Multi-Model Approach to Recommend Personalized Music Playlist

TMP – 2023 – 24 - 065

Individual Final Report

GUNASEKARA C.M – IT20665852

Supervised by – Mr. Thusithanjana Thilakarathne

B.Sc. (Hons) Degree in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

May 2024

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Student ID	**Name**	**Signature**
IT20665852	Gunasekara C.M	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

....

Mr. Thusithanjana Thilakarathne

Supervisor

2024.05.04

....

Date

..

Dr. Darshana Kasthurirathne

Co - Supervisor

2024.05.04

....

Date

ABSTRACT

Due to the growing accessibility of social media data and developments in machine learning technologies, the field of emotion recognition using voice frequency for music recommendations system yet to be revealed. This study focuses on an innovative technology that detects users' emotional moods by listening to their vocal frequencies. In addition to understanding the emotional context of users' spoken interactions, this dynamic technology provides recommendations and personalized playlist ideas that are sensitive to their emotions. The goal is to develop a unique and intelligent system that blends voice-based emotion recognition with personalized playlist production. The purpose of this system is to enhance user experiences by fostering a dynamic and emotionally resonant link between the user and the application. Our technology allows users to express themselves through speech, resulting in a unique and emotionally engaging engagement experience. It does this by combining signal processing, machine learning, and real-time analytics. When users offer voice clip, the system identifies and categorizes their emotions using mfcc techniques, based on that emotion providing them more personalized playlists . This concept represents a paradigm change in how humans and technology interact, in which technology responds to users' emotional needs and use speech and music to provide a customized and emotionally rich experience. In a nutshell, my work reflects the integration of emotion-aware technology with user-driven interaction, resulting in a system that communicates with and collecting user's voice clip and analyze frequency of the voice and then send to recommendation hub to playlist generation.

ACKNOWLEDGEMENT

We are deeply grateful to our committed supervisor, Mr. Thuthithanjana Thilakarathna, whose persistent advice and unflinching support have been the foundation of our research journey. His expert ideas and efforts have helped shape our knowledge and refine our research strategy. Equally deserving of our gratitude is our co-supervisor, Dr. Darshana Kasthurirathna, whose outstanding contributions, thorough feedback and constructive ideas have considerably improved the quality of our research. His skills and guidance have been critical in guiding us in the proper path and assuring the success of our research projects. We are grateful to our external supervisor, Dr. Kusumanjalee Thilakarathna, for graciously providing her domain-specific views and skills in sycolocolgy. Her contributions have given us a broader perspective and a better grasp of the complexities involved in our study. We would like to thank our friends and colleagues for their essential advice, support, and assistance. We would want to thank everyone who helped make this project a success, even if their names are not listed here.

Contents

List of figure

List of tables

List of Abbreviations	Description
RNN	Recurrent Neural Network
IEEE	Institute of Electrical and Electronics Engineers
AWS	Amazon Web Services
DOI	Digital Object Identifier
ML	Machine Learning
DL	Deep Learning
AI	Artificial Intelligence

1. INTRODUCTION

1.1. BACKGROUND

In today's fast-paced world, while competition is challenging and stress is common, music emerges as a powerful antidote to our collective stresses. As we navigate this never-ending race, the desire for unique audio experiences increases dramatically. To accommodate this need, researchers are digging deeply into the areas of music suggestion, customization, and user experience development.

Music recommendation is at the forefront of user experience development, providing a portal to a world of musical discovery and peace of mind. It creates a welcoming environment in which users can easily find new releases, dig into their favorite genres, and explore varied musical landscapes. Over time, many AI-driven recommendation systems have arisen, including content-based filtering, collaborative filtering, and hybrid filtering, each customized to particular aspects of user preferences.

Content-based filtering functions by evaluating a user's consumption history from the moment they join the site, adapting suggestions to their specific preferences and behaviors. Collaborative filtering, on the other hand, pools user data to detect commonalities and similarities, then makes suggestions based on shared preferences. Hybrid filtering combines the benefits of both techniques, using a detailed understanding of user behavior to provide highly tailored suggestions.

The value of such notions extends beyond entertainment, with implications in disciplines such as mental health and therapy. Platforms such as YouTube, Spotify, and iTunes demonstrate the effective incorporation of hybrid recommender systems, which have transformed the music streaming landscape by offering consumers seamless and highly customized listening experiences.

In essence, while the world moves forward, music recommendation algorithms serve as calming lighthouses, leading consumers through the clamor of modern life to find the appropriate soundtrack for each time.

In recent years, due to technological improvements and a greater emphasis on enhancing user experiences, Human-Computer Interaction (HCI) has evolved significantly. Recognition of emotions technology, a growing subject in HCI, has attracted attention for its potential to transform customized suggestion systems. This confluence represents a transformational horizon in which technology may fully engage with humans on an emotional level, as shown in sectors like voice-based interactions and playlist generating.

Traditional HCI research was largely concerned with increasing usability and functionality. However, integrating emotion detection technology adds a new level to user engagement by allowing computers to comprehend and respond to users' emotional states in real-time. This trend has resulted in unique applications such as voice-based emotion recognition and personalized playlist creation, which try to tailor user experiences to their feelings and preferences.

Voice-based emotion detection has seen remarkable advancements, leveraging sophisticated methodologies such as Mel-Frequency Cepstral Coefficients (MFCCs) and deep learning models. These techniques enable systems to accurately discern emotional states from speech signals, paving the way for emotionally intelligent interactions. Simultaneously, personalized playlist creation has evolved to cater to individual preferences, with research exploring the integration of emotional cues with music recommendations.

The existing research emphasizes the need of smoothly integrating real-time emotion detection into customized recommendation systems. However, this integration presents obstacles, such as developing emotionally coherent playlists that are relevant to users' current emotional states while respecting their preferences. Furthermore, research is ongoing into the creation of intuitive interfaces for emotion input identification and algorithms capable of creating emotionally consistent playlists.

The past study has made major contributions to this topic, ranging from investigations into emotion-aware recommendation systems to the use of facial expression detection for tailored music recommendations. These investigations have paved the way for novel techniques that seek to bridge the gap between emotion identification and playlist generation. However, multidisciplinary collaboration and a user-centric approach are still required to fully achieve the convergence's promise for improving the entire human-computer interaction experience.

1.2. Literature survey

In the rapidly developing field of human-computer interaction (HCI), the combination of emotion identification technologies with personalized recommendation systems provides a revolutionary horizon. This combination has the potential to transform user experiences, especially in voice-based interactions and playlist creation. Voice-based emotion detection has advanced significantly, leveraging complex approaches such as Mel-Frequency Cepstral Coefficients (MFCCs) and deep learning models to effectively distinguish emotional states from speech signals in real-time [1]. Simultaneously, personalized playlist creation has grown to accommodate individual preferences, with research such as Nambiar and Palaniswamy's study titled "Speech Emotion Based Music Recommendation" demonstrating the integration of emotional signals with music recommendations [1]. Furthermore, Krupa et al.'s work on "Emotion aware Smart Music Recommender System using Two Level CNN" illustrates the use of deep learning techniques for emotion-aware recommendation systems [2]. However, despite these advances, there is still a gap in seamlessly integrating real-time emotion recognition into customized recommendation systems, a task that requires a comprehensive strategy that takes into account both technological and user-centric elements [2].

One of the most difficult issues is producing emotionally cohesive playlists that reflect users' present emotional states while honoring their preferences. Existing research, such as Mounika and Charitha's work on "Mood-Enhancing Music Recommendation System based on Audio Signals and Emotions," underlines the necessity of combining sound signals and emotions to improve user experience [3]. However, this study frequently focuses on solo suggestions rather than real-time emotion recognition, which leaves a gap in understanding how such integration affects user engagement and satisfaction [3]. Furthermore, research is ongoing into the creation of intuitive interfaces for voice input recognition and algorithms capable of creating emotionally consistent playlists [1][2]. This highlights the need of multidisciplinary cooperation and a user-centered strategy in bridging the gap between emotion identification and playlist building, eventually improving the entire human-computer interaction experience.

The paper "Emotion-Based Music Recommendation System" [4] discusses the creation of a customized music recommendation system that employs neural networks to suggest music based on facial expressions and emotions. Music is noted for its deep emotional connection and ability to drastically affect a person's mood. Traditional music recommendation systems are based on user preferences over time, but this research suggests a revolutionary technique that uses facial expressions to assess a person's mood for music suggestion. Previous research has looked into numerous approaches for emotion recognition, such as Artificial Neural

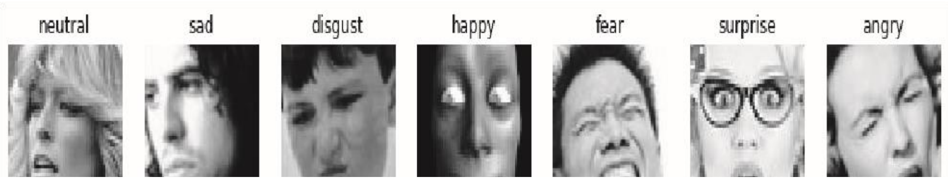


Figure 1:sample data they used

Networks (ANN). Several models, such ResNet-38 and CNN, have been created to recognize user emotions and propose music based on facial expressions. The system is separated into three major components: a face detector, an emotion detector based on a face, and a music selection system based on mood. The suggested technique seeks to automate music recommendation based on human facial expression detection via CNN. Current methods for playlist production based on human emotions are either time and storage substantial, poor in generating playlists depending on the user's emotional state or necessitate extra hardware and sensors. The study solves these challenges by creating an automated music selection system that recognizes human facial expressions. The study proposes a novel technique for music suggestion that uses facial expressions to efficiently assess a person's mood. The system uses neural networks and face recognition technologies to give consumers a more customized and engaging music listening experience. The suggested emotion-based music recommendation system provides a novel and creative approach to recommending music based on the user's mood, hence improving the entire listening experience. In conclusion, the research proposes a cutting-edge approach to music recommendation that uses facial expression recognition to detect emotions and recommends appropriate music playlists. This method is intended to give consumers a more customized and pleasurable music listening experience based on their current mood.

The research entitled "Face Emotion Based Music Recommendation System Using Modified Convolution Neural Network" [5] describes an innovative way to personalize music recommendations using real-time facial emotion analysis. The system combines three main components to provide a dynamic user experience: real-time face expression monitoring, an emotion-driven music recommendation algorithm, and personalized music suggestions. It uses modified convolutional neural networks to properly analyze facial expressions in real-time, delivering a detailed insight into the user's emotional state. This information is effortlessly integrated into a unique hybrid recommendation system, which matches music selections to the user's present feelings. Furthermore, by merging facial expression analysis with traditional approaches, the system generates individualized music choices based on the user's specific moods and interests. This comprehensive approach ensures that each user has a genuinely engaging and emotionally resonant music experience. The database contains a wide variety of facial photos expressing diverse emotions, together with training and testing data sets. Preprocessing techniques are used to standardize picture sizes, normalize pixel values, and enrich data to increase model robustness and generalizability. The research presents a strong face emotion detection model designed specifically for real-time music recommendation systems. A revolutionary music recommendation system uses face expression data to provide tailored and emotionally appropriate song recommendations. The technology prioritizes user privacy by analyzing facial expressions locally on the user's device, guaranteeing that no sensitive data is sent to other servers. The objective of this research is to bridge the gap between emotions and music suggestions by studying facial expressions and curating playlists that reflect consumers' sentiments. By utilizing facial expression analysis, the system can alleviate the cold-start problem that standard recommendation systems confront for new users. The research continues by stressing the potential for additional breakthroughs in emotion-based music recommendation systems, with a focus on user privacy, data security, and improved user experiences. Overall, the study proposes a unique strategy that uses face emotion detection and deep learning techniques to improve music recommendation systems, providing individualized and emotionally resonant music selections based on users' real-time emotional states.

The proposed research aims to solve these shortcomings by developing an integrated system that smoothly combines voice-based emotion identification with tailored playlist creation. Drawing on the approaches used in previous research, this initiative seeks to enhance and build on current frameworks. Using deep learning models, context-aware interfaces, and ethical concerns, the proposed study aims to reinvent human-computer interaction paradigms. The desired consequence is not only technological improvement, but a transformative experience that develops emotional ties between individuals and technology. By bridging the gap between emotion identification and playlist creation, this study aims to generate sympathetic and contextually appropriate user experiences,

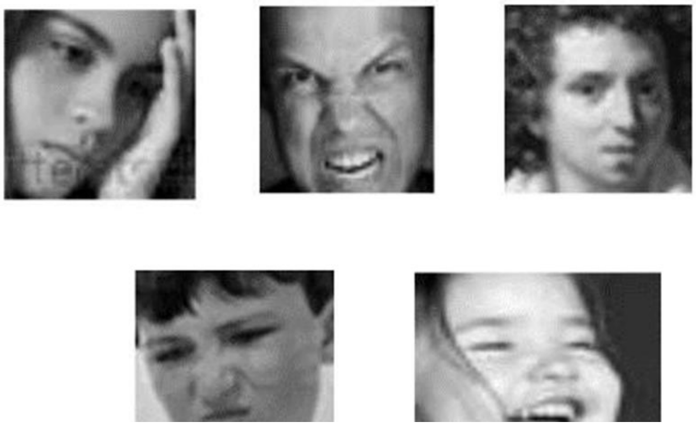


Figure 2-sample data they used

ushering in a new era of human-computer interaction.

As a summary, the combination of emotion detection technologies with tailored recommendation systems provides a promising frontier in human-computer interaction. The suggested activities attempt to solve current gaps and reinvent user experiences by drawing on lessons from previous research. Incorporating current emotion recognition utilizing voice clips into customized recommendation systems, through multidisciplinary collaboration and a user-centric approach, has the potential to improve user engagement and happiness. As academics continue to investigate this multidisciplinary subject, they pave the path for transformational advances that align with consumers' growing demands and expectations in the digital era.

1.3. Research gap

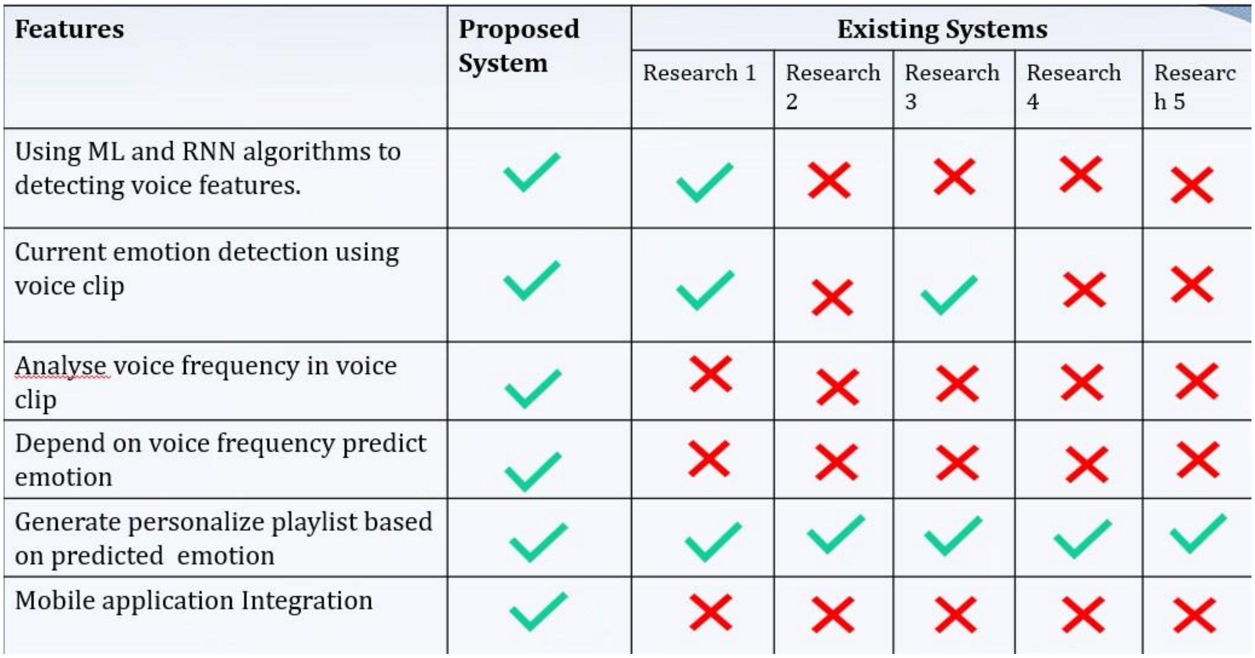
In the growing field of human-computer interaction (HCI), where the convergence of emotion recognition technology with personalized suggestion systems holds transformative promise, a critical research gap emerges in seamlessly integrating voice-based emotion recognition with personalized playlist production. Despite significant advances in both domains, current literature focuses on individual components rather than the complex dynamics of real-time interaction, coherent playlist generation, contextual recommendations, user engagement, and generalization across diverse contexts. This gap highlights the necessity for an integrated and emotionally intelligent engagement experience that seamlessly integrates various technologies.

While prior work has made significant contributions to voice-based emotion identification and music selection, it frequently overlooks the complexities of modern interaction paradigms. Emotion detection and music recommendation systems are frequently viewed as separate entities, missing the coherence required to comprehend and respond to users' emotional states in real-time while providing personalized and contextually appropriate material. As a result, there is a void in the literature about the development of integrated platforms capable of identifying, understanding, and adjusting to users' emotions while providing tailored music playlists.

Furthermore, current research focuses mostly on theoretical frameworks and algorithmic breakthroughs, with little attention on practical implementation and user-centered design. While algorithms for emotion recognition and playlist building have been significantly improved, the translation of these advances into understandable and user-friendly interfaces is still underexplored. As a result, there is a research gap in developing holistic systems that not only exploit cutting-edge technology but also prioritize user experience via intuitive interfaces and smooth interaction design.

Furthermore, the present research fails to address the complexities of contextual suggestions, which are critical for meeting users' changing requirements and preferences across diverse contexts. While some research recognizes the dynamic nature of user emotions and preferences, few focus on the creation of adaptive systems capable of contextual comprehension and tailored suggestions in real-time. As a result, there is a gap in understanding the junction of emotion identification, music selection, and user engagement across different contextual contexts.

In essence, the research gap is the lack of an integrated platform that effectively bridges the realms of voice-based emotion recognition and personalized playlist production, while also addressing the complexities of modern interaction dynamics, coherent playlist generation, contextual recommendations, intuitive interface design, and adaptive user engagement. To bridge this gap, future research should aim to create complete systems that smoothly integrate various parts, therefore pushing the boundaries of human-computer interaction and improving user experiences in the digital era.below diagram show existing researches gaps.(*figure 3*)



Features	Proposed System	Existing Systems				
		Research 1	Research 2	Research 3	Research 4	Research 5
Using ML and RNN algorithms to detecting voice features.	✓	✓	✗	✗	✗	✗
Current emotion detection using voice clip	✓	✓	✗	✓	✗	✗
Analyse voice frequency in voice clip	✓	✗	✗	✗	✗	✗
Depend on voice frequency predict emotion	✓	✗	✗	✗	✗	✗
Generate personalize playlist based on predicted emotion	✓	✓	✓	✓	✓	✓
Mobile application Integration	✓	✗	✗	✗	✗	✗

Figure 3:research gap.

1.4. research problem

Despite the widespread popularity of music streaming apps such as Spotify, iTunes, Alexa, and Amazon Music, there remains an untapped frontier in music streaming app development: the seamless integration of emotion recognition technology with personalized music recommendation systems. While existing systems provide huge libraries and individualized suggestions based on user preferences, they frequently ignore an important feature of users' present emotions, which have a substantial impact on their music listening preferences.

The primary problem we intend to solve is the creation of a novel way to improve music streaming applications by reliably predicting users' emotions using speech patterns and frequencies. Unlike earlier research, which have mostly focused on standalone emotion detection or music recommendation systems, our method aims to bridge the gap between these areas by using recurrent neural network (RNN) models for audio classification.

Specifically, our study challenge involves:

2. Building an Audio Categorization Model: We propose developing and training an RNN-based model that can analyze voice patterns and frequencies to reliably forecast users' emotions. This model will use deep learning technology to extract significant characteristics from audio inputs and categorize them into different emotional states.

2. Integrating Emotion Classification with Music Recommendation: Our goal is to combine the emotion classification model with existing music recommendation systems to produce a more cohesive mobile app experience. We want to increase user engagement and happiness by smoothly combining current emotion recognition with tailored music selections.

3. Overcoming Mood-Driven Music Listening Habits: By addressing the issue of mood-driven music listening habits, we want to give consumers a more natural and emotionally satisfying music streaming experience. Instead of depending exclusively on user preferences or previous data, our suggested system will dynamically change music recommendations in response to users' present emotional states.

4. Developing a Unified Mobile App: Our ultimate objective is to create a single mobile app that smoothly combines emotion detection technology with music streaming functions. This software will provide a platform for users to explore and discover music based on their current emotions, ultimately improving the overall listening experience.

In essence, the research problem is to use deep learning techniques to create an audio categorization model capable of predicting users' moods in real time, and then integrate this model with existing music recommendation systems to create a unified mobile app that caters to users' emotional states. By dealing with this issue, we hope to reshape the landscape of music streaming app development and give consumers a more customized and emotionally engaging musical listening experience.

2. RESEARCH OBJECTIVES

2.1. Main objective

The main aim of this recommender system is to elevate user satisfaction by providing custom music playlists finely tuned to their preferences and emotional states. This involves refining both usability and recommendation precision through a series of intricate procedures. These steps include capturing top-tier selfies, scrutinizing facial features, interpreting environmental cues, and employing voice prompts to discern user emotions. By amalgamating these inputs, the system endeavors to craft tailor-made playlists that deeply resonate with each user's distinct musical tastes and mood. Furthermore, sub goal entails utilizing voice samples to analyze the user's vocal frequency, thereby enriching the recommendation process.

2.2. Specific objectives

- Collect speech data inputs from the Kaggle dataset for an emotion classification model that can recognize various emotions.

  This sub-objective focuses on gaining speech data inputs from the Kaggle dataset for use in the emotion classification model. The technique entails obtaining pre-existing audio recordings from the Kaggle dataset, which is particularly curated for emotion categorization tasks. These audio recordings may have actors or people expressing a variety of emotions through speaking. It is critical to choose a broad and representative dataset that covers a wide variety of emotional states and has enough samples for each emotion category.

- Convert raw voice signals from the Kaggle dataset into a set of acoustic features using feature extraction.

  After collecting the vocal data inputs from the Kaggle dataset, the following step is to preprocess the raw speech signals and extract key acoustic properties. This entails using signal processing techniques to transform raw audio signals into a structured format suitable for analysis. Feature extraction methods such as Mel-frequency cepstral coefficients (MFCCs), pitch, intensity, and formants are used to capture the underlying acoustic properties of voice data. These retrieved characteristics are used as input for the emotion classification model.

Mel-Frequency Cepstral Coefficients (MFCCs) are a common approach for capturing speech spectrum features.

Brief-time Fourier Transform (STFT) decomposes a signal into its frequency components over brief time intervals.

Other aspects, such as pitch, energy, and formants, may be extracted to offer further information about the voice features.

- Train the emotion classification model on speech samples from the Kaggle dataset.

  After preparing the collected audio features, the next step is to train the emotion categorization model with machine learning or deep learning techniques. This entails choosing an appropriate model architecture (e.g., LSTM-based recurrent neural networks) and training it using labeled speech data inputs from the Kaggle dataset. The model learns to link certain auditory patterns with various emotional states, allowing it to reliably categorize previously unknown speech inputs into predetermined emotion categories. The model is trained by tweaking hyperparameters and verifying its performance with cross-validation techniques.

- Train the music recommendation model:

  In tandem with the development of the emotion classification model, this sub-objective focuses on training the music recommender model using merged datasets incorporating emotion-based speech samples. The recommender model learns to recommend tailored music playlists based on emotional states derived from speech snippets in the datasets. Training the music recommender entails using the vast range of emotional expressions collected in the datasets to suggest appropriate music tracks that correspond to the inferred emotional states. While the datasets do not expressly include user listening data, they can give useful insights into emotional expressions that may be utilized to influence the music selection process.

- Develop the mobile application accordingly:

  After training the emotion classification and music recommender models, the next step is to create the mobile application interface. The application interface should smoothly incorporate functions for recording and uploading voice snippets, showing emotion-based music suggestions, and delivering a consistent and straightforward user experience. User interface design concepts, platform-specific requirements, and usability testing are all used to guarantee that the program fulfills user expectations and usability standards.

3. REQUIREMENTS

1. Functional requirements

- Get the voice clip:

To capture current audio snippets from users, add a microphone icon to the Melowave mobile app's UI. The microphone capability allows users to record their voice straight within the program. The app should guarantee that the speech snippets it captures are of good quality to allow for accurate emotion analysis and processing. Once captured, the speech snippets are sent to a specific place for further processing and analysis by the system.

- The system should be capable of extracting sentiment from a voice clip:

After receiving the recorded voice clip, the Melowave system extracts and analyzes the emotional information inherent within it using powerful machine learning techniques such as Recurrent Neural Networks (RNNs). The system recognizes major emotional indicators in the speech, such as tone, pitch, and intensity, using signal processing and feature extraction methods like Mel-Frequency Cepstral Coefficients (MFCC). These collected elements go into the emotion identification model, allowing the system to correctly classify the emotional state indicated in the speech sample.

- The system ought to generate a customized playlist accordingly:

Based on the extracted emotional analysis of the speech sample, the Melowave system generates a personalized playlist based on the user's current emotional state. The algorithm generates and curates a playlist based on the user's indicated emotions, drawing on a huge collection of music classified by emotional qualities. The playlist-generating method considers genre preferences, pace, and lyrical themes in order to provide the listener with a unified and enjoyable listening experience.

- The user should be able to listen to the playlist using the application that was downloaded:

Once the personalized playlist is created, the Melowave app allows users to listen to the customized playlist directly from the application interface. Users may quickly explore the playlist, play individual songs, skip tracks, and change playback parameters to their liking. The app's straightforward and user-friendly design allows for a seamless and delightful listening experience, increasing user engagement and happiness.

2. Nonfunctional requirements

- Performance:

Real-Time Processing: The Melowave system must handle emotion assessment and playlist-generating requests in real-time to provide a smooth and responsive user experience. Emotion analysis and playlist construction should be performed within milliseconds after receiving the audio clip input.

Scalability: The system should be built to manage a growing number of users and concurrent requests while maintaining performance. It should be scalable to support increased user base and workload demand while maintaining peak performance levels.

- accuracy

Accuracy of Emotion Detection: The Melowave system's emotion classification model must accurately detect and categorize various emotional states based on voice inputs. To guarantee that playlist suggestions are of high quality, the model must continuously produce precise and trustworthy emotion predictions.

Playlist Relevance: The Melowave system's playlists should be very relevant and closely linked with the emotional states detected by the user's vocal inputs. To increase user satisfaction and engagement, the system must precisely match the emotional context of the user's voice clip to relevant music options.

- Security and Privacy:

Data Security: The Melowave system must follow strict data security measures to protect user voice data and preserve its confidentiality and integrity. Encryption, access restrictions, and secure data storage should be used to safeguard sensitive user information from unwanted access or alteration.

User Data Protection: Robust measures should be implemented to protect user data and privacy. To increase user trust and confidence, the system should emphasize user data protection by integrating features such as user permission processes, data anonymization, and adherence to privacy standards.

3. System requirements

- Laptop/Desktop:

The Melowave system is developed and deployed using a laptop or desktop computer as its hardware platform. It should have the minimal hardware requirements to execute the software tools and libraries efficiently. These specs normally include a powerful CPU, enough computer RAM, enough storage space, and, if necessary, a dependable graphics card. The computer offers the computing capacity and resources required for model training, data processing, and software development.

- Cameras and Audio Recorders:

High-quality cameras and audio recorders capture important inputs for the Melowave system, including selfies, photos, and voice clips. The cameras should be able to provide crisp, detailed pictures with accurate color representation and enough resolution. Similarly, audio recorders must capture high-fidelity sound recordings with minimum noise and distortion to accurately analyze and process speech inputs. These devices collect important data inputs for emotion recognition and playlist construction within the system.

- Internet Connection:

Accessing online resources, including cloud-based platforms, datasets, and APIs, requires a strong and consistent internet connection. The internet connection facilitates smooth communication between Melowave system components, allowing for real-time updates, data synchronization, and interaction with other services. Additionally, internet connectivity is required for users to use the Melowave mobile application, stream personalized playlists, and receive system updates or suggestions while on the move.

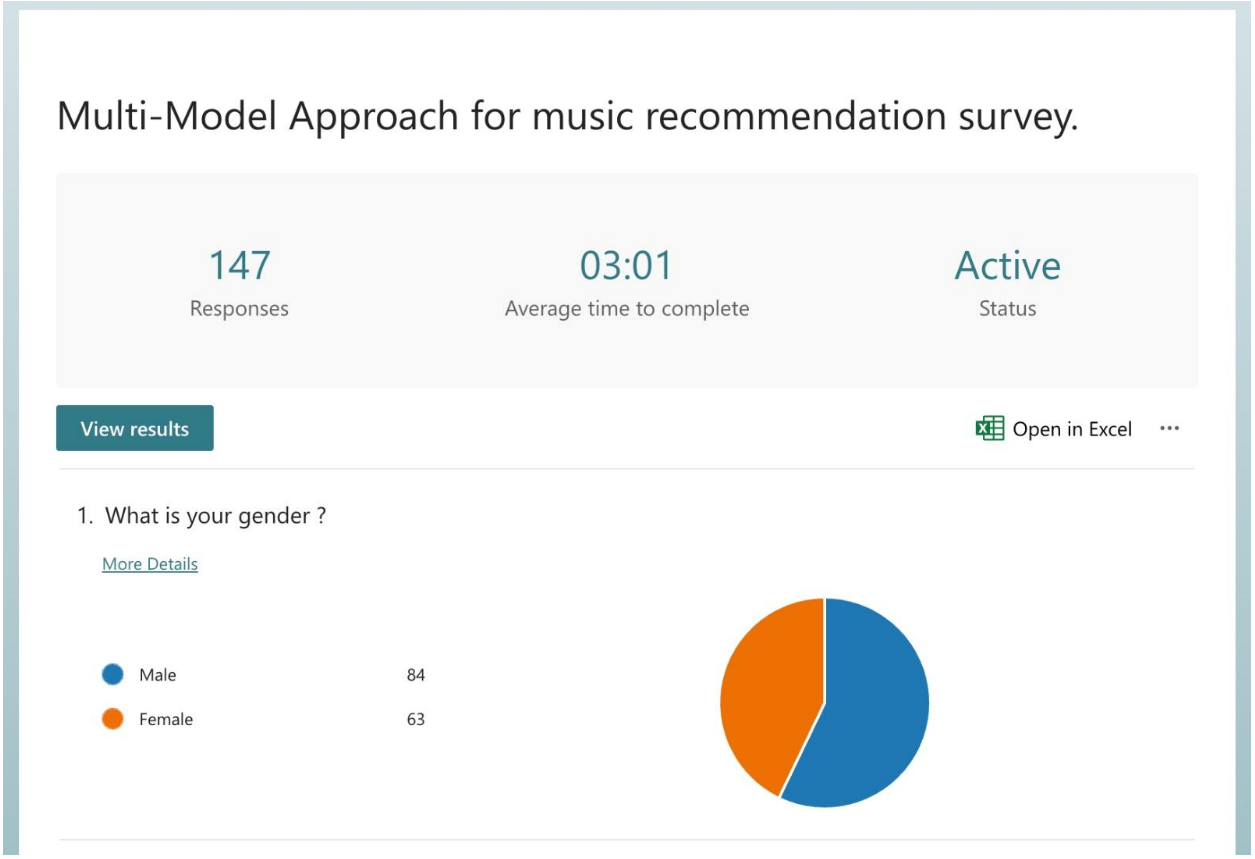
- Mobile Phone:

Melowave's mobile app delivers tailored music suggestions to consumers via their phones. It should be compatible with the system's application and support the required operating system (e.g., Android or iOS) and hardware specs. The mobile phone allows users to easily access their personalized playlists and interact with the recommendation engine from anywhere, increasing user engagement and happiness. Furthermore, mobile phones allow users to shoot pictures and voice inputs right within the app, which improves the user experience and allows for seamless interaction with the Melowave system.

4. METHODOLOGY

This chapter describes the process and methods for designing and developing emotion detection using voice frequency when it comes to train model and development of melowave mobile app. These subjects covers feasibility research, approaches, methodology, and requirement collecting. This chapter aims to present a concise technique for the study component "voice frequency-based emotion identification". In this concept, we intend to create a mobile application appropriate for use on a smartphone.

In the early stages of development, comprehensive data collecting was critical for informing the system's algorithms and ensuring its efficacy. This involved doing a thorough examination to identify significant elements impacting music tastes and mood. The research focused on data collection to determine users' subjective experiences with music. As a result, the first step was to run a poll to learn more about my component.



Multi-Model Approach for music recommendation survey.

147
Responses

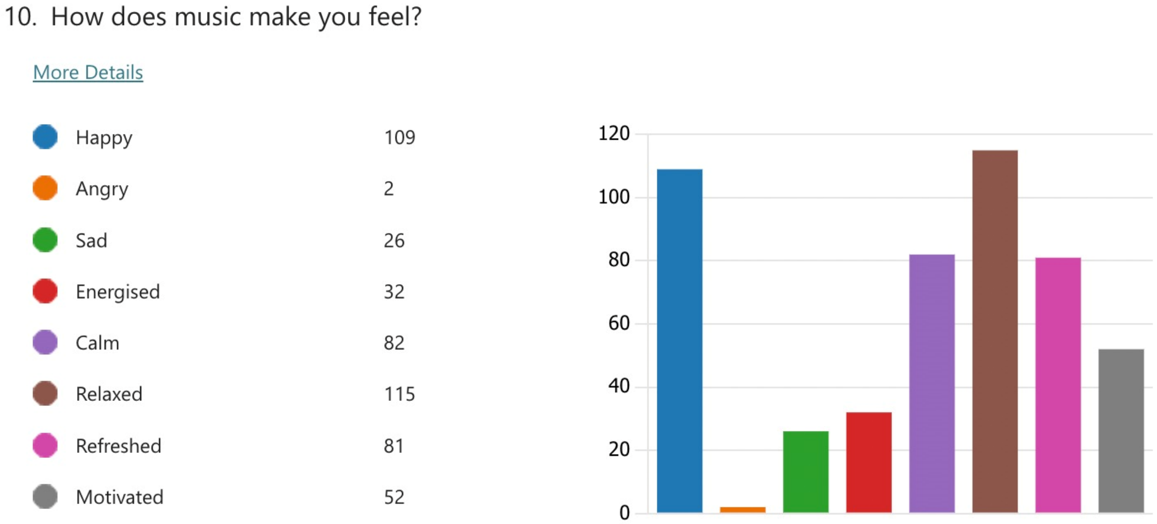
03:01
Average time to complete

Active
Status

1. What is your gender ?

● Male	84
● Female	63

Figure 4:survey results



10. How does music make you feel?

● Happy	109
● Angry	2
● Sad	26
● Energised	32
● Calm	82
● Relaxed	115
● Refreshed	81
● Motivated	52

Figure 5:survey results for feeling of music

4.1. Feasibility study

1. Technical feasibility:

The suggested methodology is technically feasible, as it utilizes publicly available datasets and existing machine learning algorithms. The system provides extensive resources for model training and assessment thanks to the use of open-source tools and frameworks such as TensorFlow and Librosa, as well as varied emotion-based voice clips from Kaggle datasets. The integration of machine learning techniques, such as deep learning with LSTM networks and signal processing for MFCC feature extraction, allows for reliable analysis and interpretation of speech data to successfully infer emotional states. Furthermore, optimization tactics such as parallel processing and lightweight deployments maximize computing resource consumption, improving scalability and responsiveness in real-time inference. These technological considerations all contribute to the viability of creating a dependable emotion categorization and music recommendation system.

2. Economic feasibility:

The suggested approach is economically feasible due to its low implementation costs and potential for long-term sustainability. Using open-source technologies and publicly available datasets decreases initial investment and eliminates the need for expensive proprietary software and data collecting. Cloud-based deployments provide scalability and flexibility while lowering infrastructure costs via pay-as-you-go pricing approaches. Furthermore, revenue-generating options, such as premium music suggestions and tailored playlists, improve the system's economic viability, ensuring its long-term sustainability and profitability.

3. Ethical feasibility:

The suggested technique comprises ethical concerns to ensure the responsible and fair deployment of the emotion categorization and music recommendation system. Maintaining data privacy and gaining user agreement for data use is critical, which may be accomplished through anonymized datasets and transparent data management policies. Addressing fairness and bias problems in algorithmic decision-making with varied and representative datasets promotes inclusion while reducing discrimination risks. Transparency regarding system capabilities and restrictions encourages user trust and accountability, allowing for ethical use and informed decision-making. By emphasizing ethical principles, the suggested technique is consistent with social norms and fosters responsible technical innovation.

4.2. Requirement gathering

The method of gathering requirements for our Music Recommendation System was tedious, with multiple critical phases to ensure the effective and ethical achievement of our study aims. Initially, we conducted a thorough review of previous research and examined accessible systems, utilizing a variety of internet resources to further our understanding.

A key aspect of our research was finding and evaluating systems similar to our proposed system, which provided important insights into common processes and technologies used in their development. Our research approach was developed through a detailed requirement-collecting process that included multiple stakeholders and information sources.

We began by gaining a clear understanding of our study objectives through meetings with experts from the music business, technological sector, and relevant academic domains. Their insightful insights guided us in developing specific criteria for evaluation, with a major focus on improving user experience through individualized music suggestions based on emotional states.

We then identified the critical technologies and tools necessary for data collecting and processing, working closely with experts in machine learning, emotion detection, and data analytics. Their knowledge was critical in selecting appropriate algorithms and libraries for a variety of applications, including emotion recognition, sentiment analysis, and user profiling.

Ethical concerns were crucial during the requirement collection phase. User privacy and ethical norms were of the highest significance. Collaborations with legal professionals and ethical review boards were critical to developing strong processes for informed consent and data anonymization.

The collection of necessary data for testing and validation required recording user interactions inside the program. This required open contact with users, the development of consent forms, and the careful selection of recording equipment.

Finally, our research methodology involved the production of an individual data set for model evaluation, which required thorough planning and execution to correctly capture a wide range of user preferences and emotional states.

The research approach was thoroughly developed to match our specific aims after soliciting requirements from industry professionals, specialists, and ethical authorities. This technique ensured that our Music Recommendation System was thoroughly evaluated while adhering to the highest ethical and research integrity requirements.

4.3. Model inplementation

1. Data collection and Preprocessing

1.1. *Dataset selection*

Any machine learning model's performance is determined by the quality and variety of the data used to train it. As a result, the datasets used for emotion identification are carefully chosen. To ensure the model's robustness and generalizability, datasets should include a diverse variety of emotional expressions, spoken languages, genders, ages, and cultural backgrounds. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), the Toronto emotional speech set (TESS), the Surrey Audio-Visual Expressed Emotion (SAVEE) database, and the Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D) are some of the most often used datasets for emotion identification.

1.2. Data Preprocessing

Before feeding the audio data into the model, numerous preprocessing processes are used to improve its quality and applicability for training:

-Feature Extraction: One important stage in audio preprocessing is feature extraction, which extracts significant characteristics from raw audio data. Mel-frequency cepstral coefficients (MFCCs) are commonly employed to depict the spectral properties of audio sources. MFCCs capture the frequency and temporal dynamics of audio waves, making them ideal for applications such as speech and emotion identification.

- Segmentation: Long audio recordings are frequently broken down into smaller, overlapping segments to capture temporal fluctuations in emotional expression. By dividing long recordings into smaller parts, the model can better catch tiny changes in sentiment over time.

- Label Encoding : Emotion labels for each audio segment are encoded numerically using techniques such as one-hot encoding. This transformation allows the model to read and learn from the emotion labels while training.

2. Model Architecture Design

2.1. LSTM-based Model Architecture

The proposed model architecture employs Long Short-Term Memory (LSTM) units, a form of recurrent neural network (RNN), to describe temporal relationships in audio sequences. LSTMs are ideal for jobs involving sequential data, such as voice and audio processing, since they can capture long-range relationships and alleviate the vanishing gradient issue found in regular RNNs.

2.2. Feature Representation and Learning

The LSTM layers are further enhanced by batch normalization, a method that normalizes the activations of the preceding layer, stabilizing and speeding up the learning process. Batch normalization, which reduces internal covariate shift, provides for faster convergence and better generalization. Following the LSTM layers, dense (completely connected) layers with Rectified Linear Unit (ReLU) activation functions are used to train higher-level representations of the audio data recovered by the LSTMs. ReLU activation functions add nonlinearity to the model, allowing it to learn complicated mappings between input characteristics and output labels.

2.3. Dropout Regularization

Dropout regularization is used to reduce overfitting and enhance model generalization in dense layers. Dropout loses a proportion of units at random during training, resulting in an ensemble of thinning networks. This stochastic regularization strategy adds redundancy to the network, forcing it to learn more robust characteristics while decreasing its dependence on individual neurons.

2.4. Output Layer and softmax Activation

The model's output layer is made up of dense units with softmax activation that generate probability distributions for the emotion classes ('angry', 'neutral','sad', and 'happy'). Softmax activation guarantees that the anticipated probabilities add up to one, allowing the model to generate accurate predictions about the emotional content of the input audio segments.

3. Model Training and Evaluation

3.1. Training Procedure

The model is trained with preprocessed audio samples and accompanying emotion labels. During training, the model uses the Adam optimizer to minimize a categorical cross-entropy loss function. Adam is an adaptive learning rate optimization technique that combines the advantages of AdaGrad and RMSProp, making it ideal for training deep neural networks.

3.2. Model evaluation

To evaluate the model's performance, it is tested on a different validation dataset using measures such as accuracy, precision, recall, and F1 score. Accuracy quantifies the fraction of properly identified cases, whereas precision and recall give information about the model's ability to reduce false positives and false negatives, respectively. The F1-score, or harmonic mean of accuracy and recall, provides a balanced evaluation of the model's performance across classes.

4. experimental setup and parameter tuning

4.1. Dataset Splitting

To guarantee sufficient data for both training and validation, the dataset is divided into training and validation sets at an appropriate ratio (e.g., 80% training, 20% validation). Stratified sampling is carefully considered to preserve class distribution throughout training and validation sets, particularly when datasets are unbalanced.

4.2. Hyperparameter Tuning

Hyperparameters like learning rate, batch size, and LSTM units are optimized for model performance utilizing approaches such as grid search or random search. The learning rate sets the step size used during gradient descent optimization, whereas batch size provides the amount of data processed before updating the model's parameters. LSTM units regulate the model's capacity and complexity, impacting its ability to grasp temporal relationships in data.

5. Results analysis and Interpretation

5.1. Performance metrics

The model's performance is assessed using numerous performance measures derived during evaluation. Confusion matrices are created to visually represent the model's classification accuracy for each emotion class, revealing the model's strengths and weaknesses across diverse emotional expressions.

5.2. Qualitative analysis

The model's predictions are qualitatively analyzed by reviewing sample audio segments and their accompanying emotion forecasts. This study identifies patterns, outliers, and probable causes of misclassification, providing useful insights into model behavior and performance.

6. Model discussion and conclusion.

6.1. Interpretation of Results

The results of the study are examined to generate conclusions about the suggested model's effectiveness for emotion identification from audio data. The strengths, limits, and potential areas of development are highlighted, offering a detailed insight into the model's performance and capabilities.

7. Ethical Considerations

7.1. Data Privacy and consent

Data privacy, informed permission, and responsible data usage are all handled to guarantee that ethical principles and legislation are followed. Individuals who provide data to the study are protected from privacy and confidentiality violations, and explicit agreement is acquired for data collection and utilization.

7.2. Fairness and bias Mitigation

Potential biases in the dataset and model predictions are detected, and steps are made to reduce bias and assure fairness, particularly for sensitive factors such as gender, ethnicity, and age.

Fairness-aware machine learning approaches, such as fairness constraints and bias detection algorithms, are used to ensure equity and inclusion in the model's predictions and results.

As a summary, the technique presented in this paper provides a systematic and rigorous framework for constructing and assessing an emotion identification model for audio data. We hope to create a strong and dependable model capable of reliably recognizing emotional states transmitted through audio recordings by combining sophisticated deep-learning techniques, rigorous experimentation, and ethical concerns. This research advances affective computing research and has major real-world implications in mental health assistance, human-computer interaction, and other areas.

4.4. Tools and technologies

- Python programming language

Python is a popular programming language for creating machine learning models due to its ease of use, adaptability, and extensive library ecosystem. Python's simple syntax and wide library support, which includes NumPy, pandas, TensorFlow, and PyTorch, gave us powerful tools for data processing, model construction, and evaluation. We were able to easily create sophisticated machine learning algorithms and prototypes using Python, which accelerated development and allowed for fast experimentation with multiple methodologies.

- Google Colab

Google Colaboratory (Colab) proved to be an invaluable resource for our project, improving our development workflow. Colab's free access to GPU and CPU resources allowed us to expedite model training and testing without incurring significant hardware costs. Furthermore, Colab's easy interface with Python, as well as built-in capabilities for data preposition, analysis, and visualization, helped to speed our collaborative development process. Its interactive platform enabled team members to participate in real-time, replace code snippets, and see findings, increasing productivity and allowing for more effective model iteration.

- Kaggle

Kaggle was our major data source platform, providing huge datasets and resources for data science and machine learning applications. Using Kaggle's various datasets gave us access to a variety of data sources relevant to our machine learning challenges, such as age, gender, weather, and voice emotion identification. Furthermore, Kaggle's community-based platform provided useful educational materials, tournaments, and forums for knowledge exchange and cooperation. We were able to improve the accuracy and resilience of our machine learning models by using Kaggle datasets.

The below mentioned dataset was taken from Kaggle for the use of model creation.

Emotion prediction

https://www.kaggle.com/datasets/dmitrybabko/speech-emotion-recognition-en/

- TensorFlow

TensorFlow is our preferred deep learning framework. TensorFlow's reliability, scalability, and complete toolbox offer us with everything we need to create effective machine learning models that are suited to our project's needs. TensorFlow enables us to fully realize the promise of deep learning for our application by providing cutting-edge algorithms and innovative approaches.

- Flutter, Visual Studio Code, and Firebase:

We leveraged Flutter, Visual Studio Code, and Firebase for building our mobile application. With Flutter's cross-platform features, we were able to design a single codebase for various platforms, assuring device compatibility and consistency. Visual Studio Code was our go-to programming environment, with functionality for developing, debugging, and deploying code easily. Meanwhile, Firebase supplied critical backend services like authentication, database management, and cloud storage, allowing us to seamlessly incorporate our machine learning models into our mobile application while maintaining scalability, security, and real-time data synchronization.

- Swagger:

We used Swagger to simplify API documentation and improve collaboration between frontend and backend teams. Swagger is an open-source software framework that helps developers design, construct, document, and consume RESTful web services effectively. Using Swagger, we were able to easily describe and test our APIs, assuring consistency and dependability throughout our application ecosystem.

Overall, Python, Google Collaboratory, Kaggle, TensorFlow, Flutter, Visual Studio Code, Firebase, and Swagger provided us with a powerful and adaptable platform for developing, deploying, and integrating machine learning models into our mobile app. With these tools and technologies at our disposal, we were able to meet our project needs efficiently and effectively, bringing our vision to reality.

4.5. Overall system diagram

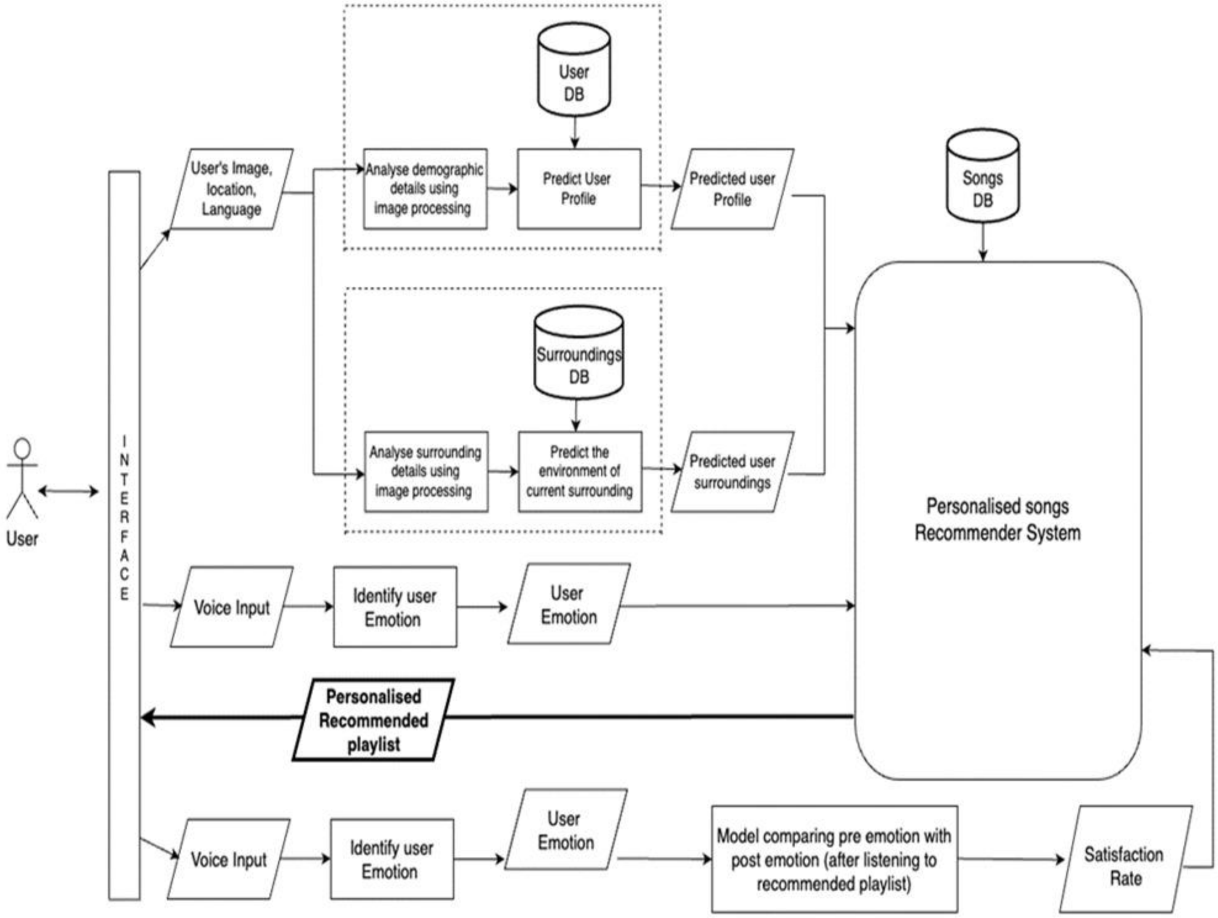

User DB
User's Image, location, Language
Analyse demographic details using image processing
Predict User Profile
Predicted user Profile
Songs DB
Surroundings DB
Analyse surrounding details using image processing
Predict the environment of current surrounding
Predicted user surroundings
User
INTERFACE
Personalised songs Recommender System
Voice Input
Identify user Emotion
User Emotion
Personalised Recommended playlist
Voice Input
Identify user Emotion
User Emotion
Model comparing pre emotion with post emotion (after listening to recommended playlist)
Satisfaction Rate

Figure 6:System overview Diagram

Figure 6 depicts our suggested system, which consists of four key sub-components that have been precisely constructed to ensure peak operation. The first component involves retrieving a user picture for processing, with a focus on obtaining critical user information such as age and gender. This data is critical input for the recommender system, allowing for individualized music selections that are highly matched to the user's demographic profile.

The second component of our system is collecting an image of the user's surroundings in order to determine current weather conditions. This information is then retrieved and communicated to the recommender system as additional input. By including meteorological information, our method improves the relevancy of music suggestions by matching them to the current climatic circumstances and user preferences.

Moving on to the third component, we collect voice input from the user to determine their current emotional state. This emotional data gives vital insight into the user's mood, allowing the recommender system to generate music selections that match their sentiments and emotions at the time.

After gathering all relevant inputs, our music recommendation models emerge into existence, creating a tailored playlist designed to enhance the user's enjoyment excursion. Using powerful machine learning algorithms, our system ensures that each recommendation is precisely tailored to the user's tastes, demographics, weather circumstances, and emotional state.

Furthermore, our system encompasses an evaluation mechanism that assesses the user's feelings after listening. This feedback loop allows for constant refining and development of the music recommendation process, ensuring that future recommendations are even more matched with the user's changing interests and emotional responses. Overall, our entire approach is designed to provide each user with a fully individualized and immersive music listening experience.

4.6. Component overview diagram

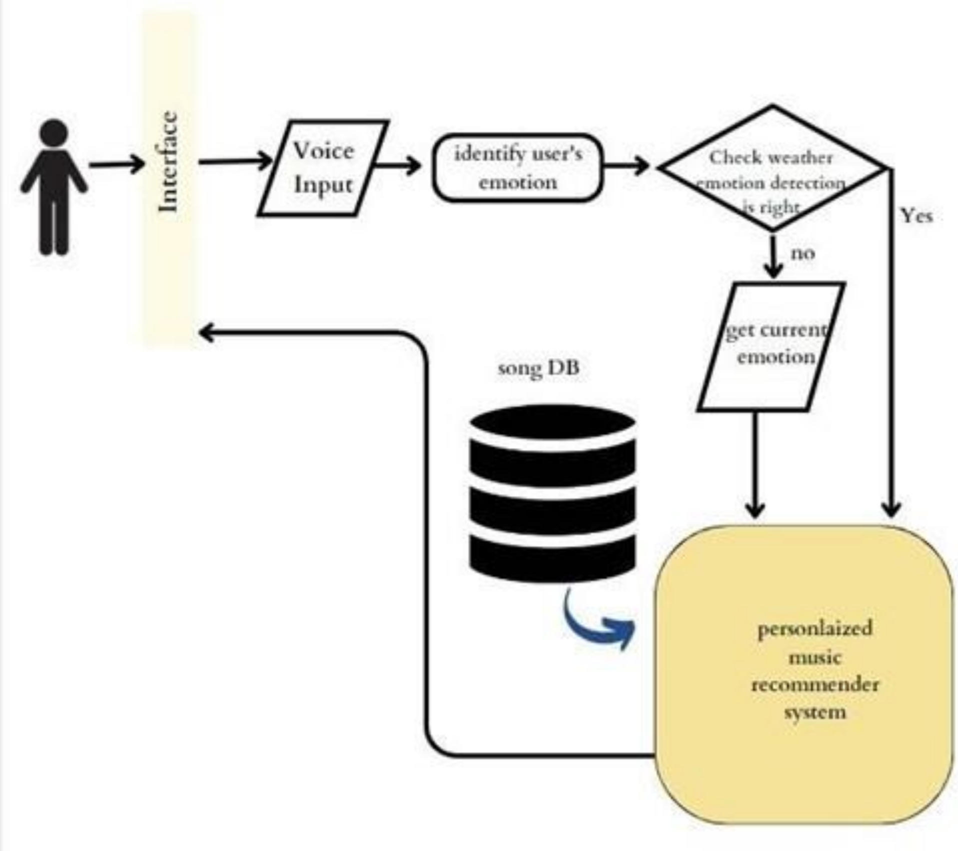

Interface
Voice Input
identify user's emotion
Check weather emotion detection is right
Yes
no
get current emotion
song DB
personlaized music recommender system

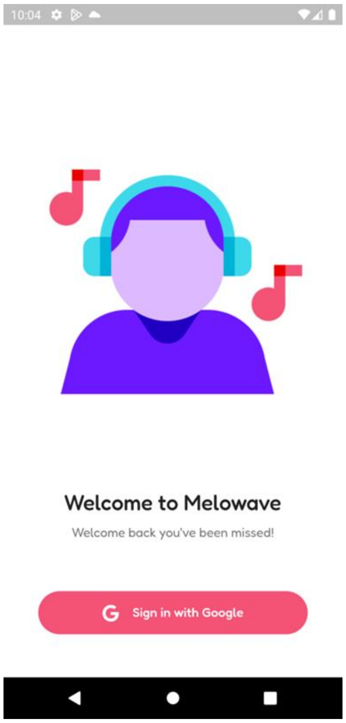
Figure 7:component overview diagram

The first part of this component involves getting the user's current voice clip, which is used as input for the emotion detection model. After acquiring the speech data, the system does frequency analysis to extract its attributes. This includes extracting pitch and applying noise cancellation. The system then uses deep learning algorithms to estimate the user's emotions. This forecast is then combined to provide a tailored playlist. Once anticipated, the emotion is smoothly integrated into the recommender system as input data. Using this expected mood, the recommendation algorithm creates a tailored playlist that reflects the predicted emotional state.

Finally, the personalized playlist is sent to the mobile app, where the user may access and enjoy carefully selected music substance based on their current emotional state.

4.7. User interfaces of the mobile application

From the user's viewpoint, the "Melowave" mobile application starts with a login box that prompts them to sign in using their Google account.(Figure 8).

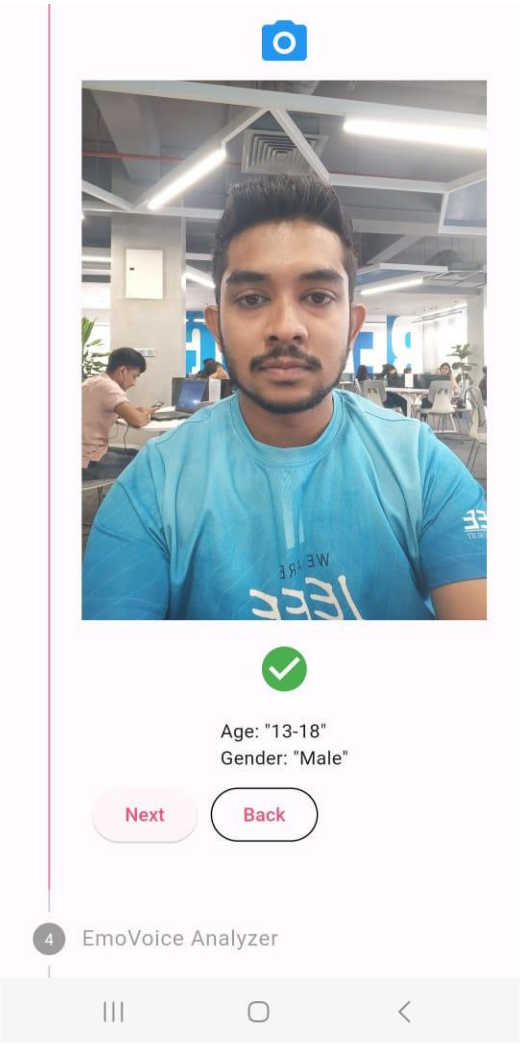


Welcome to Melowave

Welcome back you've been missed!

Figure 8:signup UI

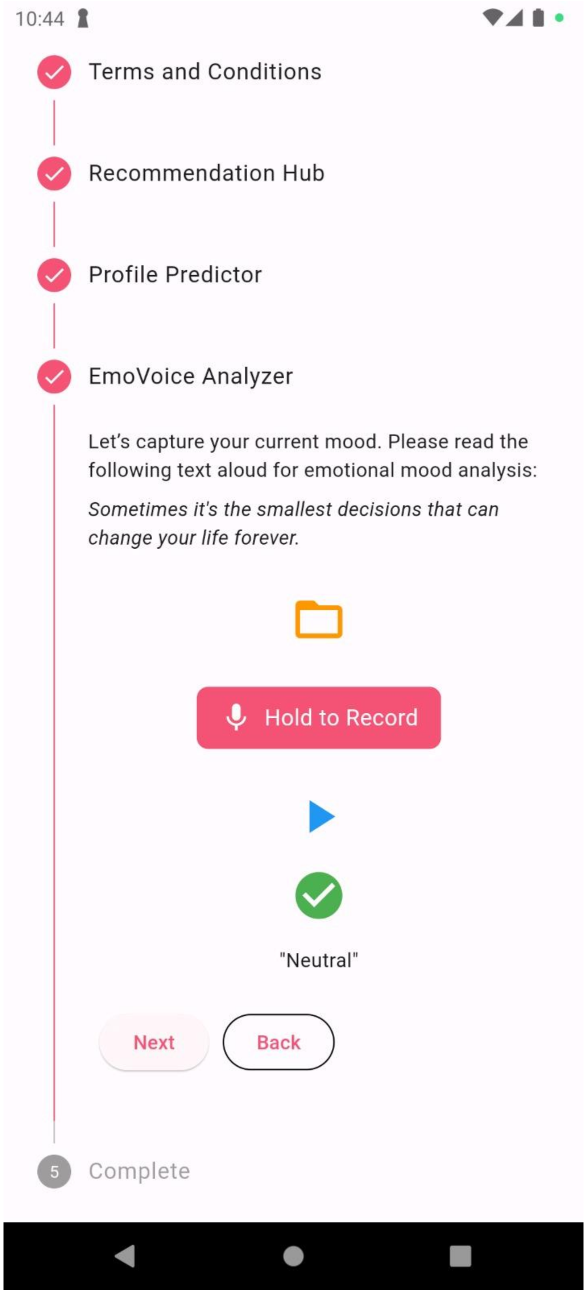
After logging in, the program guides the user through a series of stages using a stepper interface. In the first stage, a consent popup will display, seeking permission to access the camera, gallery, and microphone on the user's mobile phone. After granting the consent, the next step is to take a selfie with the device's camera, which captures the user's face traits for analysis.(figure 9)



Age: "13-18"
Gender: "Male"

Figure 9:selfie image UI

Moving on to the next phase, the program again uses the camera to record the user's surroundings and collect data about the current environment. Finally, in the third phase, the software invites the user to record a speech clip using the microphone so that the system can assess the user's present emotional state.(figure 10)



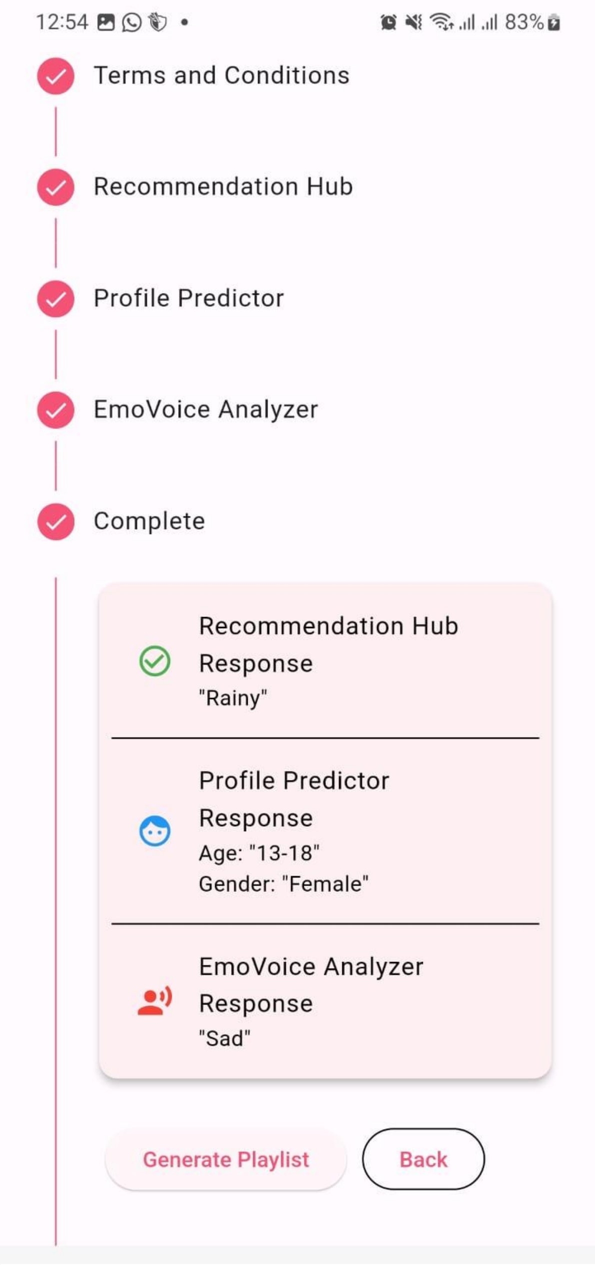
Let’s capture your current mood. Please read the following text aloud for emotional mood analysis:

Sometimes it's the smallest decisions that can change your life forever.

"Neutral"

Figure 10:voice recording UI

After hitting the "next" button, the trained machine learning models are deployed. These models include the age and gender prediction models, the weather forecast model, and the emotion prediction model. They evaluate the acquired data to provide outputs such as the user's age, gender, current weather conditions, and emotional state (see Figure 12).



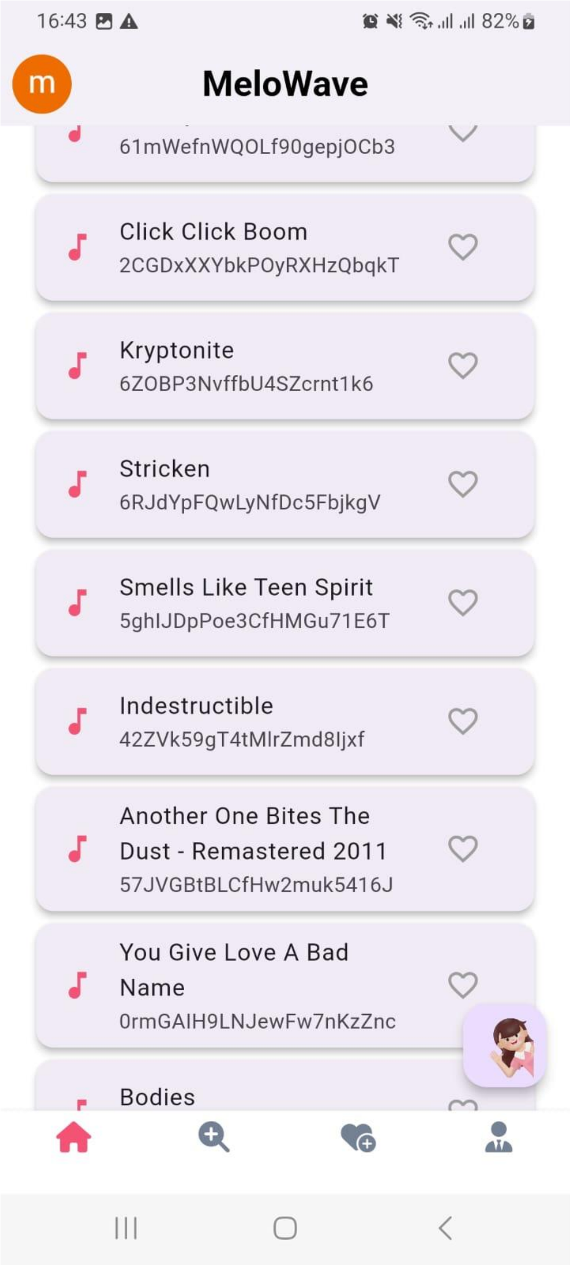
Recommendation Hub Response
"Rainy"

Profile Predictor Response
Age: "13-18"
Gender: "Female"

EmoVoice Analyzer Response
"Sad"

Figure 11:summarization of all predictions

Following that, the music recommender model generates a tailored playlist depending on the user's age, gender, current weather, and emotional state (see Figure 11). However, if the user is displeased with the music or their mood does not improve, the post-emotion categorization algorithm steps in.



16:43 82%

m **MeloWave**

61mWefnWQOLf90gepjOCb3

Click Click Boom
2CGDxXXYbkPOyRXHzQbqkT

Kryptonite
6ZOBP3NvffbU4SZcrnt1k6

Stricken
6RJdYpFQwLyNfDc5FbjkgV

Smells Like Teen Spirit
5ghIJDpPoe3CfHMGu71E6T

Indestructible
42ZVk59gT4tMlrZmd8ljxf

Another One Bites The Dust - Remastered 2011
57JVGBtBLCfHw2muk5416J

You Give Love A Bad Name
0rmGAIH9LNJewFw7nKzZnc

Bodies

Figure 12:songlist UI

5. BUDGET AND COMMERCIALIZATION

Given the widespread usage of music players in people's daily lives, this initiative has considerable economic potential. Individuals are eager to invest in an upgraded music player experience, showing significant economic value. However, with established industry giants like Spotify, iTunes, and Deezer already dominating the field, it's critical to develop a competitive and equitable pricing plan for music players.

While popular subscription models such as Spotify, Apple Music, and Deezer normally charge roughly $10 per month, some customers believe this price is too much for the value provided. As a result, a unique subscription model is presented to boost the commercialization of this mobile software.

Table 1 - Subscription Types

	Free version	**Paid version (<$10/month)**
Advertisements	Yes	No
Monthly charges for the users.	No Revenue will be generated from the advertisements showed to the user while the user is using the mobile application.	Yes Revenue will be generated from the monthly charges paid by the user.
Features	All features	All features

The final mobile application will be focused on different user groups; therefore, it will be marketed to each user group using different methods.

1. Young People – social media, gaming advertisements

2. Adults – worldwide news
3. Tech People – in-depth technical advertisements, new technologies, new trending applications

Below is the budget that has been planned for the project. Charges will be changed from time to time, and final charges will be based on the consumption of the resources used in the cloud environment.

Table 2 - Budget Plan3

Description	**Amount (USD)**
1. AWS Cloud database (S3) for facial images • To store user images collected through mobile app.	0.023 per GB / Month
2. AWS Cloud database (EFS) for user demographic data. • To store demographic data of the users.	0.30 per GB / Month

3. AWS glacier to store User logging from the mobile application.	Storing = $0.004 per GB / Month
	Retrieving = $0.01 per GB

4. Paper Publications and documentation.	50 - 100

6. TEST PLAN

Testing for the proposed system will take place at various phases of the project, assisting in the detection of issues within each component and allowing for independent resolution rather than tackling the entire project at once. As a result, the testing strategy will include several phases and processes.

Unit Testing:

Individual unit testing will be performed on each part, including the face image classification model and the music recommendation model. This method enables the isolation and correction of faults inside each part. Researchers will focus on two main dimensions:

1) Performance testing for the component: Evaluating how effectively each component accomplishes its intended function.
2) Component accuracy testing involves evaluating the precision and reliability of each component's outputs.

Integration Testing:

The integration of the components will be a critical step in this research effort. Because integration can introduce severe flaws into the system, components will be merged one at a time and tested concurrently.

Final testing

Final testing ensures optimal system performance and eliminates potential faults. The end result will be tested using a variety of test scenarios and sample data sets. During the final step of testing, the mobile application will be delivered to chosen users and feedback will be collected. Users will rate the mobile application's user experience, and depending on their feedback, changes will be made to improve the user interface, therefore improving the overall user experience for end users.

6.1. Test cases.

Table 1- Test case 01 – verify the expected response time

Test Case ID	01
Test Case	Verify the response time for given voice clip
Test scenario	Ensure the output's response time
Input	Voice clip
Expected output	Within 30 seconds, the output will be given
Actual output	Within 30 seconds, the output will be given
Status (Pass / Fail)	PASS

Table 1 - Test case 02 – analyze the voice clip's frequency.

Test Case ID	02

Test Case	Give voice clip and Analysis
Test scenario	Ensure the given voice clip's frequency matches with the emotion.
Input	A voice clip
Expected output	Successfully identified current emotion
Actual output	Successfully identified current emotion
Status (Pass / Fail)	PASS

Table 2 - Test case 03 – give happy based emotion voice clip .

Test Case ID	03
Test Case	Give voice clip based on happy state
Test scenario	Validate that given voice clip for feature extraction
Input	Give happy state-based voice clip

Expected output	Predict voice clip as a happy state
Actual output	Predict voice clip as a happy state
Status (Pass / Fail)	PASS

Table 4 - Test case 04 – give sad based emotion voice clip.

Test Case ID	04
Test Case	Give voice clip based on sad state
Test scenario	Validate that given voice clip for feature extraction
Input	Sad state-based voice clip
Expected output	Predict voice clip as a sad state

Actual output	Predict voice clip as a neutral state
Status (Pass / Fail)	FAIL

Table 5 - Test case 05 – Give low pitch based voice clip

Test Case ID	04
Test Case	Give low pitch-based voice clip
Test scenario	Validate that the system accurately analyze low pitch voice clips
Input	Predict emotion correct
Expected output	Predict voice clip accurately
Actual output	not Predict voice clip accurately

Status (Pass / Fail)	FAIL

7. RESULTS AND DISCUSSION

The results of thorough testing demonstrate the resilience of our technology. Throughout this phase, our major goal was to test the system's capacity to accurately recognize emotions.

In my portion, I developed a complex mechanism for predicting the user's present emotion. The technique was tested with 50 people, and 45 of them made commendably accurate predictions. Furthermore, the model's prediction based on frequency produced good results. Thus, despite the dataset being in English, it was also excellent at distinguishing emotions in Sinhala.The below diagram shows training and validation accuracies(figure 13).

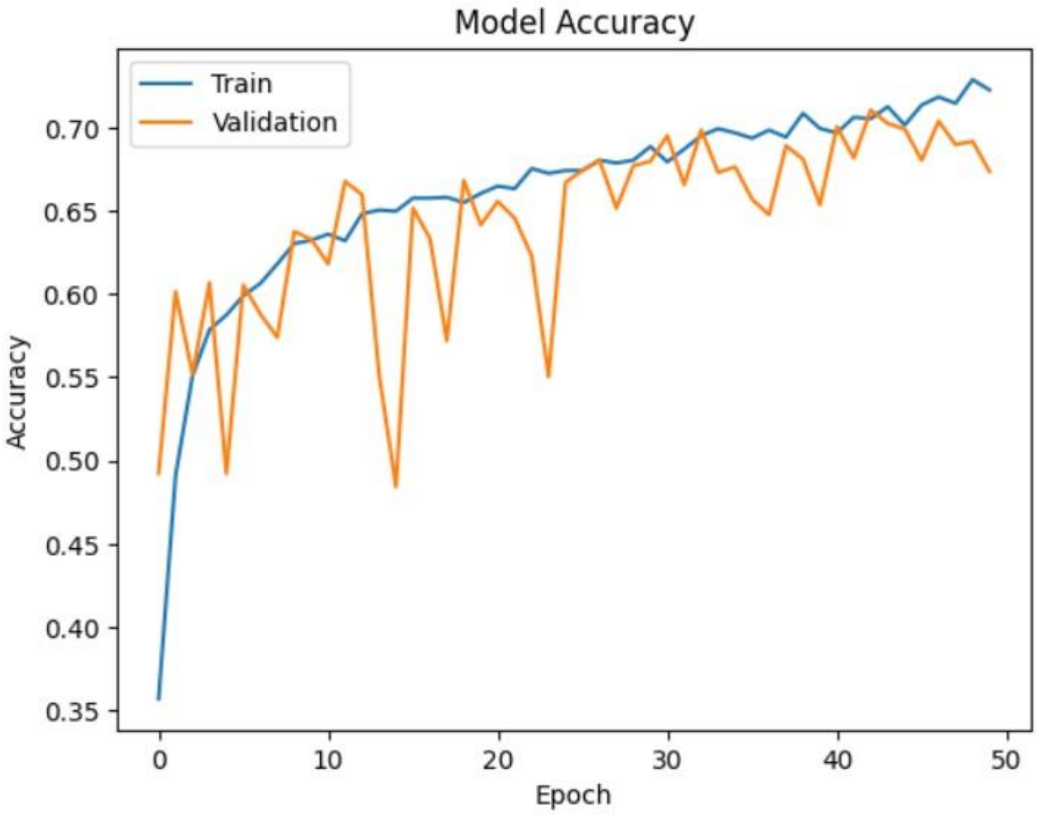


Figure 13:Training and validation accuracies

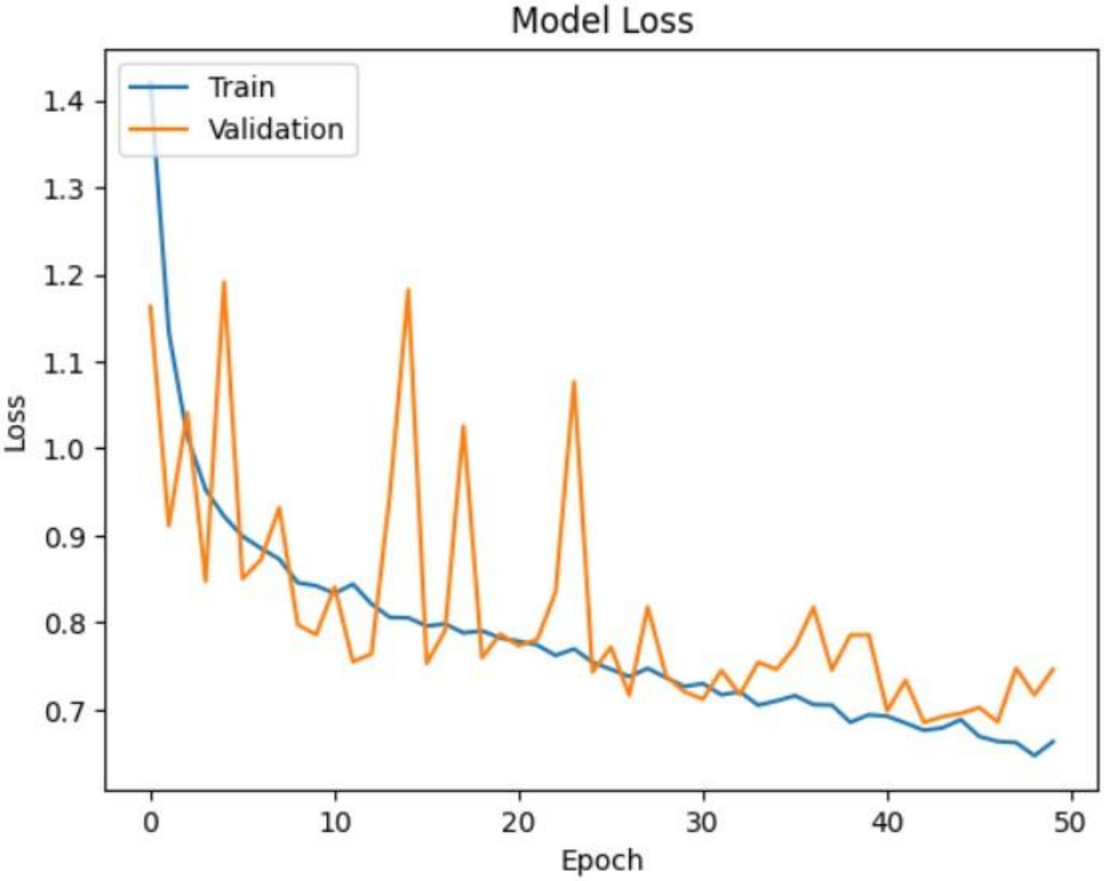


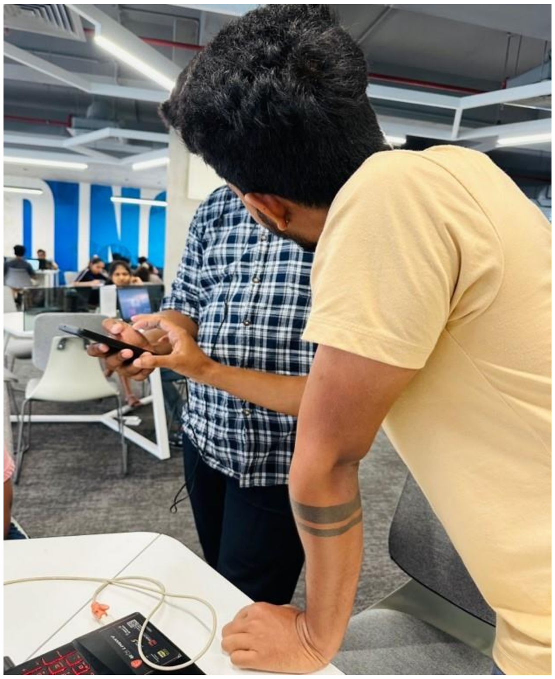
Figure 14:model loss

The Adam optimizer is used to reduce the loss function when training neural networks.

7.1. Fieldvisit And Feedbacks

Implementing and integrating the mobile application was only the first step, since user feedback was still required for testing and improvement. Field visits and feedback were critical in assuring the success of our mobile application. We interacted directly with our target consumers during field trips, examining their habits, preferences, and pain points in real-world circumstances. This direct experience gave us great insights into how our program was used and how it could be improved to better meet user demands.

Furthermore, gathering feedback from users allowed us to gain direct insight into their experiences, preferences, and ideas for development. Integrating this data into the development process allowed for iterative refining and optimization of our mobile application, ensuring that it remains relevant, useful, and user-friendly. Finally, visits and feedback were critical tools for understanding user requirements, verifying design decisions, and providing a mobile application that actually connected with its target audience.During our field visit, we identified several flaws and areas for improvement (Figure).



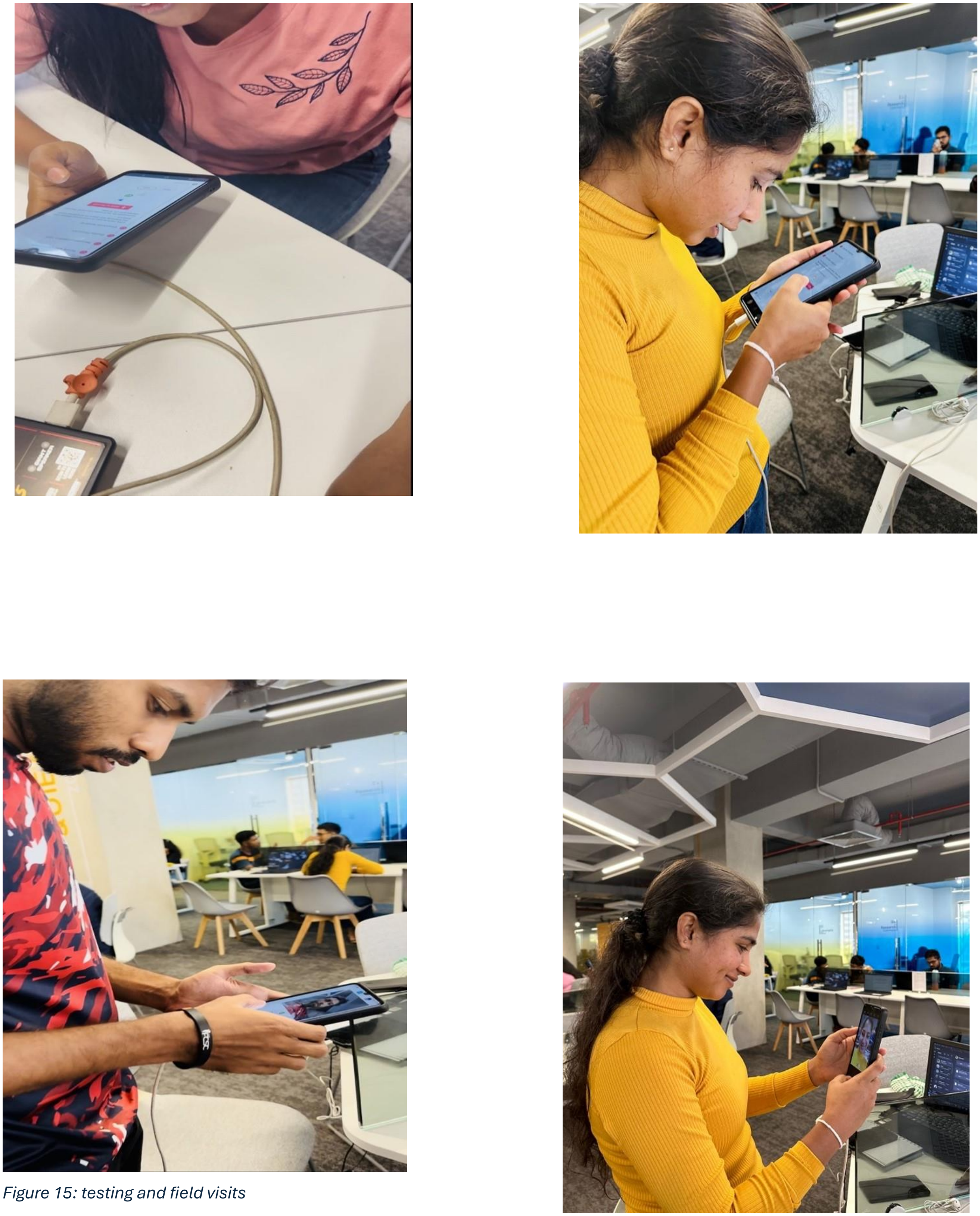
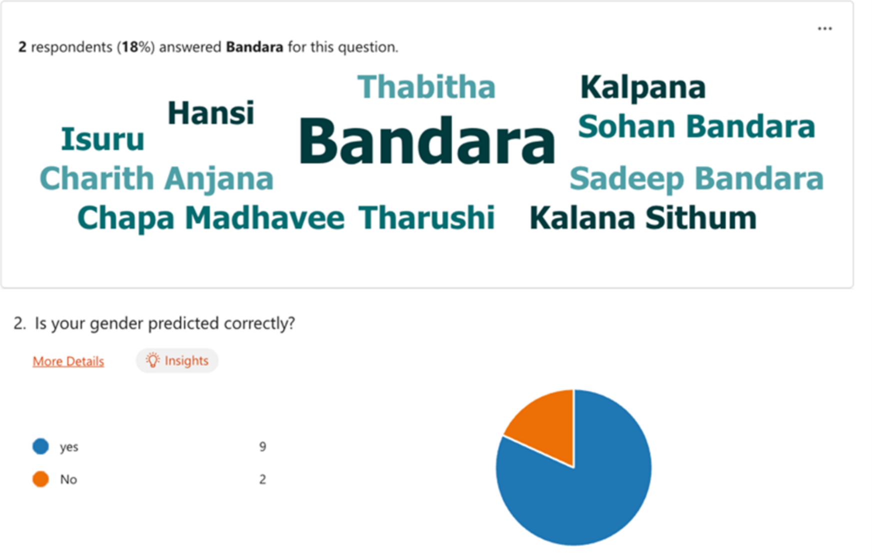


Figure 15: testing and field visits

Feedback is a vital component for the success of our mobile application, providing valuable insights into user experiences, preferences, and areas for improvement. Actively soliciting and responding to user feedback enables us to improve the overall user experience, quickly identify and fix bugs and technical glitches, validate features and functionality, tailor music recommendations to user preferences, and foster user engagement and loyalty.

Using feedback as a guiding force enables us to constantly develop and improve our product, increasing user happiness and ensuring its long-term survival in the market.

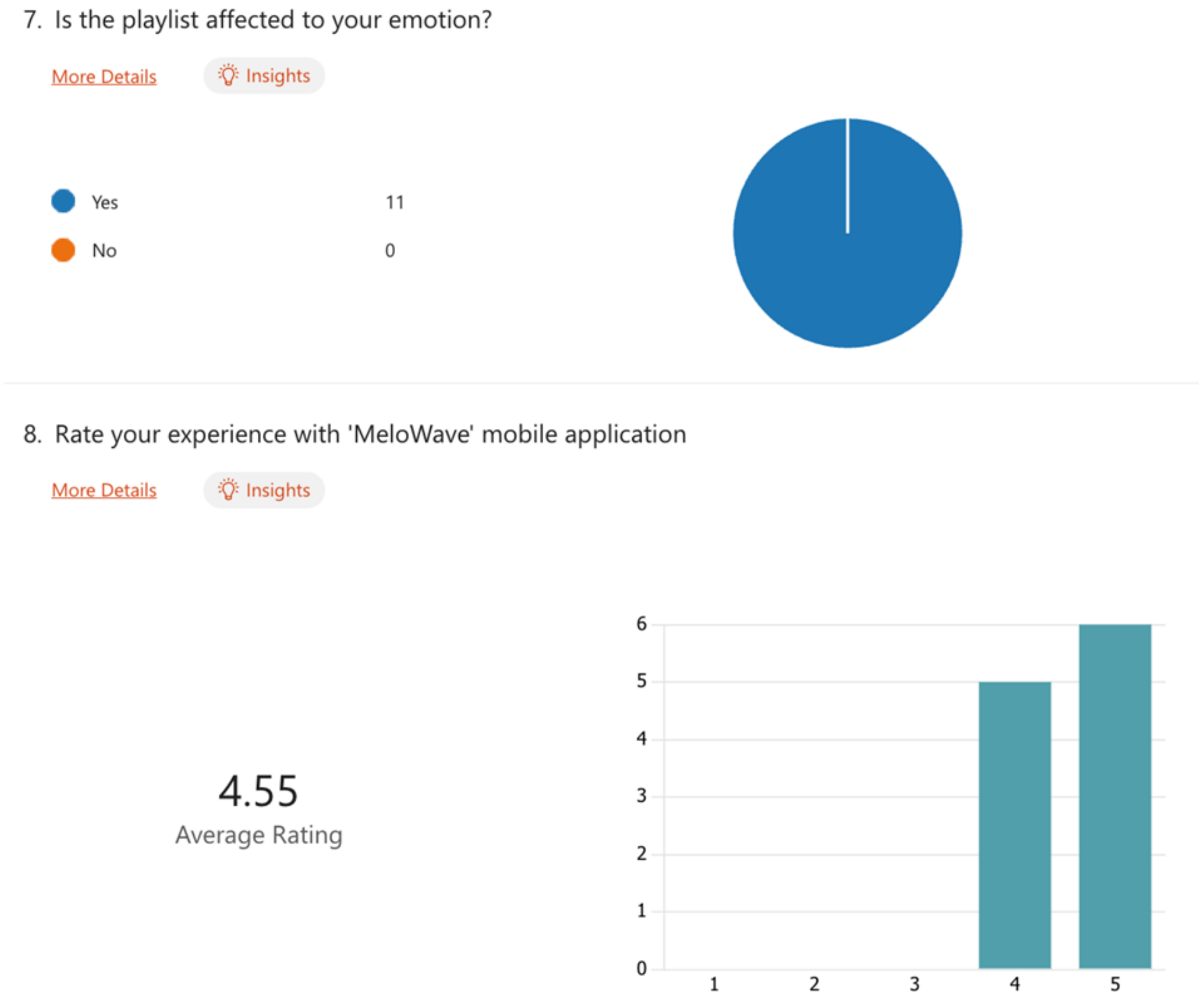
Here are some user feedback what we have received.(figure 16)



2 respondents (**18%**) answered **Bandara** for this question.

2. Is your gender predicted correctly?

● yes	9
● No	2



7. Is the playlist affected to your emotion?

More Details 💡 Insights

● Yes	11
● No	0

8. Rate your experience with 'MeloWave' mobile application

More Details 💡 Insights

4.55
Average Rating

Figure 16:user Feedbacks

7.2. Discussion

The Melowave mobile application indicates a substantial leap in speech frequency analysis, with a primary focus on reliably recognizing various emotional states. Recurrent Neural Networks (RNNs), in combination with the TensorFlow and Keras frameworks, provide a solid basis for effective emotion identification within the app. However, one of the most difficult issues encountered during the development process was model training, which was exacerbated by the vast quantity of the voice dataset. The restricted resources supplied by Colab made it difficult to efficiently train such a large dataset. Despite these limitations, significant progress was achieved in training the model. The consistent gain in accuracy with each training iteration is a good sign, even if the initial training accuracy of 72.24% and validation accuracy of 67.36% are considered typical levels. attaining 100% accuracy in training speech datasets is intrinsically challenging; yet, substantial progress was made toward attaining good outcomes. Moving forward, additional refining and optimization of the model are required to improve its accuracy and performance. Moving forward, the model will require further modification and optimization to improve its accuracy and performance. Another constraint is a combination of age, gender, weather, and emotion prediction models, which may affect the system's performance. Due to the limited resources available, processing such a combination of criteria may take some time. To overcome this constraint, we suggest a unique solution: use our own computer as a server for backend processing activities. However, it is crucial to note that this method needs the mobile application to connect to the same internet network as the server, which may be a constraint in some cases. Despite this constraint, using our own computer as a server has the ability to overcome resource constraints and increase the efficiency of backend processing activities. Continued investigation and implementation of such new ideas will be critical for overcoming problems and enhancing the Melowave mobile app's performance in the future.

8. Conclusion

In conclusion, this research provides a thorough methodology for developing an emotion-driven music recommendation system using sophisticated machine learning algorithms and publicly available datasets. The suggested system seeks to improve user experience by providing tailored playlists based on emotional indicators derived from voice inputs. The methodology's technological practicality is proved using open-source tools, powerful machine-learning algorithms, and optimization methodologies for rapid real-time inference. The cost-effective deployment choices and possibilities for income generation through value-added services also contribute to economic feasibility. Ethical factors such as data privacy, fairness, and openness are prioritized to ensure that the technology is used responsibly and fairly. By addressing these critical features, the suggested technique is consistent with ethical principles and societal norms, opening the way for the creation of a dependable and socially acceptable emotion-driven music recommendation system. Moving forward, more study and development of the process will help to expand emotion-aware technology and increase user engagement and happiness with music recommendations.

9. REFERENCES

[1] K. R. Nambiar and S. Palaniswamy, "Speech Emotion Based Music Recommendation," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-6,doi: 10.1109/INCET54531.2022.9824457.

[2] K. S. Krupa, G. Ambara, K. Rai and S. Choudhury, "Emotion aware Smart Music Recommender System using Two Level CNN," 2020 Third International Conference on SmartSystems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 1322-1327, doi: 10.1109/ICSSIT48917.2020.9214164.

[3] V. Mounika and Y. Charitha, "Mood -Enhancing Music Recommendation System based onAudio Signals and Emotions," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 1766-1772, doi:10.1109/ICICT57646.2023.10134211

[4] V. P. Sharma, A. S. Gaded, D. Chaudhary, S. Kumar and S. Sharma, "Emotion-Based Music Recommendation System," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-5, doi: 10.1109/ICRITO51393.2021.9596276. keywords: {Art;Mood;Webcams;Music;Market research;Reliability;Recommender systems;Face Expression;Emotion;music;Recommendation;CNN model},

[5] A. J. Mabel Rani, M. N. S, N. M. Jothi Swaroopan and K. Hari Kumar, "Face Emotion Based Music Recommendation System Using Modified Convolution Neural Network," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10368948. keywords: {Mood;Convolution;Face recognition;Neural networks;Music;User experience;Real-time systems;Music Recommendation;Face Emotion;Deep Learning;Data Security;Modified Convolution neural network},

10. APPENDIX