**KATHFORD INTERNATIONAL COLLEGE OF**

**ENGINEERING AND MANAGEMENT**

Balkumari, Lalitpur



A

Minor Project

On

**“Facial Expression Based Music Recommendation System using CNN”**

**Project Members**

Aavash Baral (KIC077BCT001)

Gaurav Pyakurel (KIC077BCT017)

Adish Bhattarai (KIC077BCT006)

Dikshita Poudel (KIC077BCT016)

**DEPARTMENT OF COMPUTER AND ELECTRONICS AND COMMUNICATION ENGINEERING**

**LALITPUR, NEPAL**

**AUGUST, 2023**

**KATHFORD INTERNATIONAL COLLEGE OF**

**ENGINEERING AND MANAGEMENT**

Balkumari, Lalitpur



A

Minor Project

On

**“Facial Expression Based Music Recommendation System using CNN”**

[Subject Code: CT654]

**Project Members**

Aavash Baral (KIC077BCT001)

Gaurav Pyakurel (KIC077BCT017)

Adish Bhattarai (KIC077BCT006)

Dikshita Poudel (KIC077BCT016)

A PROJECT WAS SUBMITTED TO THE DEPARTMENT OF COMPUTER AND

ELECTRONICS, COMMUNICATION & INFORMATION ENGINEERING IN

PARTIAL FULLFILLMENT OF THE REQUIREMENT FOR THE BACHELOR’S

DEGREE IN COMPUTER ENGINEERING

**DEPARTMENT OF COMPUTER AND ELECTRONICS AND COMMUNICATION ENGINEERING**

**LALITPUR, NEPAL**

**AUGUST, 2023**

**KATHFORD INTERNATIONAL COLLEGE OF ENGINEERING AND MANAGEMENT**

# PAGE OF APPROVAL

**BALKUMARI, LALITPUR**

**DEPARTMENT OF COMPUTER AND ELECTRONICS, COMMUNICATION & INFORMATION ENGINEERING**

The undersigned certify that they have read, and recommended to this department for acceptance, a project report entitled “**Facial Expression Based Music Recommendation System using CNN”**, submitted by **Aavash Baral, Gaurav Pyakurel, Adish Bhattarai and Dikshita Poudel** in partial fulfilment of the requirements for the Bachelor’s degree in Computer Engineering.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor, name of Supervisor

Title

Name of the Department

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

External Examiner, name of External

Title

Name of the Organization, he/she belongs to

------------------------------------------------------------

Coordinator,

Name of Coordinator

Title

Name of the coordinating committee

**DATE OF APPROVAL**: Day.Month.Year

# COPYRIGHT

The authors have agreed that the Library, Department of Computer and Electronics, Communication and Information Engineering, Kathford International College may make this report freely available for inspection. Moreover, the author has agreed that permission for extensive copying of this project report for scholarly purpose may be granted by the supervisors who supervised the project work recorded herein or, in their absence, by the Head of the Department wherein the project report was done. It is understood that the recognition will be given to the author of this report and to the Department of Computer and Electronics, Communication and Information Engineering, Kathford International College in any use of the material of this project report. Copying or publication or the other use of this report for financial gain without approval of to the Department of Computer and Electronics, Communication and Information Engineering, Kathford International College and author’s written permission is prohibited.

Request for permission to copy or to make any other use of the material in this report in whole or in part should be addressed to:

Head

Department of Computer and Electronics, Communication and Information Engineering, Kathford International College of Engineering and Management

Lalitpur, Nepal

# ACKNOWLEDGEMENT

We extend our sincere appreciation to the Department of Computer and Electronics, Communication & Information Engineering, IOE, Kathford International College Of Engineering and Management, for granting us the opportunity to undertake a Minor Project as part of our curriculum. We are deeply grateful to our project supervisor, **Er. Jaydeep Shah**, for his unwavering support and guidance throughout the project duration.

Additionally, we wish to express our gratitude to **Er. Pankaj Kumar Jha**, Head of the Computer and Electronics, Communication & Information Department, for his support. Our heartfelt thanks also go to all our instructors for their insightful suggestions and motivational lectures, which greatly contributed to the completion of this project.

Finally, we would like to acknowledge our colleagues for their valuable feedback and suggestions during the project.

# ABSTRACT

This project focuses on the relationship between music and emotions, presenting a Facial Expression Music-Based Recommendation System utilizing Convolutional Neural Networks (CNN). With a testing accuracy of 81.71%, the system employs a dataset comprising 35,887 grayscale images (48x48 pixels) sourced from FER2013. The model undergoes training with a 9:1 split for training and testing/validation, integrating OpenCV2 Cascade Classifier for frontal face detection and a CNN for emotional state recognition. The system's output is presented via a pygame-based Graphical User Interface (GUI), facilitating the playback of recommended music tracks.

Top of Form

**Keywords**

Emotion Detection, Convolutional Neural Network, Pygame, OpenCV, FER2013

TABLE OF CONTENTS

[PAGE OF APPROVAL ii](#_Toc159615836)

[COPYRIGHT iii](#_Toc159615837)

[ACKNOWLEDGEMENT iv](#_Toc159615838)

[ABSTRACT v](#_Toc159615839)

[LIST OF FIGURES viii](#_Toc159615840)

[LIST OF ABBREVIATIONS ix](#_Toc159615841)

[1. INTRODUCTION 1](#_Toc159615842)

[1.1 Background 1](#_Toc159615843)

[1.2 Problem Statement 2](#_Toc159615844)

[1.3 Objectives 2](#_Toc159615845)

[1.4 Scopes 3](#_Toc159615846)

[2. LITERATURE REVIEW 4](#_Toc159615847)

[3. REQUIREMENT ANALYSIS 5](#_Toc159615848)

[3.1 Feasibility Study 5](#_Toc159615849)

[3.1.1 Technical Feasibility 5](#_Toc159615850)

[3.1.2 Economic Feasibility 5](#_Toc159615851)

[3.1.3 Operational Feasibility 6](#_Toc159615852)

[3.1.4 Schedule Feasibility 6](#_Toc159615853)

[3.1.5 Legal and Ethical Feasibility 6](#_Toc159615854)

[3.2 Software Requirement Specification 7](#_Toc159615855)

[3.2.1 Data Requirements: 7](#_Toc159615856)

[3.2.2 Functional Requirements: 7](#_Toc159615857)

[3.2.3 Non-Functional Requirements: 8](#_Toc159615858)

[3.2.4 Security Requirements: 8](#_Toc159615859)

[4. PROJECT METHODOLOGY 9](#_Toc159615860)

[4.1 Development Model 9](#_Toc159615861)

[4.2 Block Diagram 10](#_Toc159615862)

[4.2.1 Image Source 10](#_Toc159615863)

[4.2.2 Face Detection and Face Normalization: 10](#_Toc159615864)

[4.2.3 Feature Extraction: 11](#_Toc159615865)

[4.2.4 Emotion Classification using CNN: 11](#_Toc159615866)

[4.2.5 Emotion-Music Mapping and Playlist Recommendation: 11](#_Toc159615867)

[5. IMPLEMENTATION PLAN 12](#_Toc159615868)

[5.1 Activity Diagram 12](#_Toc159615869)

[5.2 Use Case Diagram 13](#_Toc159615870)

[5.3 Data Flow Diagram 14](#_Toc159615871)

[5.4 Sequence Diagram 15](#_Toc159615872)

[6. WORK COMPLETED 16](#_Toc159615873)

[7. WORK REMAINING 17](#_Toc159615874)

[REFERENCES 18](#_Toc159615875)

# LIST OF FIGURES

[Figure 4.1: Iterative and Incremental Development Model 9](#_Toc157428527)

[Figure 4.2: Block Diagram of Proposed System 10](#_Toc157428528)

[Figure 5.1: Activity Diagram 12](#_Toc157428529)

[Figure 5.2: Use Case Diagram 13](#_Toc157428530)

[Figure 5.3.1: Level 0 DFD 14](file:///D:\Desktop\Minor%20Project%20Proposal.docx#_Toc157428531)

[Figure 5.3.2: Level 1 DFD 14](file:///D:\Desktop\Minor%20Project%20Proposal.docx#_Toc157428532)

[Figure 5.4 Sequence Diagram 15](#_Toc157428533)

# LIST OF ABBREVIATIONS

AI Artificial Intelligence

CNN Convolutional Neural Network

FER2013 Facial Expression Recognition 2013

ML Machine Learning

UI User Interface

# INTRODUCTION

## Background

Music is a powerful medium that is intricately interconnected with human emotions, capable of uplifting one's mood. Evolutionary speculations have tended to focus on single-source causes such as music as an indicator of biological ﬁtness, music as a means for social and emotional communication, music as social glue, music as a way of facilitating caretaker mobility, music as a means of tempering anxiety about mortality, music as escapism or transcendental meaning, music as a source of pleasure, and music as a means for passing time [1]. The rise of various digital musical platforms has led to an increased demand for personalized music systems. As a result, consumers now expect music services that cater to their individual preferences, offering tailored playlists, recommendations, and a seamless user experience.

In recent years, advances in artificial intelligence and deep learning have opened new possibilities for improving music recommendation systems. Musical preferences have been demonstrated to be highly related to personality traits and moods [2]. To introduce an innovative approach that utilizes facial emotion recognition for automatic music selection based on the user's emotional state, we propose an innovative approach that leverages facial emotion recognition to automatically play music based on the user's emotional state. Facial expressions are a powerful indicator of human emotions and can provide valuable cues for understanding how individuals feel at any given moment. By employing Convolutional Neural Networks (CNN), a class of deep learning algorithms well-suited for image recognition tasks, we can detect and analyze facial expressions in real-time.

The FER2013 dataset, a widely used benchmark for facial emotion recognition, will serve as a valuable resource for training our CNN model. This dataset contains a diverse range of facial expressions labeled with emotions such as happiness, anger, sadness and neutrality. By utilizing this dataset, we can build a robust emotion detection model that can accurately identify users' emotional states through their facial expressions.

## Problem Statement

The current landscape of music selection systems lacks a comprehensive and efficient approach to cater to users' real-time emotional states, resulting in suboptimal music recommendations. Existing methods rely on manual input, wearable devices, or audio-based classification, which may not accurately capture the user's emotional context at a given moment. This limitation leads to a less engaging and personalized music experience for users.

Traditional music selection techniques based on manual sorting or audio features fail to consider the dynamic nature of human emotions. Wearable devices, though capable of monitoring certain physiological indicators, may not capture facial expressions, which are powerful indicators of emotions. Additionally, audio-based classification might not accurately discern the user's emotional state in real-time.

## Objectives

The main objective of this project is to develop an automatic music selection system that utilizes facial emotion recognition through Convolutional Neural Networks (CNN) to enhance the user's music experience based on their real-time emotional states. The project aims to achieve the following specific objectives:

* To create a emotion detection model using CNN
* To create a music player that plays songs based on emotion

## Scopes

* **Personalized music recommendations**: This system has the potential to personalize music recommendations to a much greater extent than traditional methods. By analyzing users' facial expressions in real-time, it can adapt to their changing moods and preferences, creating a more immersive and emotionally resonant listening experience.
* **Accessibility for non-verbal users**: This system could be a valuable tool for non-verbal individuals who have difficulty expressing their emotions verbally. By using facial expressions as input, it can open up a new avenue for them to discover and enjoy music that resonates with their mood.
* **Mental health applications**: Monitoring facial expressions in conjunction with music preferences could contribute to research and development of tools for mental health monitoring and interventions. By identifying patterns in emotional responses to music, this system could potentially serve as an early warning system for mood disorders or provide personalized music therapy recommendations.

# LITERATURE REVIEW

In recent years, the convergence of AI and music recommendation systems has ushered in a new era of personalized and engaging music experiences for users. Extensive research has explored the potential of utilizing facial expression analysis to enhance the relevance and emotional resonance of music recommendations.

Scholars have investigated the intricate relationship between facial expressions and emotional responses to music. Research by Yang et al. [3] demonstrated a strong correlation between specific facial expressions and distinct emotional categories, suggesting the feasibility of utilizing facial cues to infer emotional preferences. Preema et al. [4] stated that manually segregating the list of songs associated, generating acceptable playlist supported an individual’s emotions could be a terribly tedious, time overwhelming, intensive and upheld task.

The paper by Ramya Ramanathan et al. [2] presents an innovative contribution in the domain of intelligent music players by employing emotion recognition. The study delves into the development and implementation of an intelligent music player system that harnesses the capabilities of emotion recognition technology. This work explores the intricate relationship between user emotional states and musical preferences, advocating for a more personalized and engaging music listening experience.

In summary, the existing literature underscores the viability of AI-driven music recommendation systems that incorporate facial expression analysis. The studies mentioned above collectively lay the foundation for our project, aiming to enhance the user experience by providing personalized music suggestions aligned with users' emotional states.

# REQUIREMENT ANALYSIS

## 3.1 Feasibility Study

### 3.1.1 Technical Feasibility

* **Availability of Technology and Data**

The technology required for the emotion-based music recommendation app, including ML libraries like TensorFlow, Keras, and PyTorch, is widely available and accessible. PyQt, as a Python library, facilitates the creation of a user-friendly graphical user interface, enabling users to interact seamlessly with the application. The FER2013 dataset, containing labeled facial expression images, is publicly available and serves as the primary dataset for training the emotion recognition model.

* **Expertise and Skill Set**

Adequate expertise in deep learning, computer vision, and software development is essential. The project team possesses the required skills or can acquire them through training and collaboration.

### 3.1.2 Economic Feasibility

* **Development Costs**

To minimize development costs, we will leverage open-source tools, libraries, and frameworks. Our primary investment will be the time and effort contributed by team members.

* **Hardware Infrastructure**

To access advanced hardware resources without significant expenses, we will utilize Google Colab, a free cloud-based platform that offers GPU support for machine learning tasks.

### 3.1.3 Operational Feasibility

* **Integration and Compatibility**

Our emotion-based music recommendation system's success depends on seamlessly integrating the CNN-based emotion detection model and PyQt user interfaces. The CNN model will operate in the backend, enabling real-time facial emotion analysis, while the PyQt-based interface will offer a user-friendly and visually appealing experience. Compatibility across devices and platforms, including laptops, desktops, and mobile devices, will be ensured.

* **User Acceptance**

User feedback and testing will assess the system's usability and whether users find the automatic music selection based on facial emotions to be beneficial and enjoyable.

### 3.1.4 Schedule Feasibility

* **Realistic Project Timeline**

We will set a realistic and achievable project timeline, breaking down the development process into manageable phases. Each phase will have specific milestones and deadlines to monitor progress effectively.

* **Task Prioritization**

Tasks will be prioritized based on their dependencies and criticality, ensuring that essential components are addressed first. This approach will enable the completion of foundational elements before moving to subsequent phases.

### 3.1.5 Legal and Ethical Feasibility

Compliance with data protection regulations and ensuring user data privacy is crucial. Proper consent from users for data collection and usage must be obtained.

## 3.2 Software Requirement Specification

### 3.2.1 Data Requirements:

* **Type:** Images of human faces expressing various emotions.
* **Source:** FER2013 dataset (preferred) or any other dataset containing labeled facial expressions (neutral, happy, sad, angry, surprised, fearful, disgust).
* **Format:** Grayscale images of fixed size (e.g., 48x48 pixels).
* **Volume:** 35,900 images for training and sufficient images for validation and testing.
* **Additional Data:** Music metadata including: genre, tempo, mood, lyrics (optional).

### 3.2.2 Functional Requirements:

* **Facial Expression Recognition:**
  + Capture real-time webcam footage or analyze pre-recorded images.
  + Employ a pre-trained CNN model (e.g., using FER2013) to predict the user's dominant facial expression from the visual input.
  + Achieve an accuracy of at least 75% in recognizing the seven basic emotions.
* **Music Recommendation:**
  + Based on the recognized emotion, recommend a song or playlist that matches the user's current mood.
  + Utilize a pre-defined mapping between emotions and musical characteristics (e.g., happy - upbeat tempo, sad - slow tempo, etc.).
* **User Interface:**
  + Provide a user-friendly interface for accessing the system.
  + Display captured webcam footage or uploaded images.
  + Show recognized emotions with confidence scores.
  + Display recommended songs with relevant information (title, artist, genre, etc.).

### 3.2.3 Non-Functional Requirements:

* **Performance:**
  + Real-time facial expression recognition with minimal latency.
* **Reliability:**
  + Accurate facial expression recognition even with slight variations in lighting and facial features.
  + Recommend relevant songs consistently based on recognized emotions.
* **Security:**
  + User privacy should be protected.
  + Webcam footage and personal data should be securely stored and transmitted.

### 3.2.4 Security Requirements:

* Implement secure protocols for data transmission and storage.
* Follow ethical guidelines for data collection and usage

# PROJECT METHODOLOGY

## Development Model

The Iterative and Incremental Development Model is suitable approach for the proposed system, as it perfectly fits our dynamic team and flexible project requirements. In this model, we can collectively contribute to different aspects of the system, allowing us to learn and experiment collaboratively. We can divide the project into smaller, manageable increments, with each increment adding new functionalities and improvements to the system.

Since we all have diverse skill sets and interests, the Iterative and Incremental approach enables us to work on different components based on our strengths. For instance, some team members can focus on facial expression analysis and emotion detection, while others can work on the UI part. As we progress through iterations, we can integrate our individual efforts to create a coherent and functional system.

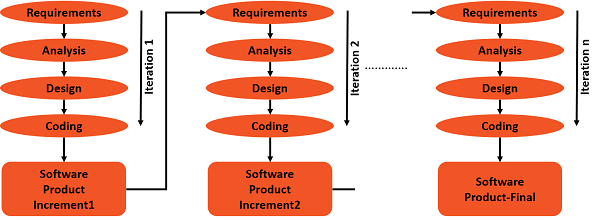


Figure 4.1: Iterative and Incremental Development Model

## Block Diagram

Figure 4.2: Block Diagram of Proposed System

### 4.2.1 Image Source

The system continuously captures live video frames (webcam feed) as its primary image source.

### 4.2.2 Face Detection and Face Normalization:

OpenCV2 Haarcascade algorithms then efficiently detect and isolate the frontal face region within each frame, ensuring emotion recognition focuses on relevant features.

To prepare the extracted face for accurate analysis, several key normalization techniques are applied:

* **Grayscale Conversion**: This reduces computational complexity and removes potential biases or distractions related to skin tone or color variations.
* **Scaling**: The face image is resized to a fixed dimension (48x48 pixels) to achieve consistency for the CNN model's input layer.
* **Cropping**: Focusing on the central area containing key facial features like eyes, nose, and mouth improves the extraction of emotion-related information.

### 4.2.3 Feature Extraction:

Meaningful features are extracted from the normalized face image to provide the CNN model with a concise and informative representation for emotion classification. These features encompass:

* **Geometric features**: Relative positions and distances of facial landmarks, analogous to understanding the spatial relationships between letters in a word.
* **Texture features**: Variations in pixel intensities and patterns, revealing wrinkles, lines, or muscle movements associated with certain emotions.
* **Statistical features**: Measurements like mean, standard deviation, or entropy across different regions of the face, capturing subtle changes in appearance.

### 4.2.4 Emotion Classification using CNN:

The extracted features are passed through the pre-trained convolutional neural network (CNN) model. This model, built upon a labeled dataset of facial expressions, compares the features against its internal knowledge base and outputs the predicted dominant emotion from the seven basic categories (neutral, happy, sad, angry, surprised, fearful, disgust).

### 4.2.5 Emotion-Music Mapping and Playlist Recommendation:

A pre-defined mapping database links each predicted emotion to its corresponding musical characteristics (tempo, genre, lyrics, mood keywords). Based on this mapping, the system dynamically generates a playlist of suitable songs from a pre-classified music library, considering additional factors like user preferences and song popularity.

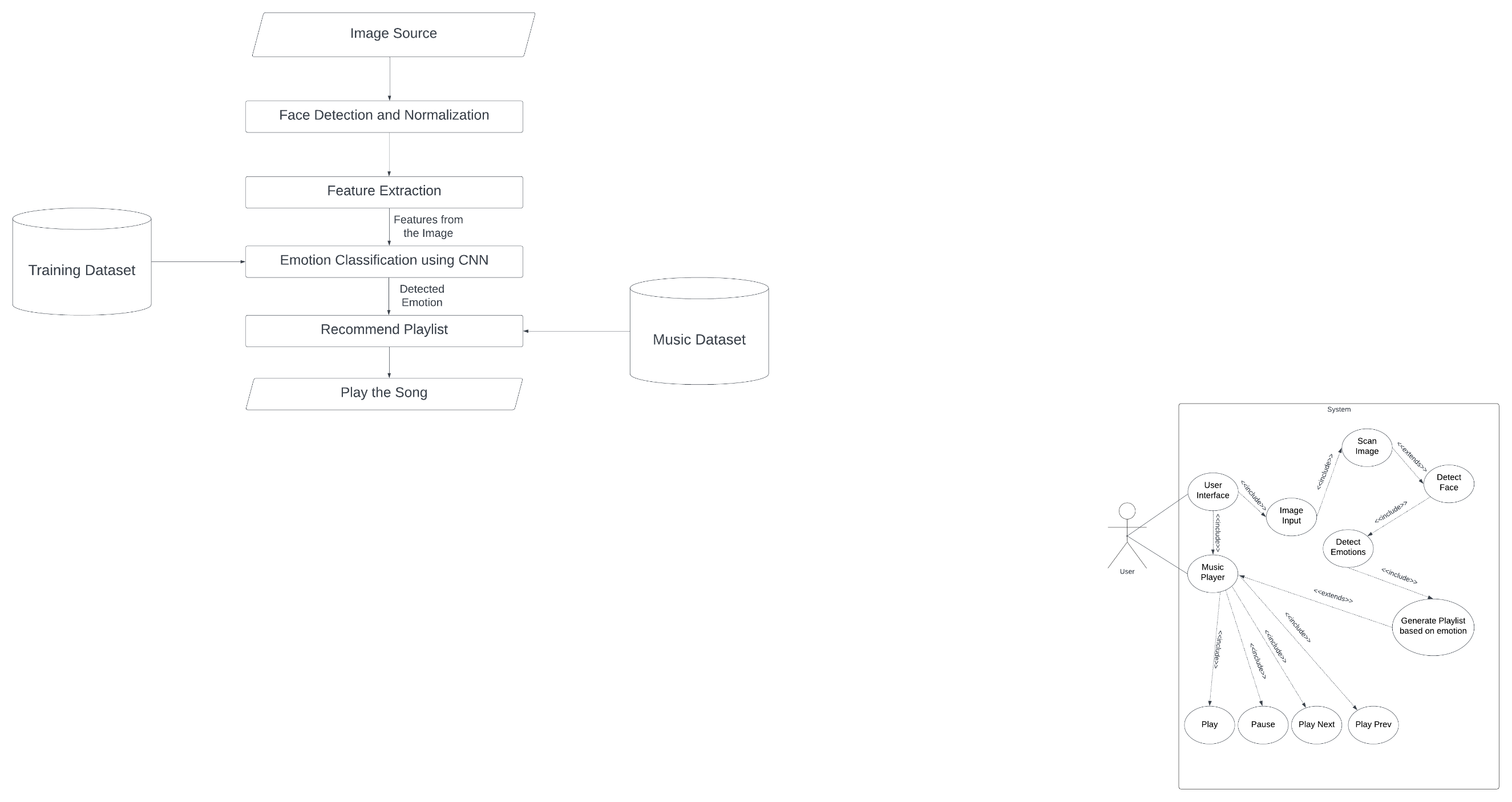
# IMPLEMENTATION PLAN

In our project,we've taken a function-focused approach.This helped us organize tasks clearly and made our findings easier to understand and use.

## Activity Diagram

Figure 5.1: Activity Diagram

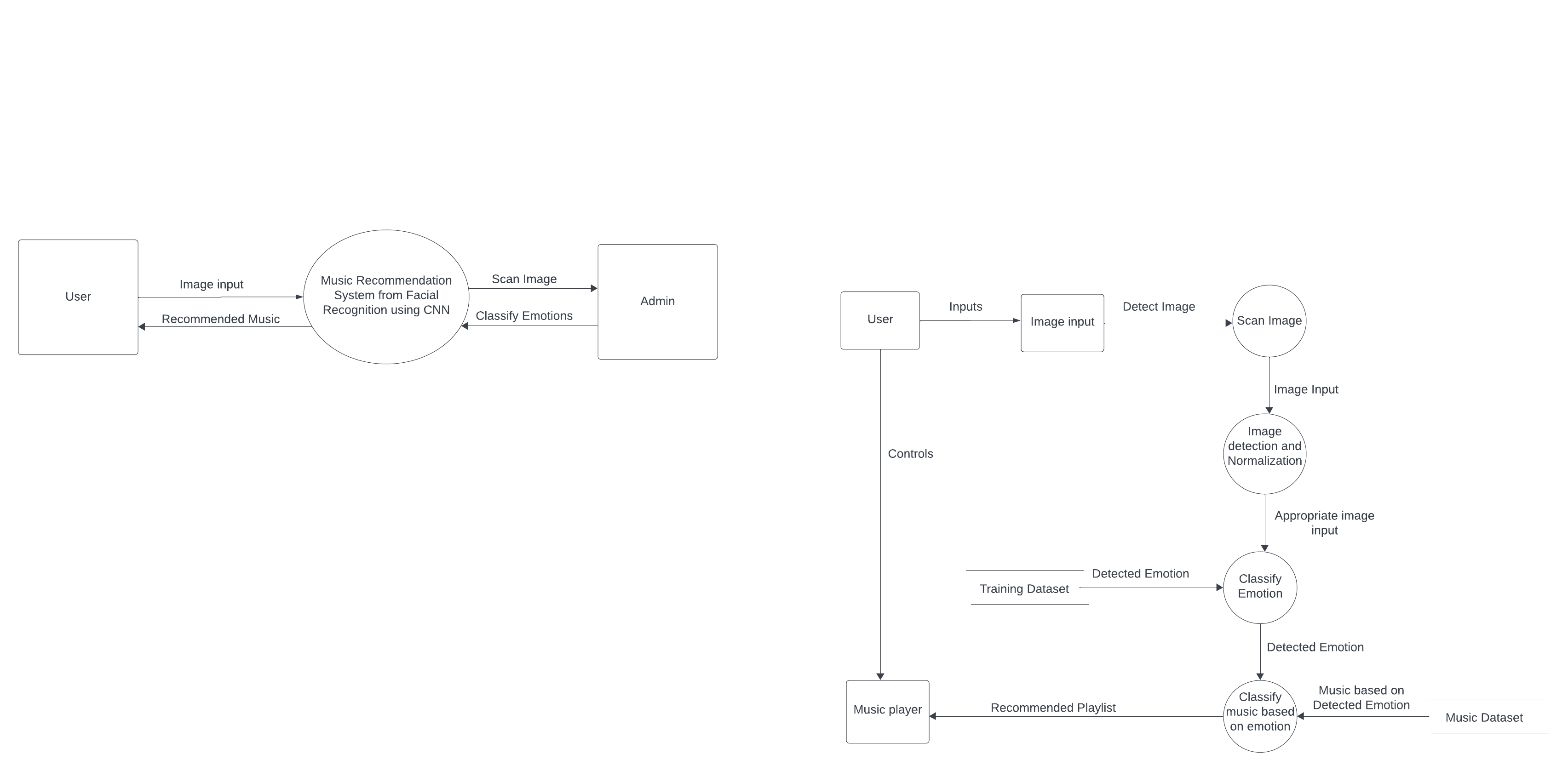
## Use Case Diagram



## 

Figure 5.2: Use Case Diagram

## Data Flow Diagram



## 

### 

Figure 5.3.1: Level 0 DFD

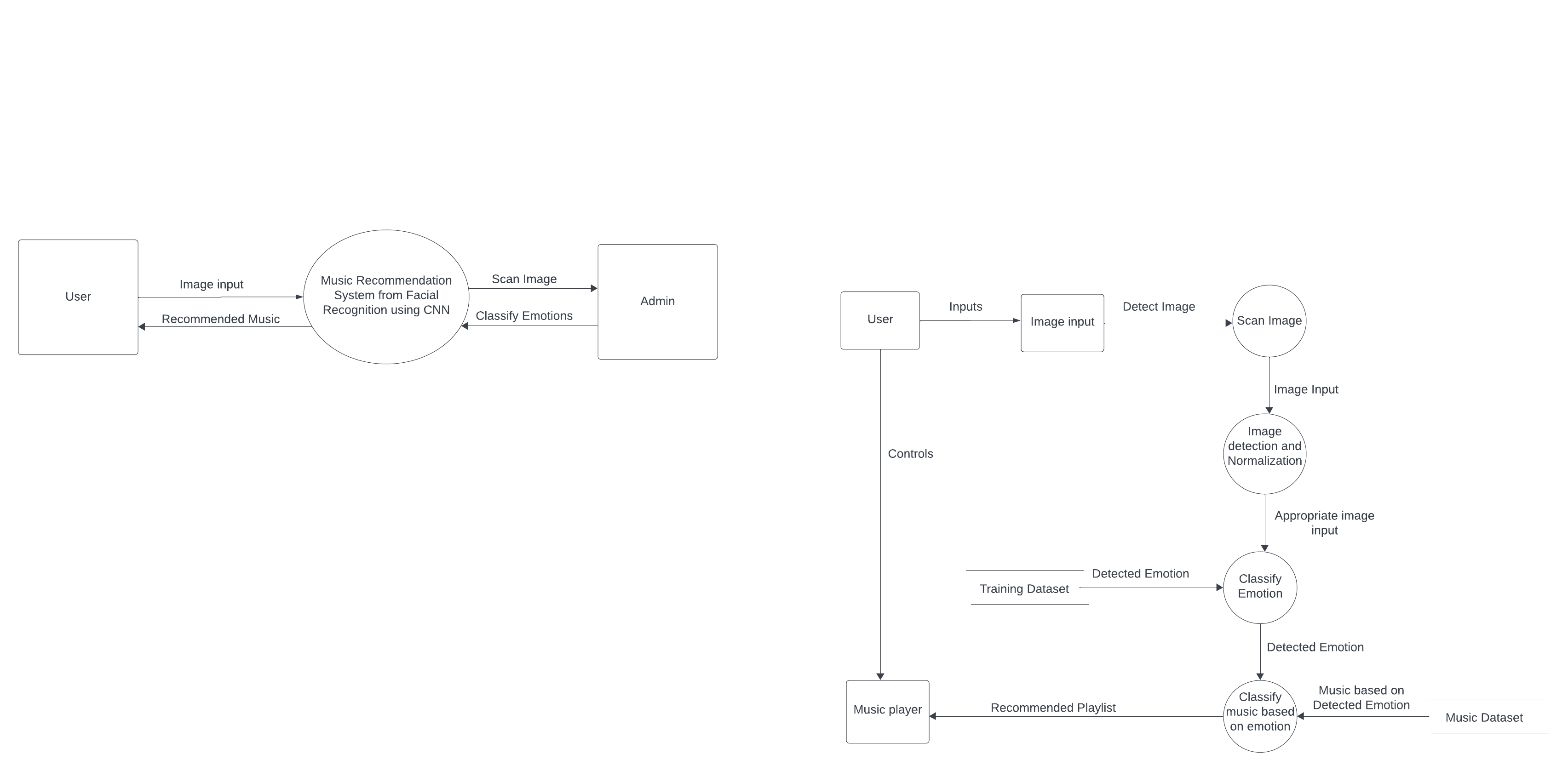


Figure 5.3.2: Level 1 DFD

## Sequence Diagram

Figure 5.4 Sequence Diagram

# WORK COMPLETED

* **Data acquisition and preprocessing:**  We’ve collected the necessary facial expression data (using FER2013) and prepared it for training CNN model.
* **Model training:** We've successfully trained a CNN model to recognize the seven basic emotions with an training accuracy of 72% .
* **Music classification:** We've categorized the available music based on the identified emotions, ensuring appropriate matches for each mood

# WORK REMAINING

* **Model fine-tuning**: Further optimize the CNN model to improve its accuracy and robustness against variations in facial expressions and lighting. Consider employing techniques like data augmentation, hyperparameter tuning, and ensemble learning.
* **Music player integration**: Develop a functional music player within the system that allows users to listen to recommended songs directly. This includes features like playback controls, playlist creation, and integration with existing music libraries.
* **Document and finalization:** Prepare comprehensive documentation outlining the system design, implementation details, and user instructions. Ensure it's clear, concise, and accessible for future reference.

# REFERENCES

|  |  |
| --- | --- |
| [1] | T. Schäfer, P. Sedlmeier, C. Städtler, D. Huron, "The psychological functions of music listening," *Frontiers in psychology,* vol. 4, p. 3, 2013. |
| [2] | R. Ramanathan, R. Kumaran, R. Ram Rohan, R. Gupta and V. Prabhu, "An Intelligent Music Player Based on Emotion Recognition," in *2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, Bengaluru, 2017. |
| [3] | M. Liu, Y. Duan, R. A. A. Ince, C. Chen, O. G. B. Garrod, P. G. Schyns, and R. E. Jack, "Facial expressions elicit multiplexed perceptions of emotion categories and dimensions," *Current Biology,* vol. 32, no. 1, p. 203, 2022. |
| [4] | Preema J. S, Rajashree, Sahana M, Savitri H, and Shruthi S. J, "Review on Facial Expression Based Music Player," *International Journal of Engineering Re-search & Technology (IJERT),* vol. 6, no. 15, 2018. |