methodologies of Problem solving and efficiency issues

1. Rule-Based Systems (Heuristic Approach): Uses manually defined rules to map speech features (e.g. pitch, volume) or user input (e.g., mood buttons) to emotion labels. If pitch is high and intensity is high then emotion is Happy so it suggests to play upbeat playlist.
2. Machine Learning (Without Deep Learning): Use classic ML algorithms (e.g., SVM, KNN, Decision Trees, Random Forest) for classifying emotion based on features like MFCC, pitch, tempo. Once emotion is classified, songs are recommended using a mapping
3. Deep Learning-Based Approaches: CNN is used to process extracted features like MFCC, capturing patterns that help classify emotions like happiness, sadness, Disgust, etc. However, Training CNNs on high-dimensional input data like speech features can be computationally intensive.
4. Issues in Efficiency: Real-Time Performance, Noise in Speech Input, Emotion Ambiguity, Scalability of Song Mapping, Personalization vs Generalization

4.4. Outcome

1. Increased Personalization and Emotional Resonance: The system provides music suggestions that are in sync with the user's current emotional state, providing an extremely personalized experience. Since it knows how the user feels, it picks songs that are emotionally close to them, which establishes a stronger bond between the listener and the material, thus raising satisfaction and emotional welfare.
2. Enhanced User Engagement and Retention: By providing emotionally applicable playlists in the moment, the system retains users more actively involved and disposed to revisit. In contrast to static recommenders, such an emotion-based technique tracks shifting moods, leading users to perceive themselves as being recognized and understood, which greatly increases user loyalty and long-term platform use.

4.5. Applications

1. Personalized Music Streaming: Music streaming platforms like Spotify, Apple Music, and YouTube Music can use CNN-based emotion detection from the speech to personalize playlists according to the user’s real time emotion.

2. Mental Health Support: Emotion-aware music recommendation systems can be applied in mental health and therapy settings. Music has been proven to help manage emotional stress, anxiety, and depression. This usage of Speech emotion detection can uplift the mood of the patient to overcome the sufferings.
3. Voice-controlled Virtual Assistants: Virtual assistants like Amazon Alexa, Google Assistant, and Siri can incorporate speech emotion detection to provide emotionally appropriate music recommendations.

4.6. Hardware Resources Required

Sr. No.	Parameter	Minimum Requirement	Justification
1	CPU Speed	8-core (3.0 GHz or higher)	A powerful multi-core CPU is required to handle the intensive computations. More cores and higher speeds improve overall performance and response time.
2	RAM	16 GB	A minimum of 16 GB of system RAM is required to manage system operations. It ensures the system can handle loading large datasets, and prevent memory bottlenecks.
3	Storage	512 GB SSD (NVMe preferred)	SSD storage ensures faster read/write speeds when loading and saving large model weights, datasets, and logs. NVMe drives are preferred for their superior data transfer rates, which reduce latency in model loading.
4	GPU	NVIDIA RTX 3050/3060 (6 GB+ VRAM)	Required for deep learning tasks like emotion recognition using models such as Wav2Vec or CNNs. Speeds up inference and training time significantly.

Table 4.1 Hardware Requirements

4.7. Software Resources Required

1. Operating System: Ubuntu 20.04+ / Windows 10+, Linux-based systems (Ubuntu) are preferred due to better support for development tools, libraries, and optimized performance for machine learning frameworks. Windows is also supported for ease of use.

2. IDE: VS Code / PyCharm / Jupyter, Visual Studio Code are preferred for their rich support for Python, machine learning libraries, and integrated debugging features. Jupyter notebooks are recommended for model experimentation and iterative development.

3. Programming Language: Python (3.8 or higher), Python is the primary programming language due to its extensive support for machine learning frameworks like TensorFlow, PyTorch, Hugging Face, and libraries for natural language processing and computer vision.

5. Project Plan

5.1. Project Estimates

This section provides an overview of the estimated costs and time required for the successful completion of the Project, ensuring proper resource allocation and project management.

5.1.1. Reconciled Estimates

Reconciled estimates consolidate the projected costs and time for the project, taking into consideration any adjustments based on team discussions, resource availability, and potential risks.

a. Cost Estimate

i. Cloud Services: Currently using the free tier for Firebase and financial news APIs, which incurs no immediate costs.

ii. Training/Fine-Tuning Machine Learning Models: Estimated costs associated with training and fine-tuning machine learning models necessary for expense categorization, risk assessment, and investment suggestions.

iii. Hosting Costs: Estimated costs for hosting the trained machine learning models on a cloud service to ensure scalability and accessibility.

iv. Total Cost Estimate: Sum of the above categories.

b. Time Estimates

i. Project Planning and Research

- Duration: 2 months (Weeks 1-6)
- Activities: Idea Discussion, Topic Finalization, Information Gain, and Research

ii. Development

- Duration: 4-5 months (Weeks 7-25)
- Activities: Data Collection, Data Curation, Designing Phase, Data Preprocessing, Feature Extraction, Model Training, Algorithm Development, Frontend and API Development

iii. Testing and Quality Assurance

- Duration: 1 month (Week 26)
- Activities: Deployment and Documentation

iv. Deployment

- Duration: 1 month (Week 27)
- Activities: Final Deployment and Documentation

v. Feedback and Iteration

- Duration: 1 month (Week 28)
- Activities: Gathering Feedback, Making Iterations

vi. Total Time Estimate: 8 months

5.1.2. Project Resources

The project utilizes various resources across hardware, software, and tools to support memory sharing, inter-process communication (IPC), and concurrency, managed by a team of four people.

1. People
   - Project Manager: Manages timelines, deliverables, and progress.
   - AI/ML Engineer: Develops and fine-tunes AI models for code analysis and hardware diagnostics.
   - Software Developer: Implements the application (frontend and backend) and handles system architecture.
   - Tester: Validates the system through tests for accuracy and performance.
2. Hardware
   - CPU: 8-core (3.0 GHz or higher) for multitasking.
   - GPU: NVIDIA RTX 3050 / RTX 3060 or higher (6 GB+ VRAM)
   - RAM: 16 GB to manage large datasets and concurrency.
   - Storage: 512 GB SSD for fast model loading.
3. Software
   - Operating System: Ubuntu 20.04+ / Windows 10+ for concurrency support.
   - IDE: VS Code, PyCharm, or Jupyter.
   - Programming Language: Python 3.8+ for AI/ML and JavaScript for frontend.
4. Tools
   - Version Control: GitHub for collaboration.

- Concurrency Tools: Python libraries like tensor-flow and multiprocessing for concurrent task management.

5.2. Risk Management with respect to NP Hard analysis

5.2.1. Risk Identification

1) Have top software and customer managers formally committed to support the project?

   Answer: Yes, top management has been engaged from the outset, providing formal commitments and resources necessary for project success. Regular meetings and updates ensure ongoing support and alignment with project goals.

2) Are end-users enthusiastically committed to the project and the system/product to be built?

   Answer: The end-users have expressed interest in the project, but their commitment varies. To Enhance enthusiasm, regular workshops and feedback sessions are conducted to involve them in the development process.

3) Are requirements fully understood by the software engineering team and its customers?

   Answer: The initial requirements have been discussed; however, further clarification is needed. The team conducted additional workshops with stakeholders to ensure a comprehensive understanding of all requirements.

4) Have customers been involved fully in the definition of requirements?

   Answer: Customers have been consulted during the requirements gathering phase, but there is room for improvement.

5) Do end-users have realistic expectations?

   Answer: Some end-users have unrealistic expectations regarding the timeline and capabilities of the system. Clear communication about project timelines and deliverables are established to manage these expectations effectively.

6) Does the software engineering team have the right mix of skills?

   Answer: The current team has a good foundation of skills, but additional expertise in blockchain technology and machine learning is beneficial. A skills assessment is conducted, and necessary training or hiring is implemented.

7) Are project requirements stable?

   Answer: The requirements are currently subject to some changes due to evolving user needs and technology considerations. A change management process is established to address and control scope changes efficiently.

8) Is the number of people on the project team adequate to do the job?

   Answer: The current team size is manageable, but additional resources may be needed as the project progresses, especially during critical phases of development and testing.

5.2.2. Risk Analysis

In the context of developing a system for detecting and classifying the users mood based on the speech input of user, risk analysis is crucial, particularly given the NP-hard nature of related computational problems. NP-hard problems involve complex decision-making processes that can significantly impact system performance and reliability. The challenges include the potential for high false positive and negative rates in emotion classification, which could lead to wrong songs recommendation to the user. A thorough risk analysis helps identify vulnerabilities and informs the development of robust solutions to mitigate these risks.

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Commitment of managers unclear, leading to delays.	Low	High	High	High
2	Lack Of end-user enthusiasm resulting in low adoption.	Medium	High	Medium	Medium
3	Incomplete understanding of requirements by the team.	Medium	High	High	High
4	Inadequate customer involvement during requirement definition.	Medium	High	High	High
5	Unrealistic expectations from end-users.	Low	Medium	Medium	Medium
6	Insufficient skill mixes in the engineering team.	Medium	High	High	High
7	Frequent changes in project requirements.	High	High	Medium	High
8	Insufficient team size to handle project workload.	Medium	Medium	High	Medium
9	Disagreement on project importance among stakeholders.	Medium	Medium	High	Medium

Table 5.1 Risk Table

Probability	Value	Description
High	Probability of occurrence is >75%	Indicates a significant risk that is likely to happen and could severely impact the project
Medium	Probability of occurrence is 26 - 75%	Indicates a moderate risk that may occur, requiring monitoring and possible mitigation.
Low	Probability of occurrence is <25%	Indicates a minimal risk that is unlikely to happen and may not require immediate action.

Table 5.2 Risk Probability definitions

5.2.3. Overview of Risk Mitigation, Monitoring, Management

This section outlines strategies for mitigating, monitoring, and managing each identified risk in the project. Effective risk management involves proactive planning to minimize potential impacts on the project's schedule, quality, and overall success.

5.3. Project Schedule

5.3.1. Project task set

Task ID	Task Description	Impact	Value	Description
1	Idea Discussion and Topic Finalization	Very High	> 10%	Schedule impact or unacceptable quality if requirements are unclear.
2	Paper Publication	High	5-10%	Potential schedule impact or some parts may have low quality if not aligned.
3	Information Gain and Research	Very High	> 10%	Major impact on schedule or unacceptable quality if model is poorly developed.
4	Data Collection	High	5-10%	Schedule impact or low quality in certain features if not implemented correctly.
5	Data Curation	Very High	> 10%	Critical for overall quality; major impact if not performed adequately.
6	Data Preprocessing and Feature Extraction	Medium	5%	Minimal impact on schedule; essential for user adoption but barely noticeable degradation in quality if missed.
7	Designing Phase	High	5-10%	Can lead to schedule impact or some quality issues if not executed properly.
8	Model Development	Very High	> 10%	Major impact on schedule or unacceptable quality if model is poorly developed.
9	Project Review and Closure	Low	<5%	Impact on schedule is minimal; ensures project objectives are met, and quality checks are confirmed.

Table 5.3.2 Risk Impact definitions

Risk ID	1
Risk Description	Commitment of managers unclear, leading to delays.
Category	Management
Source	Project Scope Document
Probability Low	Low
Impact	High
Response	Mitigate
Strategy	Schedule regular meetings with management to ensure alignment.
Risk Status	Occurred

Table 5.3.3 Risk A

Risk ID	2
Risk Description	Lack of end-user enthusiasm resulting in low adoption
Category	Operational
Source	User Engagement Surveys
Probability Low	Medium
Impact	High
Response	Mitigate
Strategy	Involve end-users early in the development process for feedback.
Risk Status	Identified

Table 5.3.4 Risk B

Risk ID	3
Risk Description	Incomplete understanding of requirements by the team
Category	Technical
Source	Requirements Specification Document
Probability Low	Medium
Impact	High
Response	Mitigate
Strategy	Conduct requirement workshops to clarify expectations
Risk Status	Identified

Table 5.3.5 Risk C

Task ID	Task Description	Duration	Assigned To	Status
Task 1	Requirement Gathering	2 weeks	Mayank Adhav	Completed
Task 2	Design System Architecture	3 weeks	Pratik Hagawane, Mayank Adhav	Completed
Task 3	Model Development	4 weeks	Aaditya Meher	Completed
Task 4	Application Development	5 weeks	Jeet Harne	Completed
Task 5	Testing and Quality Assurance	3 weeks	Aaditya Meher, Jeet Harne	In Progress
Task6	Documentation	1 week	Pratik Hagawane, Mayank Adhav	Completed

Table 5.3.6 Task Assigned

5.4 Description of Tasks:

1) Task 1 - Requirement Gathering: Collect and document the project requirements from stakeholders to ensure clarity and alignment.
2) Task 2 - Design System Architecture: Create a comprehensive architecture design that outlines the system's components and their interactions.
3) Task 3 - Model Development : Develop machine learning models that powers the functionality of the project.
4) Task 4 - Application Development: Build the application based on the design specifications and integrate the developed models.
5) Task 5 - Testing and Quality Assurance: Conduct thorough testing of the application to ensure it meets quality standards and functions as intended.

5.5 Task network

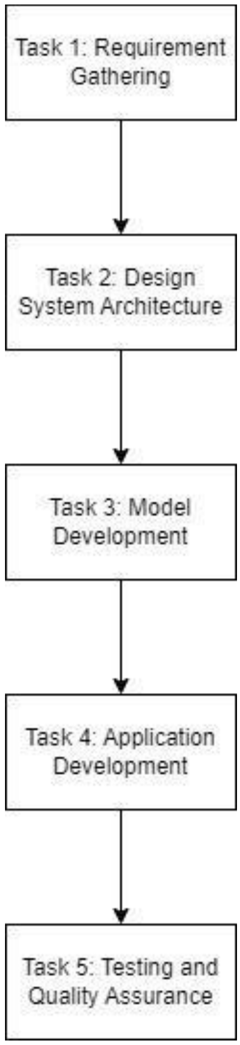


Fig. 5.5.1 Task Network

5.6 Timeline Chart

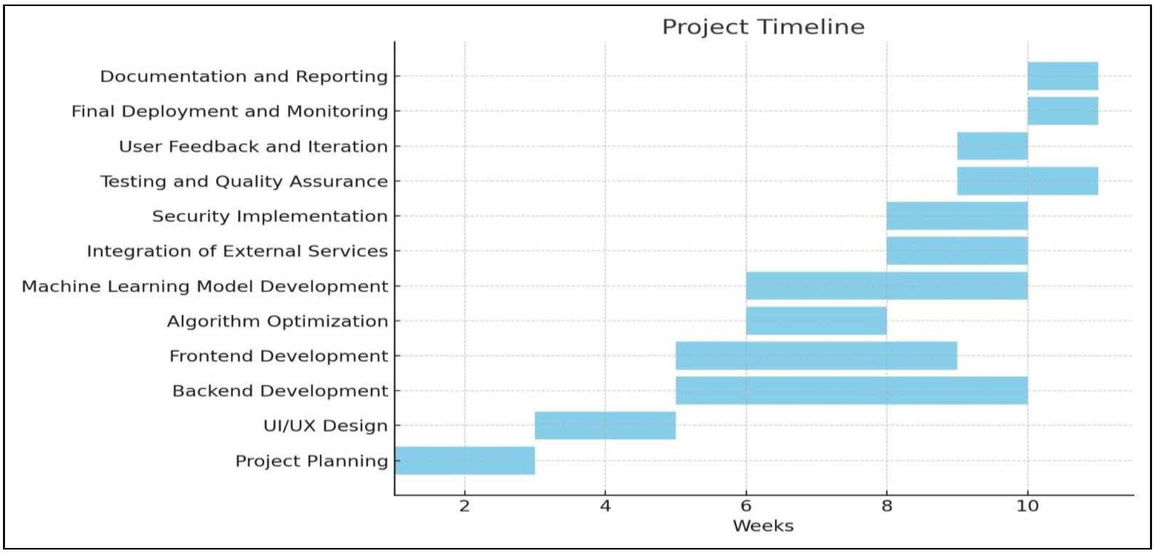


Fig. 5.6.1 Gantt Chart

5.7 Team Organization

Our strategy is to divide the tasks equally amongst the four of us. We decided a deadline for each task. In the end we combine the results of individuals into one single outcome.

5.7.1 Team structure

1) Mayank Adhav	B400080500
2) Aaditya Meher	B400080501
3) Jeet Harne	B400080469
4) Pratik Hagawane	B400080468

5.7.2 Management reporting and communication

Dedicated Team of four developers working together and taking ownership of different actions to be performed with project progress is the attitude followed by following ways:

- Daily Scrums
- Inter team communications via virtual platforms
- GitHub Collaboration
- Guidance from Mentors on monthly basis.

6 Software requirement specification

6.7 Introduction

6.7.1 Purpose and Scope of Document

The purpose of this Software Requirement Specification (SRS) is to define the functional and non-functional requirements for the development of the "Speech Emotion based Song Playlist Recommendation System using Deep Learning" system. This document outlines the project objectives, system architecture, user roles, and interactions, providing a comprehensive overview of the system's requirements. The scope includes the detection of emotion of user using machine learning (ML), enhancing the reliability of users for the emotion driven music listening.

6.7.2 Overview of responsibilities of Developer

The developer's responsibilities include:

1. Designing system architecture: Create detailed designs of the system architecture, user interfaces, and model integration
2. Model Development: Develop and train machine learning models to facilitate the functionality, ensuring they meet performance metrics.
3. Implementing real-time processing: Ensure the system processes audio input in real-time for immediate classification.
4. Testing and evaluation: Conduct unit, integration, and system testing to identify and rectify defects, ensuring the system functions as intended.
5. Ensuring system reliability and security: Enhance user trust by providing secure and accurate emotion classification.

6.8 Usage Scenario

The system can be applied in the following scenarios:

1. Personalized Music Streaming: Music streaming platforms like Spotify, Apple Music, and YouTube Music can use CNN-based emotion detection.
2. Mental Health Support: Emotion-aware music recommendation systems can be applied in mental health and therapy settings.

3. Voice-controlled Virtual Assistants: Virtual assistants like Amazon Alexa, Google Assistant, and Siri can incorporate speech emotion detection to provide emotionally appropriate music recommendations.

6.8.1 User profiles

1. Admin: Responsible for managing and configuring system settings, monitoring system performance, and handling data storage.
2. End-user: Uses the system for listening the music based on their mood by giving the audio input to the system. Their role primarily involves being the subject and use the system.
3. Developer: Manages the development and maintenance of the system, ensuring it functions efficiently across various platforms.

6.8.2 Use-cases

The use cases describe the interactions between the system and actors.

Sr No	Use Case	Description	Actors	Assumptions
1	Audio Input Capture	The system captures real-time audio from the user.	End-user	User has enabled the microphones.
2	Analyze Emotional Trends	The system analyzes patterns of user emotional states and song preferences over time to improve the system's recommendations.	Developer	Emotion analysis is successful.
3	Feature Extraction	Extracts key features from the captured audio to classify it.	System	Preprocessing is completed.
4	Classification	The system classifies the audio into one of the predefined emotions category.	System	Feature extraction is successful.

5	Result Display	Displays the result of the classification, giving the relevant song playlist.	Admin, End-user	Classification is successful.

Table 6.1 Use Cases

‘

6.8.3 Use Case View

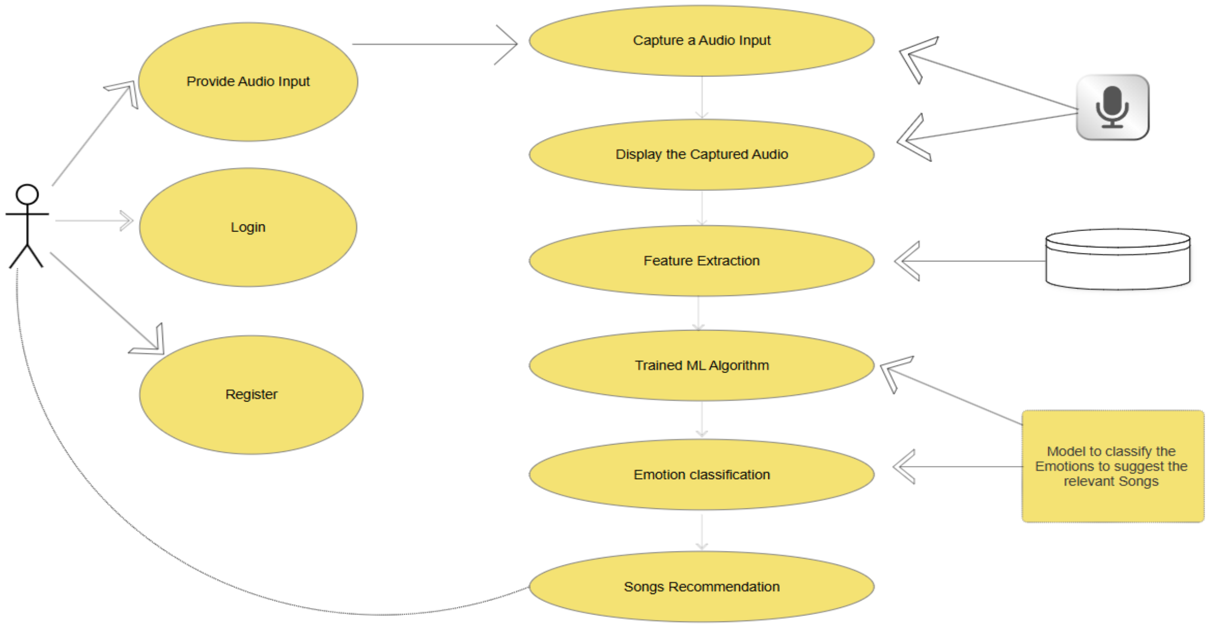


Fig. 6.1 Use case diagram

6.9 Data Model and Description

6.9.1 Data Description

In this section, we describe the data objects that the system manages and manipulate. The primary goal of the system is to process audio inputs and extract relevant features to classify the emotions. The database entities includes various types of speech in different emotions and extracted feature vectors. The major data objects in this system include:

1. User Voice Input:
   i. This data object stores the raw voice input provided by the user for emotion recognition.

ii. Attributes:
   a. Voice Input ID (Primary Key).
   b. Audio File Path.
   c. Timestamp.
   d. User ID (Foreign Key).

2. Pre-Processed Voice Data:
   i. This object stores the features extracted from the user's voice input after pre-processing (e.g., silence removal, normalization).
   ii. Attributes:
      a. Processed Data ID(Primary Key).
      b. Voice Input ID (Foreign Key).
      c. MFCC Features
3. Emotion Recognition Output:
   i. Description The recognized emotional state generated from the pre-processed voice data using classifiers like CNN.
   ii. Attributes:
      a. Emotional Output ID.
      b. Recognized Emotion (Happiness, sadness, surprise, disgust, neutral etc).
      c. Confidence Score.
      d. Processed Data ID (Foreign Key).
4. Song Playlist:
   i. This object holds information about the songs in the system's database, including genre, mood, and their association with specific emotions.
   ii. Attributes:
      a. Song ID (Primary Key).
      b. Song Title.
      c. Song Genre.
      d. Associated Emotions.

6.9.2 Data objects and Relationships

The relationships between the data objects are illustrated using an Entity-Relationship Diagram (ERD) format. The following entities and relationships are key to the system's data flow:

i. User Voice Input has a one-to-one relationship with Pre-processed Voice Data, where each voice input is processed to extract features.
ii. Pre-processed Voice Data is connected to Emotion Recognition Output via a one-to-one relationship, as each processed voice data generates one emotion classification.
iii. Emotion Recognition Output is related to Recommended Playlist via a one-to-one relationship, as one emotional state results in one playlist recommendation.
iv. Song Playlist has a many-to-one relationship with Emotion Recognition Output, as multiple songs are associated with a specific recognized emotion.

Relationships:

- User Voice Input Data → Extracts → Feature Data
- User Data → Logs → Authentication Log
- Feature Data → Classifies→ Playlist Generation Process

6.10 Functional Model and Description

The system involves multiple software functions that work together to process input audio, extract features, classify them, and suggest songs. The major software functions and their data flows are detailed below:

1) Input Audio Processing: This function takes the raw input Audio, normalizes it, and make it ready for feature extraction. It is responsible for receiving user input and ensuring compatibility with the system's requirements.
2) Feature Extraction: The key function of this system, where features are extracted using convolutional layers. The system extracts spectral details, tonality patterns, pitch of the voice, and other relevant attributes. This step results in feature vectors that are fed into the classification function.
3) Classification: The extracted feature vectors are passed through a trained classification model (e.g., CNN). Based on the feature data, this function decides whether the mood is happy, sad, disgust, surprised etc.

4) System Recommendation for Songs: Based on the emotion detected the best suitable combination of songs is created to uplift the mood of the user and help user to use system as personalized music system.

6.10.1 Data Flow Diagram

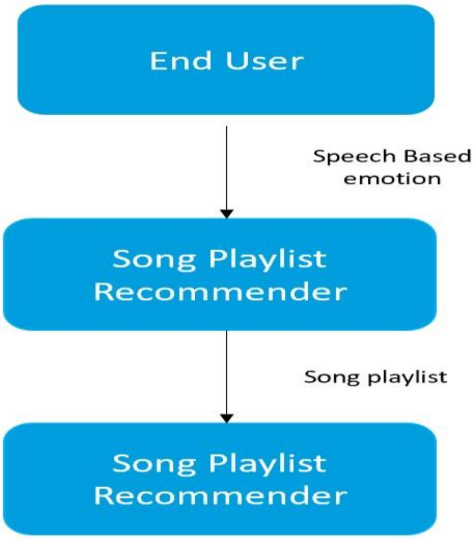


Fig. 6.2 Data Flow Diagram

6.10.2 Activity Diagram:

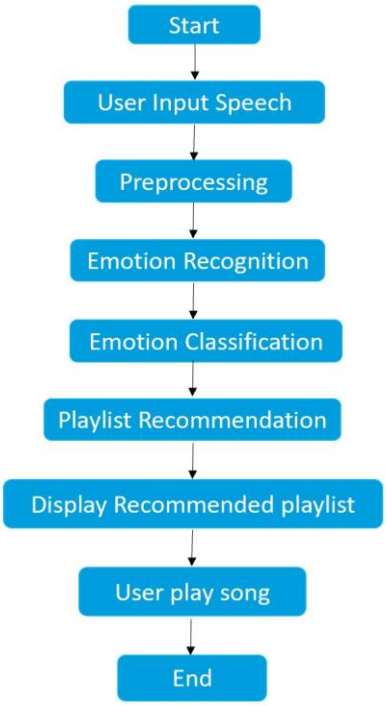


Fig. 6.3 Activity Diagram

6.10.3 Non-Functional Requirements:

1. Interface Requirements:
    a. The system should provide a clean and intuitive user interface.
    b. Allow users to input voice queries and view the song playlist recommendations easily.
    c. The system must give clear visuals without requiring any technical expertise.
2. Performance Requirements:
    a. Accuracy: The system should classify emotions with at least 95% accuracy in real-time.
    b. Latency: The system must be capable of processing the user's voice input, detecting emotion, and generating a playlist recommendation within 2 seconds for 90% of the requests.

6.10.4 State Transition Diagram:

Fig.6.4 example shows the state transition diagram of system components. The transitions from one state to the other are represented by arrows. The Figure shows important states and events that occur while creating a new project.

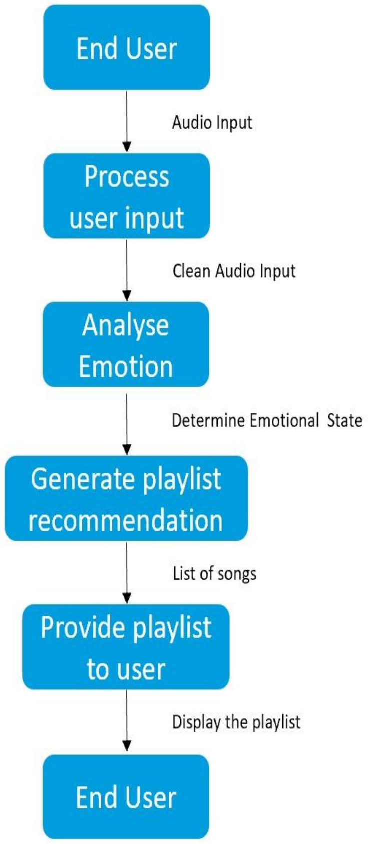


Fig. 6.4 State transition diagram

6.10.5 Design Constraints

The design of the Emotion detection system has several constraints that impact the system's subsystems. These constraints include:

1. Real-Time Processing: The system must process and classify audio emotion in real-time, requiring optimized algorithms and hardware to meet performance goals. This limits the complexity of the models and forces the use of efficient architectures like CNN with minimal layers.
2. Hardware Limitations: The system is constrained by the available hardware (CPU and memory). If the hardware is not powerful enough, model optimization may be required to maintain performance standards.
3. Scalability: The system needs to support a large number of users simultaneously. To achieve this, the architecture must be scalable, and models must be able to handle concurrent requests without performance degradation.

4. Accuracy vs. Latency Tradeoff: The system's design must balance classification accuracy and processing speed. High-accuracy models might introduce latency, whereas fast models may sacrifice accuracy. This tradeoff influences the choice of algorithms and hardware accelerators.

6.10.6 Software Interface Description

The software interface in the Speech Emotion detection system involve connections with external systems, users, and devices. These interfaces are critical for the system's integration into a larger ecosystem.

1. User Interface (UI):
   a) Input: The UI allows users to upload audio input. It should be provided with 5 to 6 seconds time period.
   b) Output: After classification, the UI displays the playlist of song according to the relevant emotion detected.
   c) Requirements: The UI must be intuitive and responsive to ensure a seamless user experience. Response should be provided in real-time.
2. External API Interface:

   Integration with Other Systems: The system's backend API (exposed via Flask) must facilitate seamless interaction with external services, particularly a third-party music streaming platform (e.g., Spotify) to fetch and recommend playlists. It also provides endpoints for client-side applications to handle user authentication and emotion prediction requests.
3. Database Interface:
   a) Input: The system reads and write data (e.g., authentication logs, user audio data, and classification results) to a connected database.
   b) Output: All classification outcomes and user interaction logs must be stored in the database for future auditing and analysis.
   c) Requirements: The database must support fast read/write operations, encryption for sensitive data, and be highly available to handle real-time queries.

7 Detailed Design Document using Appendix A and B

7.7 Introduction

This Detailed Design Document outlines the technical and architectural framework for the system that detects whether the audio input given by user is happy, sad, neutral, surprised, etc. The system captures a short audio with users awareness and analyzes it to make this distinction. Key elements such as audio processing, detection and classification in one of the emotions in real-time to ensure the system's efficiency.

7.8 Architectural Design

The architecture design focuses on a system for detecting the Emotions of the user. The system uses a combination of feature extraction algorithms, MFCCs, and machine learning models to make decisions based on the input data. System is composed of three main components: the User Interface, Emotion Detection Module, and Playlist Generator. These components are integrated to ensure a seamless experience from speech input to playlist recommendation.

1. User Interface (UI): The user interface receives inputs in the form of an audio, which can represent various emotions of the user. It acts as the interaction point for users and collects input data to be processed by the system.
2. Feature Extraction Algorithm: The central part of the system is the feature extraction module, which processes the input data. It uses a combination of techniques to extract significant features from the audio.
3. Playlist Generator: Based on the detected emotion, this component selects an appropriate playlist from a predefined music library. It maps emotions to genres or songs that are likely to resonate with the user's current mood.

7.9 Data Design using Appendices A and B

7.3.1 Internal Software data structure

1. Audio Data: A structure representing the speech input from the user, containing fields such as audio-id, audio , sample-rate, and time-stamp.

2. Emotion Data: A structure that holds the emotion detection results, including fields for emotion type(e.g., happy, sad, angry), confidence score, and detection-time.
3. Playlist Data: Contains the list of songs generated by the Playlist Generator, with fields like playlist-id, song list, and genre.

7.3.2 Global data structure

1. User Profile Data: Stores user-specific information, such as user-id, preferred-genres, listening-history, and default-emotion-playlist-map.
2. Music Library: A global data repository that holds song metadata, including song-id, title, artist, genre, and emotion-association.

7.3.3 Temporary data structure

1. Session Data: Stores information for each user session, including session-id, current-emotion, and playlist-recommendations. This data is cleared when the session ends.
2. Log Data: Temporary logs generated during emotion detection and playlist generation, such as detection-logs and recommendation-logs, which are cleared after each process.

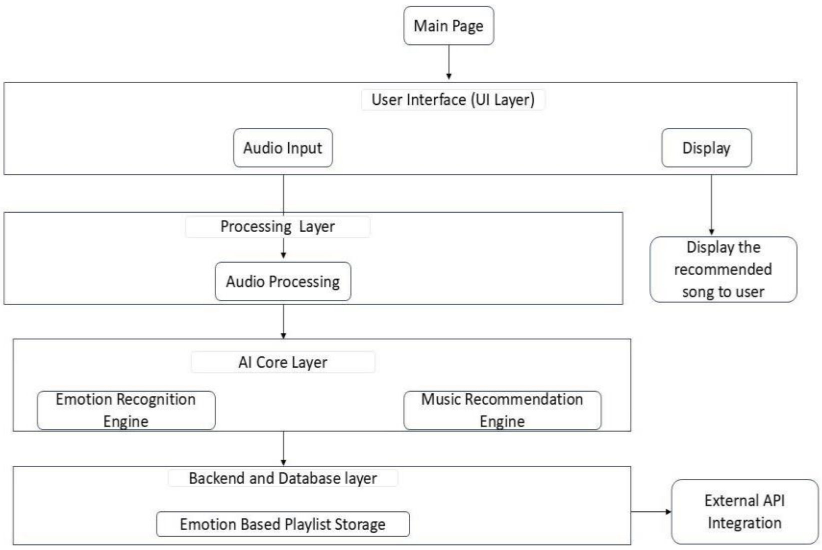


Fig. 7.1 Architecture diagram

7.4 Component Design

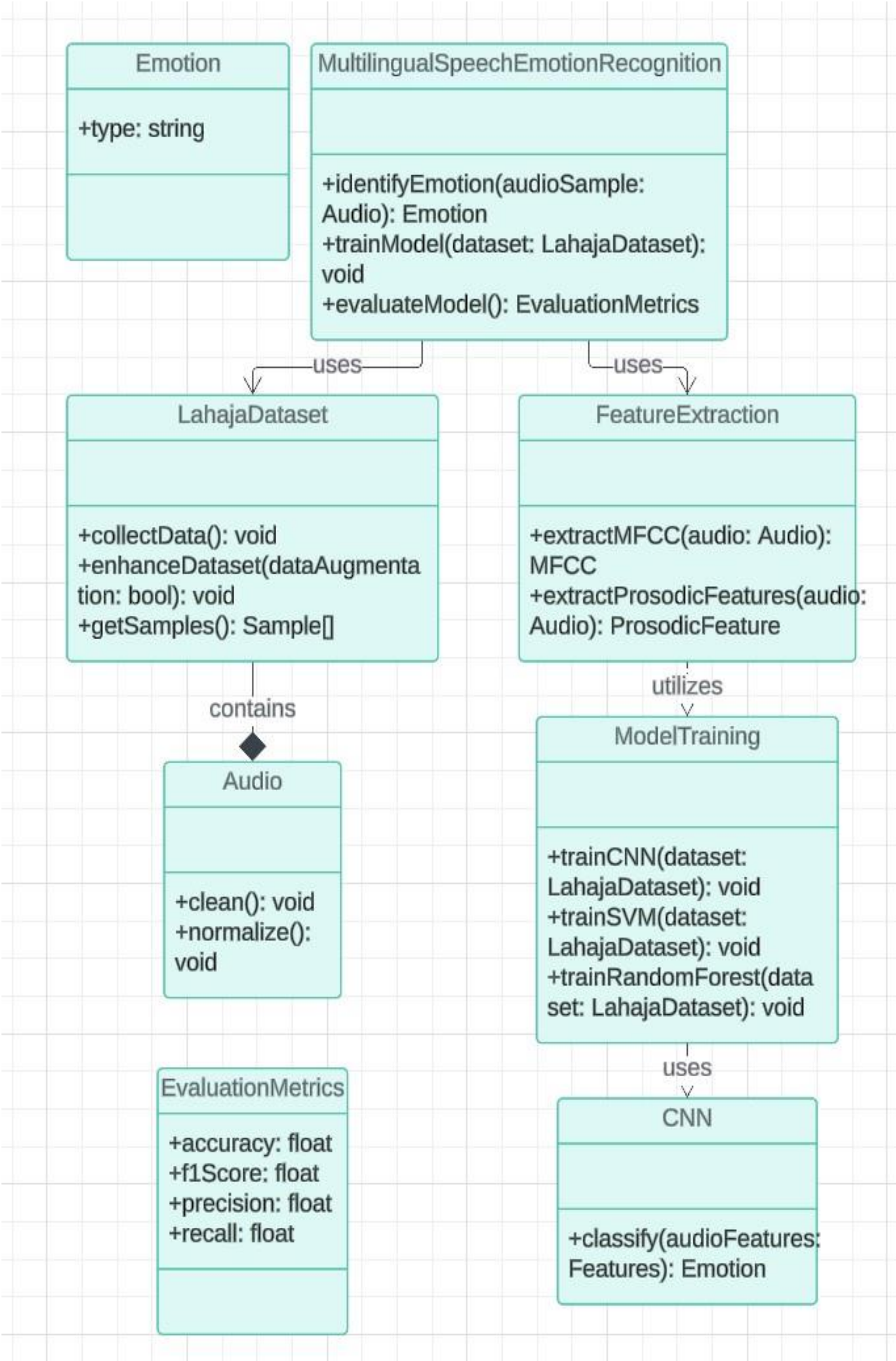


Fig. 7.2 Class Diagram

8 Project Implementation

8.1 Introduction

The implementation phase focuses on developing a system of the Speech Emotion-Based Song Playlist Recommendation system to focus on leveraging audio signal processing and machine learning techniques to recognize the emotions in a user's voice and recommend an appropriate playlist. This section outlines the tools, technologies, and methodologies used to build the system, from emotion detection to song recommendation. The system uses a modular approach with components for the detection of emotions. The system aims to deliver accurate, real-time results while ensuring efficiency and seamless user interaction.

8.2 Tools and Technologies Used

1) Programming Languages:
   - Python: Used for core logic implementation and data processing tasks such as audio analysis.
   - JavaScript: For front-end development, enabling user interaction and a responsive interface.
2) Technologies:
   - PyTorch: For handling deep learning tasks like feature extraction and model training to detect the emotions.
   - Deep Learning Models (e.g., CNNs): For classifying emotional tones and matching them to music moods.
3) Other Tools:
   - Git/GitHub: For version control, enabling team collaboration and code management.

8.3 Methodologies/Algorithm Details

The project employs several algorithms to process audio inputs, recognize emotions, and generate song recommendations based on the detected emotions.

:

8.3.1 Algorithm 1 / Pseudo Code: Emotion Recognition from Speech

1) Objective: Detects the emotion expressed in the user's voice input by analyzing audio signals and extracting relevant features.

2) Input: Captured audio from the Microphone.
3) Output: Detected Emotion Label.
4) Process:
   i. Initialize the Captured Audio.
   ii. Preprocess the Voice Audio Signal.
   iii. Perform noise reduction and normalize audio.
   iv. Extract audio features (MFCC, Spectrogram, etc.) using Librosa.
   v. Pass the extracted features to the pre-trained emotion recognition model.
   vi. Model outputs a probability distribution over emotion classes.
   vii. Choose the emotion with the highest probability as the detected emotion.
   viii. Return Detected Emotion Label.

8.3.2 Algorithm 2 / Pseudo Code: Playlist Recommendation

1) Objective: Once the emotion is recognized, this algorithm recommends a song playlist that aligns with the user's emotional state.
2) Input: Detected Emotion Label
3) Output: Recommended Song Playlist
4) Process:
   i. Load the pre-trained CNN model.
   ii. Retrieve Songs by Emotion by Query the song database with the detected emotion as the filter.
   iii. Retrieve songs with matching mood tags or genres.
   iv. Sort the songs based on user preferences and generate playlist.
   v. Compile top-ranked songs into a playlist.
   vi. Return the Recommended Song Playlist.

8.4 Verification and Validation for Acceptance

1. Verification

   Verification ensures the system is being developed correctly and according to design specifications. The following verification activities were conducted:

   a. Code Review and Static Analysis

All modules (audio capture, CNN model) were reviewed for coding standards, correctness, and maintainability.

b. Unit Testing

Each component (e.g., emotion detection, CNN inference) was individually tested using dummy inputs and boundary cases to verify accurate functionality.

c. Integration Testing

Verified the interaction between modules such as audio capture, emotion detection, and models prediction to ensure data consistency and control flow.

d. System Compatibility Check

The application was tested to verify consistent deployment and execution.

2. Validation

Validation confirms that the system meets the end-user requirements and real-world expectations. The following validations were performed:

a. Live vs Pre-recorded Audio Input Validation

The system was thoroughly tested by using both spontaneous live speech input directly recorded from the user's microphone and a diversified set of pre-recorded audio samples.

b. Robustness against Varied Voices and Accents

The baseline emotion classification model was rigorously tested on a broad variety of speech samples from numerous speakers, with a deliberate attempt to include varied vocal tones, speaking patterns, and linguistic accents.

c. Non-Speech/Noise Input Discrimination

The system was tested with different types of non-emotional or irrelevant sound inputs, such as chunks of pure silence, background noise, and sounds without human speech.

d. Performance Evaluation

Key performance indicators were accurately gauged, such as audio processing latency, emotion classification speed by the deep learning algorithm, and overall responsiveness of the playlist creation process.

e. User Acceptance Testing (UAT)

Exhaustive demonstrations and hands-on sessions were done with a representative group of test users. They interacted with the web-based user interface of the system, giving voice input and checking the resulting playlists.

9 Software Testing

9.7 Type of Testing Used

1) Unit Testing:
    a. Objective: Test individual functions, components, and modules in isolation.
    b. Tools Used: PyTest (for Python), Unit(for testing emotion classification).
    c. Example: Testing the function that processes and classifies emotions.
2) Integration Testing:
    a. Objective: Ensure that different modules (e.g., Speech Emotion Recognition, Playlist Generator) work together as expected.
    b. Tools Used: Postman (for API testing).
    c. Example: Verifying that the emotion recognition module correctly passes the identified emotion to the playlist generator to create an appropriate song list.
3) System Testing:
    a. Objective: Test the entire system as a whole, simulating real-world usage scenarios.
    b. Tools Used: Manual testing for real-world scenarios along with automation tool if applicable.
    c. Example: Ensuring that the system can process various speech inputs, classify emotions accurately, and provide relevant song recommendations based on those emotions.
4) Performance Testing:
    a. Objective: Assess the system's performance, especially in handling large audio files or multiple concurrent requests.

b. Example: Evaluating the system's response time under heavy load.

5) User Acceptance Testing (UAT):
   a. Objective: Ensure that the system meets user expectations and requirements.
   b. Involves Real-world users testing the system by providing voice inputs and checking if the recommended playlist matches their emotional state.
   c. Feedback was gathered to improve the accuracy of emotion detection and playlist relevance.

9.8 Test Cases and Test Results

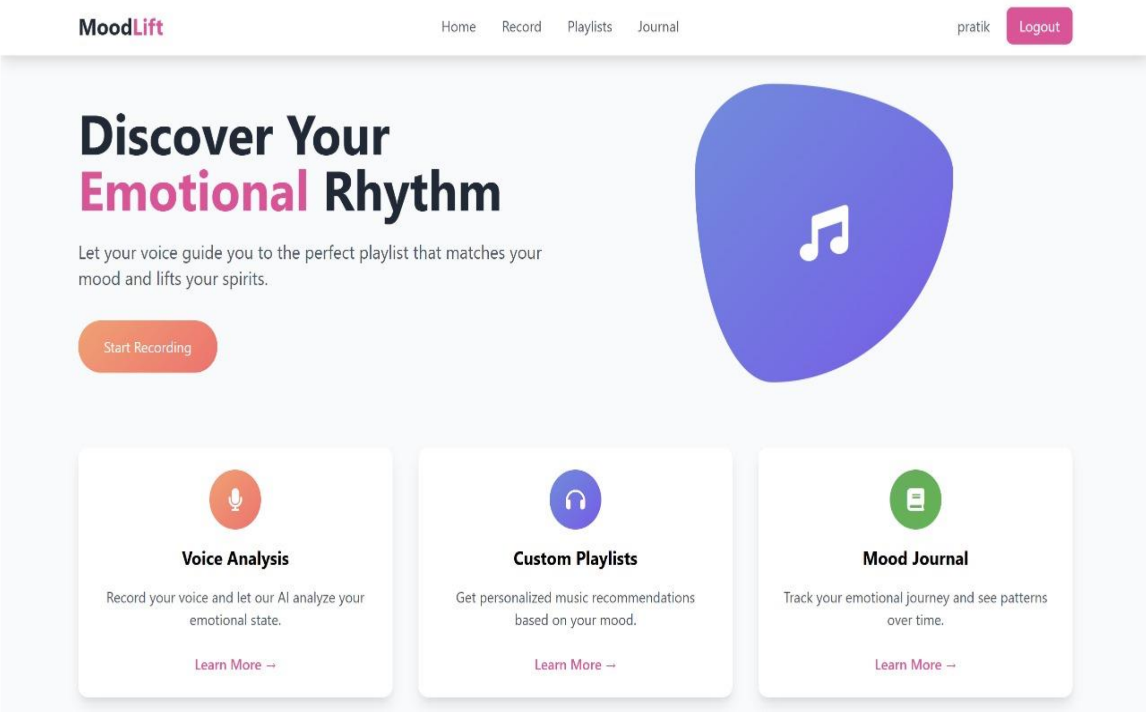
Test Case No.	Test Description	Expected Output	Actual Output	Result
TC001	Test playlist generator function for "happy" emotion	Returns a playlist with upbeat songs	Upbeat playlist generated	Pass
TC002	Test playlist generator function for "sad" emotion	Returns a playlist with mellow songs	Mellow playlist generated	Pass
TC003	Test playlist generator function for "angry" emotion	Returns a playlist with intense songs	Neutral playlist generated	Fail
TC004	Verify emotion recognition to playlist generator integration.	Emotion identified and passed correctly to playlist generator	Emotion Passed correctly	Pass
TC005	Test integration between voice input and emotion classification.	Voice input processed and emotion identified	Emotion identified accurately	Pass
TC006	Test system response with end-to-end flow	Accurate playlist generated based on identified emotion	Playlist generated expected	Pass
TC007	Ensure system handles 50 simultaneous voice inputs	All queries are processed successfully	Some are not processed	Fail

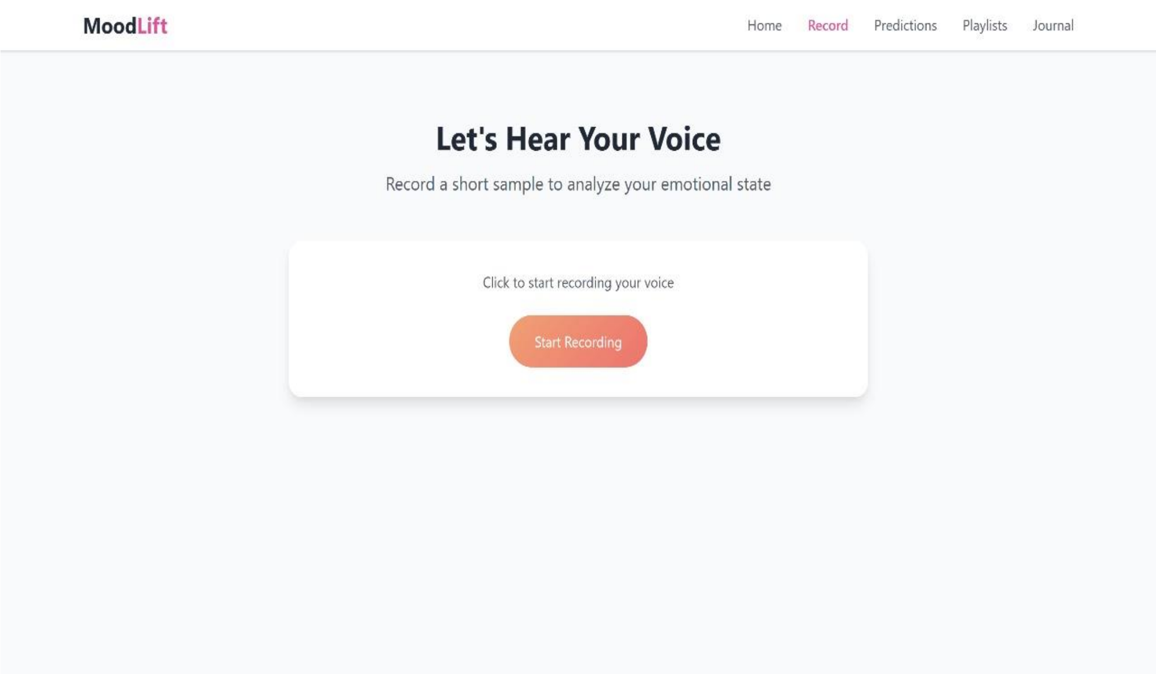
TC008	Ensures system maintain accuracy under load	Correct emotion and playlist generated	Correctly generated	Pass
TC009	Test response time for complex or lengthy speech input	Response time < 5 seconds	4 seconds	Pass
TC010	Test system memory usage with audio processing	Memory usage stays below threshold	Within limits	Pass

Table 9.1 Test Cases

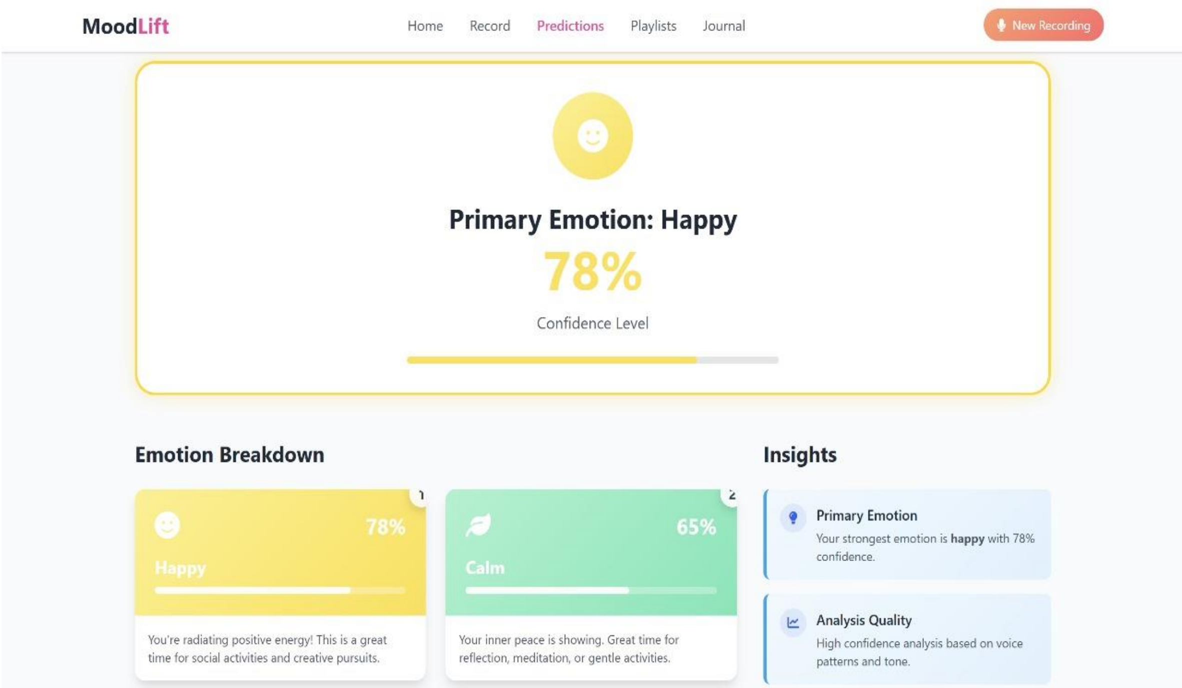
10 Results

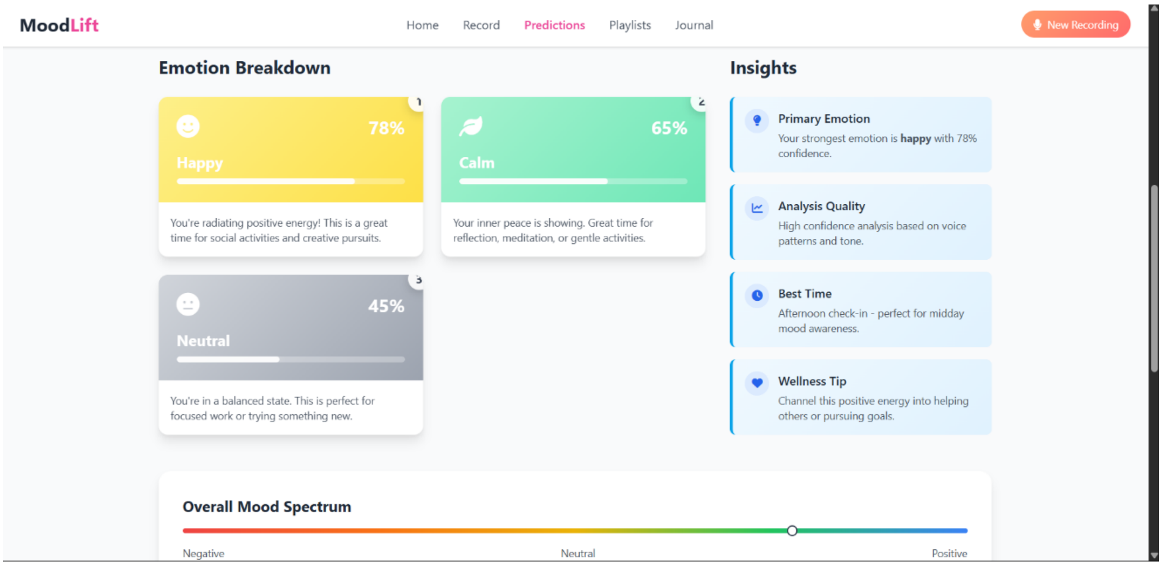
10.7 Screen shots

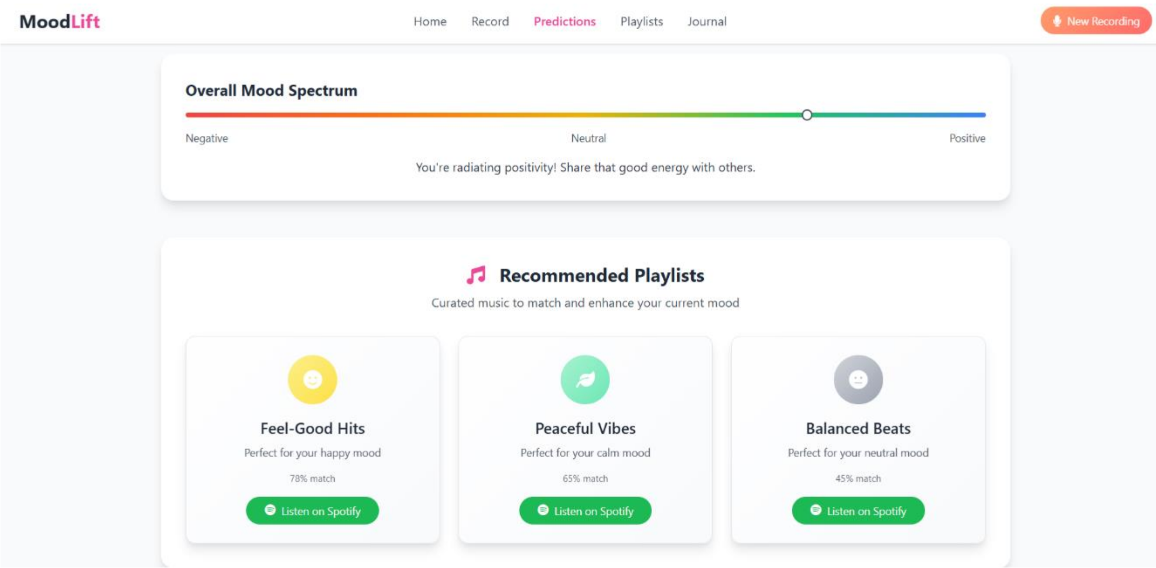




10.2. Outputs


MoodLift
Home
Record
Predictions
Playlists
Journal
New Recording
Primary Emotion: Happy
78%
Confidence Level
Emotion Breakdown
78%
Happy
You're radiating positive energy! This is a great time for social activities and creative pursuits.
65%
Calm
Your inner peace is showing. Great time for reflection, meditation, or gentle activities.
Insights
Primary Emotion
Your strongest emotion is happy with 78% confidence.
Analysis Quality
High confidence analysis based on voice patterns and tone.


MoodLift
Home
Record
Predictions
Playlists
Journal
New Recording
Emotion Breakdown
78%
Happy
You're radiating positive energy! This is a great time for social activities and creative pursuits.
65%
Calm
Your inner peace is showing. Great time for reflection, meditation, or gentle activities.
45%
Neutral
You're in a balanced state. This is perfect for focused work or trying something new.
Insights
Primary Emotion
Your strongest emotion is happy with 78% confidence.
Analysis Quality
High confidence analysis based on voice patterns and tone.
Best Time
Afternoon check-in - perfect for midday mood awareness.
Wellness Tip
Channel this positive energy into helping others or pursuing goals.
Overall Mood Spectrum
Negative
Neutral
Positive


MoodLift
Home
Record
Predictions
Playlists
Journal
New Recording
Overall Mood Spectrum
Negative
Neutral
Positive
You're radiating positivity! Share that good energy with others.
Recommended Playlists
Curated music to match and enhance your current mood
Feel-Good Hits
Perfect for your happy mood
78% match
Listen on Spotify
Peaceful Vibes
Perfect for your calm mood
65% match
Listen on Spotify
Balanced Beats
Perfect for your neutral mood
45% match
Listen on Spotify

11 Conclusion and future scope

Conclusion:

The developed Speech Emotion-based Song Playlist Recommendation system presents a significant advancement in personalized entertainment, uniquely aligning music selections with a user's detected emotional state. By effectively leveraging sophisticated voice analysis and robust deep learning models, this system provides a highly adaptive and mood-sensitive playlist generation capability that considerably enhances user satisfaction and engagement. This methodology not only heightens the importance of music suggestions but also deeply showcases the viability of combined artificial intelligence and live emotion sensing in interactive entertainment platforms. In conclusion, the successful system implementation provides users with an enriched and empathetic listening experience, making it a worthwhile contribution to the developing persona of customized music streaming services.

Future Scope:

The future scope of this project includes several potential enhancements and applications:

1) Increased Emotional Granularity: Future studies will investigate adding a larger range of subtle emotional states in addition to the existing set (e.g., excitement, stress, contemplation, boredom). Adding this scope is intended to allow for a more detailed and accurate mapping between rich human emotion and customized music recommendations, and for higher and more effective levels of personalization.
2) Continuous Adaptive Personalization: Future versions will emphasize the application of sophisticated machine learning models that learn continuously from user-specific preferences, explicit feedback (e.g., likes/dislikes), and longitudinal emotional patterns. This can enable the system to dynamically adapt song recommendations to remain long-term relevant and to change in response to user tastes.
3) Real-Time Mood Adjustment & Dynamic Playlists: The platform could be made to constantly evaluate subtle mood shifts during a listening session. This feature would allow dynamic real-time adjustments to the prevailing playlist, shifting between songs unobtrusively to adjust according to changing emotional states and potentially lead the listener towards the intended mood state, providing an Adaptive experience.

4) Multi-Service and Platform Integration: Integration in future development will focus on unified collaboration with more music streaming services than available APIs, guaranteeing compatibility across all devices such as specific mobile apps, web platforms, and smart home devices. This will drastically expand the availability of the system and its exposure across multiple user ecosystems.
5) Ethical Implications and Privacy Improvements: Since the system handles sensitive individual information (voice and emotional status), future research can focus on privacy mitigation through strong anonymization and encryption mechanisms. This involves investigating responsible AI deployment frameworks and transparency in data utilization to establish and sustain user confidence in such individualized systems.

12 Bibliography

1. S. Bhosale, C. Kothawade, G. Jagtap, V. Chavan, and P. K. Relusinghani, "Speech Emotion Based Music Recommendation System," International Journal of Innovative Science and Research Technology, vol. 9, no. 4, Apr. 2024. Doi: 10.38124/ijisrt/IJISRT24APR882.
2. C. Kothawade, G. Jagtap, P. K. Relusinghani, V. Chavan, and S. S. Bhosale, "A Survey on Speech Emotion Based Music Recommendation System," International Journal of Scientific Research in Engineering and Management (IJSREM), vol. 8, no. 4, Apr. 2024. Doi: 10.55041/IJSREM30150.
3. Keras Documentation. *Keras: The Python Deep Learning API.* https://keras.io/
4. TensorFlow Documentation. *TensorFlow: An End-to-End Open Source Machine Learning Platform.* https://www.tensorflow.org/
5. Scikit-learn Documentation. *Machine Learning in Python.* https://scikit-learn.org/
6. RAVDESS Emotional Speech Audio Dataset. https://zenodo.org/record/1188976

References

[1] S. Bhosale, C. Kothawade, G. Jagtap, V. Chavan, and P. K. Relusinghani, "Speech Emotion Based Music Recommendation System," International Journal of Innovative Science and Research Technology, vol. 9, no. 4, Apr. 2024. Doi: 10.38124/ijisrt/IJISRT24APR882.

[2] C. Kothawade, G. Jagtap, P. K. Relusinghani, V. Chavan, and S. S. Bhosale, "A Survey on Speech Emotion Based Music Recommendation System," International Journal of Scientific Research in Engineering and Management (IJSREM), vol. 8, no. 4, Apr. 2024. Doi: 10.55041/IJSREM30150.

[3] Y. Sahare, K. Kamble, T. Bhakte, R. Warkade, A. Besekar, S. Patil, and H. Gorewar, "Emotion Based Music Recommendation System," International Journal of Research Publication and Reviews, vol. 4, no. 12, pp. 818-821, Dec. 2023.

[4] K. Tharmikan, N. Heisapirashoban, M. A. M. A. Riza, R. R. S. Dinoshan, and T. Thilakarthna, "Music Moody - Facial Recognition And Voice Recognition To Detect Mood And Recommend Songs.," International Journal of Advanced Research and Publications, vol. 6, no. 10, Oct. 2023.

[5] A. Rajesh, T. Pavani, R. V. S. Naveen, P. Manikanta, and P. Sandeep, "SONG RECOMMENDATION BASED ON VOICE TONE ANALYSIS," Journal of Engineering Sciences, vol. 15, no. 02, pp. 210-211, 2024.

[6] V. K. Bongirwar and A. Potnurwar, "Song recommendation using speech emotion recognition," International Journal of Health Sciences, vol. 6, no. S1, pp. 10428–10434, 2022. Doi: 10.53730/ijhs.v6nS1.7498.

[7] G. Deshmukh, A. Gaonkar, G. Golwalkar, and S. Kulkarni, "Speech based Emotion Recognition using Machine Learning," International Journal of Research and Analytical Reviews (IJRAR), 2021.

[8] A. A. Ansari, A. K. Singh, A. Singh, and H. Murugan, "Speech Emotion Recognition Using Convolutional Neural Network (CNN)," International Journal of Psychosocial Rehabilitation, vol. 24, no. 08, pp. 2409-2414, 2020. Doi: 10.37200/IJPR/V24I8/PR280260.

[9] S. Lukose and S. S. Upadhya, "Music Player Based on Emotion Recognition of voice signals," in Proc. 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), pp. 1751-1754, 2017.

[10] S. Adebiyi, "An Emotion Based Music Recommender System Using Deep Learning," MSc Research Project, National College of Ireland, 2020.

Laboratory assignments on Project Analysis of Algorithmic Design

I **Increase/Improve**	D **Drive/Deliver**	E **Educate/Evaluate**	A **Accelerate/Associate**
Increase automation capabilities for task completion.	Drive innovation by integrating new features (e.g., voice recognition).	Educate users on the functionalities and benefits of the assistant.	Accelerate development through the use of optimized algorithms and threading.
Improve the AI's ability to from interactions learn user and improve performance over time.	Deliver a seamless user experience by ensuring low latency and high accuracy in results.	Evaluate the System's performance regularly through user feedback and technical testing.	Associate with other tools to enhance usability and efficiency.
Ignore low priority features that don't directly contribute to user productivity.	Decrease complexity by simplifying the interface and ensuring intuitive design.	Eliminate errors by implementing robust testing strategies for different modules.	Avoid unnecessary dependencies or tools may slow down development or compromise performance.

Table 1 IDEA Matrix

Laboratory assignments on Project Quality and Reliability Testing of Project Design

Assignment 1: Divide and Conquer Strategy for Distributed/Parallel/Concurrent Processing

1) Objective: To apply divide and conquer strategies to exploit distributed/parallel/concurrent processing for identifying objects, morphisms, function overloading, and functional relationships in the Personal Developer Assistant project.
2) Tasks:
   i. Identify the primary objects in the system (e.g., User queries, Assistant actions, Task management, etc.).
   ii. Use Venn diagrams to represent relationships between objects and morphisms (functions transforming one object into another).
   iii. State Diagrams: Create state diagrams for the different stages the assistant goes through (e.g., idle, processing query, fetching data, responding).
   iv. Analyze functional relationships between inputs and outputs using function relations.
   v. Examine the overload of functions and their impact on system performance.
3) Outcome: Develop a detailed functional model using the divide and conquer approach that can be parallelized or distributed for optimal efficiency.

Assignment 2: Functional Dependency Graphs and Software Modeling

1) Objective: Use identified objects, morphisms, and functional relationships to draw functional dependency graphs and apply relevant software modeling techniques.
2) Tasks:
   i. Create UML diagrams including class diagrams, sequence diagrams, and activity diagrams using tools like Visual Paradigm.
   ii. Draw dependency graphs showing the interactions between different modules and how they depend on each other.
   iii. Ensure the UML diagrams represent key software components and their interaction patterns (user input -> system processing -> output generation).
3) Outcome: Comprehensive software model that visualizes project architecture

Assignment 3: Testing the Project Using Generated Test Data

1) Objective: Generate test data based on mathematical models and perform functional testing of the project components using black box testing principles.
2) Tasks:
   i. Use tools like Mathematica or equivalent open-source tools to generate realistic test data for project functions.
   ii. Conduct function testing by selecting appropriate testing tools for functional testing, such as PyTest (for Python), or Selenium (for GUI based testing).
   iii. Write black-box test cases for each identified function:
   iv. Test input variations to check how the system handles edge cases.
   v. Validate output based on expected behavior and system response time.
   vi. Test the UML diagram reliability by checking for consistency in the design, dependency checks, and accurate representation of system behavior.
3) Outcome: A report on the system's quality and reliability based on functional tests, with detailed test cases and results.

Additional Assignments by the Guide:

Entrepreneurship-Oriented Tasks:

1) If the project leans toward entrepreneurship, assignments may focus on creating business models, analyzing market potential, or developing a go-to market strategy.
2) Use of tools such as Business Model Canvas or SWOT Analysis for business feasibility assessments

Project Planner

1. Create the Project Board
    a. Project Title: Speech Emotion Based Song Playlist Recommendation System Using Deep Learning.
    b. Board Columns:
        i. Backlog: For tasks planned but not yet started.
        ii. In Progress: For tasks currently being worked on.
        iii. Review: For tasks completed and awaiting review.
        iv. Done: For tasks that have been completed and approved.
2. Define Project Stages/Phases
    a. Planning and Feasibility Analysis
    b. Architectural Design
    c. Component and Data Design
    d. ML Algorithm Development
    e. Backend and UI Implementation
    f. Testing and Validation
    g. Documentation
    h. Final Review and Handover
3. Task Assignment
    a. Mayank Adhav: UI Designing and Development, user interface creation, Documentation.
    b. Jeet Harne: Backend Development, server-side logic, API development.
    c. Aaditya Meher: AI/ML Development, model training, and ML logic.
    d. Pratik Hagawane: Data collection, preprocessing, and analysis, Integration.
4. Set Milestones and Deadlines
    a. Phase 1: Requirements finalization and feasibility analysis (Week 1–2)
    b. Phase 2: Architecture and component design (Week 3–5)
    c. Phase 3: Implementation (Week 6–9)
    d. Phase 4: Testing and validation (Week 10–11)
    e. Phase 5: Final review and documentation (Week 12)
5. Add Task Dependencies
    Link dependent tasks to ensure proper sequencing:
    a. Example: Backend development depends on the completion of ML algorithm design.

b. Example: UI design starts after the architecture is finalized.

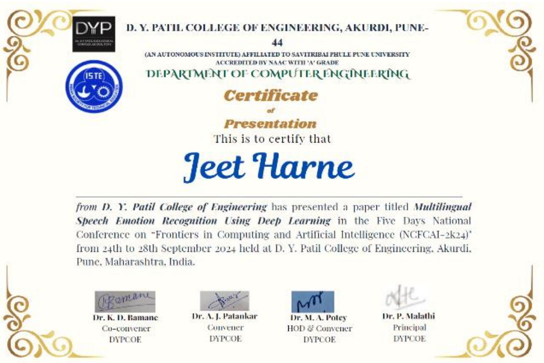
6. Monitoring Progress
   a. Use color-coded labels to highlight task priorities (e.g., high priority, low priority).
   b. Add due dates for each task and set reminders for approaching deadlines.
7. Collaboration

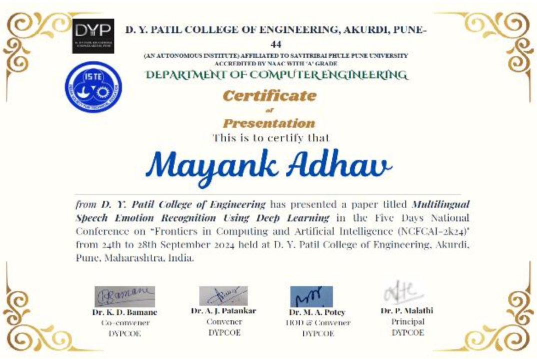
   Enable notifications and updates to keep all team members aware of task changes, assignments, or when a task is completed.

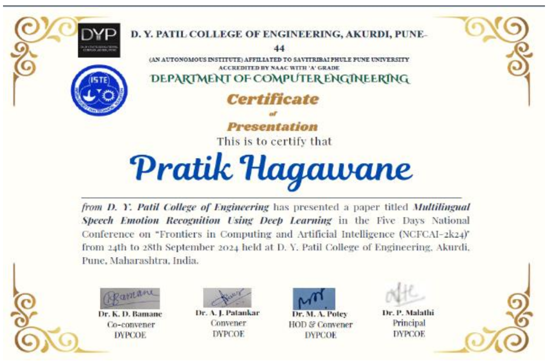
Reviewers Comments of Paper Submitted

- Paper Title: “Multilingual Speech Emotion Recognition using Deep Learning”
- Name of the Conference/Journal where paper submitted : IJSRCSEIT.
- Paper accepted/rejected : Accepted
- Review comments by reviewer : Single language but for music application.
- Corrective actions if any :Usage of this system for Music recommendation.


D. Y. PATIL COLLEGE OF ENGINEERING, AKURDI, PUNE-44
(AN AUTONOMOUS INSTITUTE) AFFILIATED TO SAVITRIBAI PHULE PUNE UNIVERSITY
ACCREDITED BY NAAC WITH 'A' GRADE
DEPARTMENT OF COMPUTER ENGINEERING
Certificate of Presentation
This is to certify that
Aaditya Meher
from D. Y. Patil College of Engineering has presented a paper titled Multilingual Speech Emotion Recognition Using Deep Learning in the Five Days National Conference on "Frontiers in Computing and Artificial Intelligence (NCFCAI-2k24)" from 24th to 28th September 2024 held at D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India.
Dr. K. D. Bamane, Co-convener, DYPCOE | Dr. A. J. Patankar, Convener, DYPCOE | Dr. M. A. Potey, HOD & Convener, DYPCOE | Dr. P. Malathi, Principal, DYPCOE


D. Y. PATIL COLLEGE OF ENGINEERING, AKURDI, PUNE-44
(AN AUTONOMOUS INSTITUTE) AFFILIATED TO SAVITRIBAI PHULE PUNE UNIVERSITY
ACCREDITED BY NAAC WITH 'A' GRADE
DEPARTMENT OF COMPUTER ENGINEERING
Certificate of Presentation
This is to certify that
Jeet Harne
from D. Y. Patil College of Engineering has presented a paper titled Multilingual Speech Emotion Recognition Using Deep Learning in the Five Days National Conference on "Frontiers in Computing and Artificial Intelligence (NCFCAI-2k24)" from 24th to 28th September 2024 held at D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India.
Dr. K. D. Bamane, Co-convener, DYPCOE | Dr. A. J. Patankar, Convener, DYPCOE | Dr. M. A. Potey, HOD & Convener, DYPCOE | Dr. P. Malathi, Principal, DYPCOE


D. Y. PATIL COLLEGE OF ENGINEERING, AKURDI, PUNE-44
(AN AUTONOMOUS INSTITUTE) AFFILIATED TO SAVITRIBAI PHULE PUNE UNIVERSITY
ACCREDITED BY NAAC WITH 'A' GRADE
DEPARTMENT OF COMPUTER ENGINEERING
Certificate of Presentation
This is to certify that
Mayank Adhav
from D. Y. Patil College of Engineering has presented a paper titled Multilingual Speech Emotion Recognition Using Deep Learning in the Five Days National Conference on "Frontiers in Computing and Artificial Intelligence (NCFCAI-2k24)" from 24th to 28th September 2024 held at D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India.
Dr. K. D. Bamane, Co-convener, DYPCOE | Dr. A. J. Patankar, Convener, DYPCOE | Dr. M. A. Potey, HOD & Convener, DYPCOE | Dr. P. Malathi, Principal, DYPCOE


D. Y. PATIL COLLEGE OF ENGINEERING, AKURDI, PUNE-44
(AN AUTONOMOUS INSTITUTE) AFFILIATED TO SAVITRIBAI PHULE PUNE UNIVERSITY
ACCREDITED BY NAAC WITH 'A' GRADE
DEPARTMENT OF COMPUTER ENGINEERING
Certificate of Presentation
This is to certify that
Pratik Hagawane
from D. Y. Patil College of Engineering has presented a paper titled Multilingual Speech Emotion Recognition Using Deep Learning in the Five Days National Conference on "Frontiers in Computing and Artificial Intelligence (NCFCAI-2k24)" from 24th to 28th September 2024 held at D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India.
Dr. K. D. Bamane, Co-convener, DYPCOE | Dr. A. J. Patankar, Convener, DYPCOE | Dr. M. A. Potey, HOD & Convener, DYPCOE | Dr. P. Malathi, Principal, DYPCOE

Plagiarism Report

iThenticate Page 1 of 76 - Cover Page Submission ID trn:oid:::3618:100012892

Mayank Adhav

Mayank BE_REPORT.docx

D Y Patil College of Engineering, Akurdi

Document Details

Submission ID

trn:oid:::3618:100012892

Submission Date

Jun 9, 2025, 4:07 PM GMT+5:30

Download Date

Jun 9, 2025, 4:10 PM GMT+5:30

File Name

Mayank BE_REPORT.docx

File Size

2.0 MB

72 Pages

12,371 Words

75,116 Characters

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Small Matches (less than 15 words)

Match Groups

- **38 Not Cited or Quoted 8%**
  Matches with neither in-text citation nor quotation marks
- **1 Missing Quotations 0%**
  Matches that are still very similar to source material
- **6 Missing Citation 1%**
  Matches that have quotation marks, but no in-text citation
- **1 Cited and Quoted 0%**
  Matches with in-text citation present, but no quotation marks

Top Sources

9%	Internet sources
1%	Publications
1%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Term-II Project Laboratory Assignments

1. Review of design and necessary corrective actions taking into consideration the feed- back report of Term I assessment, and other competitions/conferences participated like IIT, Central Universities, University Conferences.
2. Project workstation selection, installations along with setup and installation report preparations.
3. Programming of the project functions, interfaces and GUI (if any) as per 1st Term term-work submission using corrective actions recommended in Term-I assessment of Term-work.
4. Test tool selection and testing of various test cases for the project performed and generate various testing result charts, graphs etc. including reliability testing. **Additional assignments for the Entrepreneurship Project:**
5. Installations and Reliability Testing Reports at the client end.

Information of Project Group Member 1

Name : Mayank Adhav

Date of Birth : 01/10/2002

Gender : Male

Permanent Address : Gokulam Society, Patil Nagar, Chikhali 411062

E-Mail : adhavm6381@gmail.com

Mobile/Contact No. : 7509997076

Placement Details : Capgemini, LTIMindtree

Paper Published : Yes

Information of Project Group Member 2

Name : Aaditya Meher

Date of Birth : 16/11/2003

Gender : Male

Permanent Address : Near Akurdi Railway Station, Gurudwara Chowk, Akurdi, Pune

E-Mail : aadityameher16@gmail.com

Mobile/Contact No. : 7219688659

Placement Details : In Progress

Paper Published : Yes

Information of Project Group Member 3

Name : Jeet Harne

Date of Birth : 10/10/2003

Gender : Male

Permanent Address : Near Akurdi Railway Station, Gurudwara Chowk, Akurdi, Pune

E-Mail : jeetharne723@gmail.com

Mobile/Contact No. : 9405713734

Placement Details : In Progress

Paper Published : Yes

Information of Project Group Member 4

Name : Pratik Hagawane

Date of Birth : 23/01/2004

Gender : Male

Permanent Address : Near Akurdi Railway Station, Gurudwara Chowk, Akurdi, Pune

E-Mail : pratikhagawane745@gmail.com

Mobile/Contact No. : 7385165104

Placement Details : In Progress

Paper Published : Yes