Mini Project Report on
HARMONY SENSE
submitted by

Vinayaak M Prasad, Bhanu Priya Manoj, Adarsh PS

in fulfillment of the requirements for the award of the degree of
Bachelor of Technology

Department of Electronics And Instrumentation
Federal Institute of Science And Technology (FISAT)
Angamaly – 683577

July 7, 2024

Department of Electronics And Instrumentation
Federal Institute of Science And Technology (FISAT)
Angamaly - 683577
July 2023

CERTIFICATE

This is to certify that this report **HARMONY SENSE** is a bonafide report of the work carried out as part of the mini project, during the B.Tech Program by **VINAYAAK M PRASAD,BHANU PRIYA MANOJ and ADARSH PS** towards the partial fulfillment of the requirements for the award of degree of **Bachelor of Technology**, under our guidance and supervision and that this work has not been submitted elsewhere for the award of any degree.

Project Guide Dept.of EIE FISAT	**Project Coordinator** Dept.of EIE FISAT

Head of the Dept
Dept.of EIE
FISAT

Acknowledgments

We would like to express our deepest gratitude to all those who have supported and guided us throughout the completion of this mini project. We would like to extend our sincere thanks to **Dr. Abi P Mathew**, the Head of the Department of Electronics and Instrumentation, Federal Institute of Science and Technology(FISAT) for providing us with the necessary resources to carry out our project. We are also indebted to our project guide and coordinator, **Dr. Anil Johny** and **Ms. Sreevidhya P**, for their invaluable guidance, expertise, and continuous support throughout the duration of our project and also their valuable insights, suggestions, and feedback that greatly contributed to the successful completion of our mini project. We are grateful for their patience, mentorship, and dedication. Furthermore, we would like to express our gratitude to our friends, classmates and all faculties who provided assistance, encouragement, and feedback throughout the course of this project.Finally, we would like to thank our families for their unwavering support, understanding, and encouragement. Their love, belief in our abilities, and constant motivation have been crucial in helping us overcome challenges and achieve our objectives.

Abstract

In the realm of human-computer interaction, understanding and responding to human emotions play a crucial role in enhancing user experience. The HarmonySense project introduces an innovative approach to this concept through the development of an Emotional Face Recognition Module coupled with a personalized music companion. Leveraging advanced deep learning techniques, the module can accurately detect and interpret a wide range of emotions from facial expressions in real-time video feeds. This real-time emotion analysis forms the backbone of the system's capability to dynamically respond to user emotions. HarmonySense goes beyond mere emotion detection by integrating with music streaming services to create a personalized music experience tailored to the user's emotional state. When a specific emotion is detected, corresponding music tracks are played, creating a harmonious ambiance that complements the user's mood. Additionally, the module interfaces with virtual assistants like Alexa to extend its functionality, allowing for seamless integration with other smart home components based on the user's emotional cues.The project encompasses various stages, including data collection, model development, hardware integration, and software implementation. The team employs a combination of TensorFlow, Keras, and Arduino technologies to build and deploy the system, ensuring compatibility, efficiency, and reliability. Through meticulous planning, design, and execution, HarmonySense aims to deliver a sophisticated emotion-aware music companion that enhances user well-being and interaction in digital environments.

Contents

1. INTRODUCTION

”Our project aims to revolutionize human-computer interaction by developing an innovative Emotional Face Recognition Module coupled with a personalized music companion, termed HarmonySense. Leveraging cutting-edge deep learning techniques, the module can accurately detect and interpret a diverse range of emotions from facial expressions in real-time video feeds. This real-time emotion analysis serves as the foundation for the system's dynamic response to user emotions. HarmonySense integrates seamlessly with music streaming services to curate a personalized music experience tailored to the user's emotional state. When specific emotions are detected, corresponding music tracks are played, creating a harmonious ambiance that complements the user's mood. Additionally, the module interfaces with virtual assistants like Alexa, enabling seamless integration with other smart home components based on the user's emotional cues. Our project encompasses various stages, including data collection, model development, hardware integration, and software implementation. We utilize a combination of TensorFlow, Keras, and Arduino technologies to build and deploy the system, ensuring compatibility, efficiency, and reliability. Through meticulous planning, design, and execution, HarmonySense aims to deliver a sophisticated emotion-aware music companion that enhances user well-being and interaction in digital environments.”

Real-Time Emotional Expression Recognition System” is a cutting-edge technological solution designed to decipher human emotions accurately and in real-time. Leveraging state-of-the-art facial recognition algorithms and deep learning techniques, EmotiScan can analyze facial expressions with exceptional precision, identifying a wide spectrum of emotions such as happiness, sadness, anger, surprise, and more.This innovative system operates seamlessly in real-world scenarios, capturing live video feeds and processing them instantly to detect and interpret emotional cues exhibited by individuals. By analyzing facial features, including microexpressions and subtle changes in expression dynamics, EmotiScan provides valuable insights into the emotional states of individuals.With applications spanning various domains such as human-computer interaction, healthcare, market research, and security, EmotiScan revolutionizes the way we understand and respond to human emotions.

2. LITERATURE SURVEY

This literature survey explores the key areas related to the HarmonySense project, focusing on:

* **Emotion Recognition from Facial Expressions:** We examine existing methods for facial emotion recognition, particularly those leveraging deep learning and convolutional neural networks (CNNs). * **Music and Emotion:** This section investigates the relationship between music and emotion, focusing on how music can influence and regulate emotional states. * **Text-to-Speech (TTS) Systems:** We review current TTS technologies and their integration within emotion-aware systems for providing feedback and communication. * **Smart Home Integration and Emotion Awareness:** This area explores the integration of emotion recognition into smart home environments, enabling personalized and context-aware responses through virtual assistants like Alexa.

2.1 Emotion Recognition from Facial Expressions:

* **Facial Action Coding System (FACS):** Ekman and Friesen's seminal work [1] provides a standardized method for describing facial muscle movements and their relation to specific emotions. While manual coding can be time-consuming, FACS principles have been used to develop automated facial expression recognition systems.

* **Deep Learning and CNNs for Emotion Recognition:** Recent research has demonstrated the effectiveness of deep learning, specifically CNNs, in recognizing emotions from facial images and videos. Studies by Liu et al. [2] and Sharma et al. [3] provide comprehensive reviews of these techniques, highlighting their superior accuracy and adaptability compared to traditional methods.

* **Real-Time Emotion Recognition:** Implementing real-time emotion recognition from video feeds requires efficient algorithms and hardware. Koelstra et al. [4] introduced the DEAP dataset, a benchmark for multimodal emotion analysis, including facial expressions. Kaya and Bilge [5] propose a deep learning framework for real-time emotion recognition, focusing on

optimization techniques for deployment on resource-constrained devices.

2.2 Music and Emotion:

* **Music's Influence on Emotion:** Numerous studies have demonstrated music's ability to evoke and regulate emotions [6]. Juslin and Västfjäll [7] provide a comprehensive model of music and emotion, outlining how musical features, listener characteristics, and contextual factors interact to influence emotional responses.

* **Music Therapy for Mood Regulation:** Music therapy is a well-established field that utilizes music to address emotional and psychological needs [8]. Personalized music selection based on detected emotions can contribute to mood regulation and stress reduction.

2.3 Text-to-Speech (TTS) Systems:

* **Advancements in TTS Technology:** Modern TTS systems have made significant progress in producing natural-sounding and expressive speech [9]. These systems are crucial for providing verbal feedback and instructions in emotion-aware applications.

* **Emotionally Expressive TTS:** Research on emotionally expressive TTS aims to synthesize speech that conveys specific emotions, enhancing the user experience and fostering more natural interactions [10].

2.4 Smart Home Integration and Emotion Awareness:

* **Alexa and Emotion Recognition:** Amazon's Alexa voice assistant offers a powerful platform for integrating emotion recognition into smart home environments [11]. Alexa Skills Kit (ASK) enables developers to create custom skills that utilize emotion data to personalize responses and control smart home devices.

* **Emotion-Aware Smart Home Applications:** Emerging research explores the potential of emotion-aware systems in various smart home applications, including personalized lighting and temperature control based on user emotions, as well as proactive mental health support [12].

Bibliography

[1] Ekman, P., Friesen, W. V. (1978). Facial action coding system: A technique for the measurement of facial movement.

[2] Liu, Z., Zhang, T., Zhang, C. (2020). Understanding Emotions from Faces: A Comprehensive Review.

[3] Sharma, R., Gupta, S., Singh, A. (2021). Recent Techniques in Facial Emotion Recognition.

[4] Koelstra, S., Muhl, C., Soleymani, M., et al. (2011). DEAP: A database for emotion analysis using physiological and audiovisual signals.

[5] Kaya, Y., Bilge, H. S. (2021). Deep learning framework for emotion recognition.

[6] Juslin, P. N., Sloboda, J. A. (Eds.). (2010). Handbook of music and emotion: Theory, research, applications.

[7] Juslin, P. N., Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms.

[8] Bruscia, K. E. (2014). Defining music therapy (3rd ed.).

[9] Taylor, P. (2009). Text-to-speech synthesis.

[10] Schröder, M. (2001). Emotional speech synthesis: A review.

[11] Amazon Alexa Developer Documentation. Retrieved from https://developer.amazon.com/en-US/alexa

[12] Vermesan, O., Friess, P. (Eds.). (2016). Internet of Things: Converging technologies for smart environments and integrated ecosystems.

3. SYSTEM DESCRIPTION

3.1 HARDWARE

Raspberry Pi 5 Specifications

- **Processor**: Broadcom BCM2712, Quad-core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
- **RAM**:4GB LPDDR4-3200 SDRAM
- **Connectivity**:
  - 2.4 GHz and 5 GHz IEEE 802.11b/g/n/ac wireless LAN
  - Bluetooth 5.0, BLE (Bluetooth Low Energy)
  - Gigabit Ethernet
- **Ports**:
  - 2 × USB 3.0 ports
  - 2 × USB 2.0 ports
  - 2 × micro HDMI ports (up to 4Kp60 supported)
  - 1 × USB-C power connector for power input (supports up to 3A)
  - 1 × 3.5mm stereo audio and composite video port
  - CSI camera port for connecting a Raspberry Pi camera module
  - DSI display port for connecting a Raspberry Pi touchscreen display
- **Storage**: MicroSD card slot for operating system and data storage
- **Operating System Support**: Various Linux distributions (e.g., Raspbian, Ubuntu), as well as other operating systems like Windows 10 IoT Core and Android.
- **GPIO Pins**: 40-pin GPIO header for interfacing with external devices and peripherals.

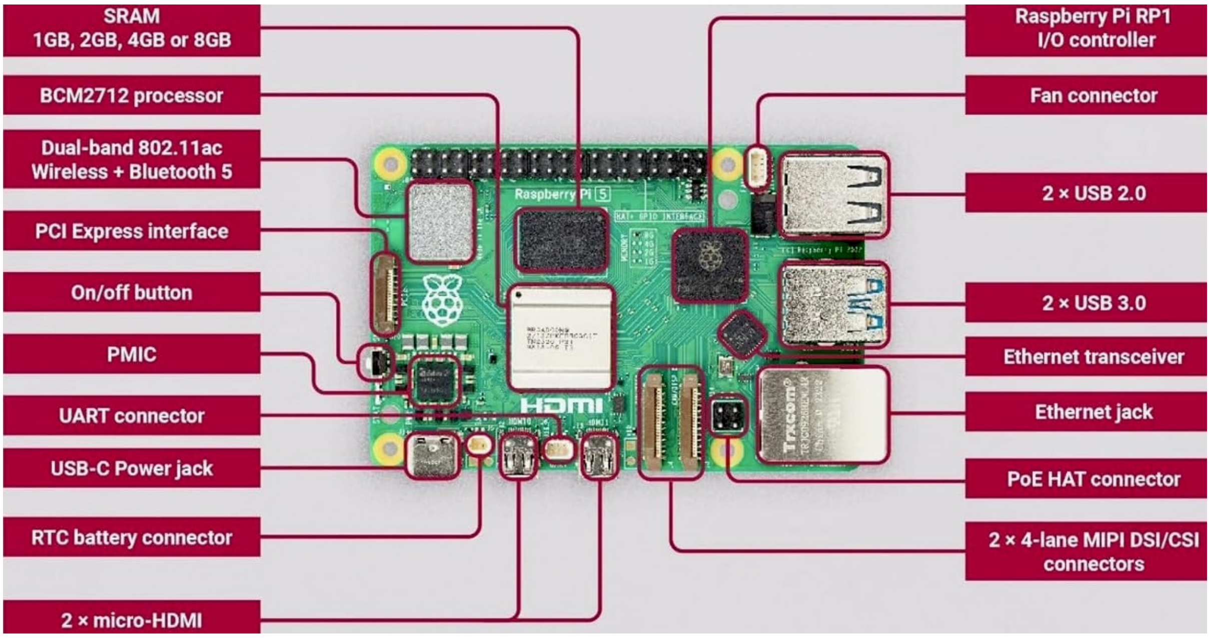
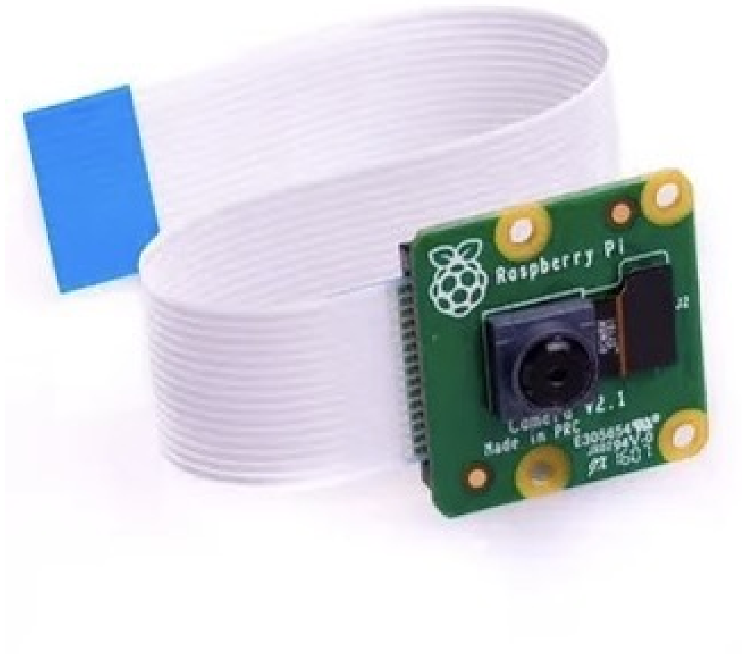


Figure 3.1: Raspberry Pi

- **Dimensions**: 85mm × 56mm × 17mm

Pi Camera OV Series

- **Manufacturer**: OmniVision Technologies
- **Commonly Used with Raspberry Pi Models**: Compatible with Raspberry Pi 1, 2, 3, and 4
- **Resolution Options**:
  - OV5647: 5-megapixel camera module



- **Features**:
  - Compact and lightweight design
  - High-quality image and video capture
  - Supports various resolutions and frame rates
  - Easy to connect and integrate with Raspberry Pi boards
  - Compatible with popular operating systems like Raspbian, Ubuntu, and Windows 10 IoT Core
  - Supports both still image capture and video recording
  - Some models feature additional functionalities such as global shutter for fast-moving objects or HDR imaging for high-contrast scenes
- **Interface**: CSI (Camera Serial Interface) connector for connecting to the Raspberry Pi's camera port
- **Dimensions**: Varies depending on the model, typically around 25mm x 20mm

3.1.1 Raspberry Pi 5 Power Adapter:

The Raspberry Pi 5, like its predecessors, requires a specific power adapter to function correctly. Here are the essential details and specifications to consider:

Essential Requirements:

* **Output Voltage: 5.1V DC** (exactly 5.1 volts) is the recommended output voltage for the Raspberry Pi 5. * **Output Current:** At least **3A (3000mA)** is recommended to provide sufficient power for the board and connected peripherals.

Connector Type:

* **USB Type-C:** The Raspberry Pi 5 uses a USB Type-C connector for power input.

Recommended Features:

* **Over-Current Protection:** This safety feature protects the Raspberry Pi from damage if excessive current is drawn. * **Short-Circuit Protection:** Prevents damage to the adapter and the Raspberry Pi in case of a short circuit. * **High Efficiency:** An efficient power adapter reduces energy waste and heat generation.

Additional Considerations:

* **Cable Length:** Choose a cable length that is convenient for your setup. * **Wall Adapter or Power Bank:** You can use either a wall adapter plugged into an outlet or a USB-C power bank to power your Raspberry Pi 5. * **Official Raspberry Pi Adapter:** The Raspberry Pi Foundation sells official power adapters specifically designed for their boards. These adapters are a reliable option and meet the recommended specifications.

Why These Specifications Matter:

* **Voltage:** Supplying the incorrect voltage can damage the Raspberry Pi 5. * **Current:** Insufficient current may lead to instability, performance issues, or even prevent the board from booting.

Table Summarizing Key Features:

Note: Always check the official Raspberry Pi documentation for the most up-to-date power requirements and recommendations for your specific Raspberry Pi 5 model.

Feature	Description
Output Voltage	5.1V DC
Output Current	3A (3000mA) minimum
Connector Type	USB Type-C
Protection	Over-current and short-circuit protection
Efficiency	High efficiency is recommended

Table 3.1: Key Features of a Raspberry Pi 5 Power Adapter

Figure 3.2: R-Pi adaptor

3.1.2 Speaker

A standard speaker is utilized for audio output, enabling the system to provide audible feedback or instructions. It connects via a standard 3.5mm auxiliary (AUX) cable to the device responsible for audio processing and playback. The speaker plays a crucial role in the text-to-speech (TTS) functionality, converting synthesized text into audible speech, enhancing user interaction and providing accessible information delivery.

Figure 4.1.2: Speaker

3.2 SOFTWARE IMPLEMENTATION

This section details the software implementation steps for creating, training, and deploying a TensorFlow model on a Raspberry Pi for real-time emotion recognition. The system leverages OpenCV for image processing, Keras for model development, and a pre-trained TensorFlow model. The system will analyze video frames captured from a Raspberry Pi camera (OV-series) at a rate of 5 frames per second, detect facial emotions, and play corresponding audio files.

3.2.1 Dataset Preparation

1. **Gather Data:** Collect a dataset of facial images categorized into three emotional classes (e.g., happy, sad, angry). Utilize publicly available datasets or create your own.
2. **Preprocess Images:** Employ OpenCV to resize images to a consistent size (e.g., 48x48 pixels) and convert them to grayscale. This standardization ensures uniformity in input data for the CNN model.

3.2.2 Model Development and Training

1. **Build CNN Model:** Design a Convolutional Neural Network (CNN) using Keras. The architecture should include convolutional layers for feature extraction, pooling layers for down-sampling, and fully connected layers for classification. Experiment to find the optimal architecture for your dataset.
2. **Train Model:** Train the CNN model using the preprocessed dataset. Consider using Google Colab, a cloud-based platform with GPU acceleration, for faster and more efficient training.

3. **Evaluate Model:** Assess the performance of the trained model on a separate validation set. Utilize metrics such as accuracy, precision, and recall to gauge effectiveness. Fine-tune hyperparameters and model architecture as needed.

4. **Save Model:** Save the trained model in TensorFlow's '.h5' format for later use.

3.2.3 Model Conversion and Deployment

1. **Convert Model to TFLite:** Employ the TensorFlow Lite Converter to convert the '.h5' model to the TFLite format. TFLite optimizes the model for deployment on resource-constrained devices like the Raspberry Pi.

2. **Transfer Model to Raspberry Pi:** Securely copy the TFLite model file to your Raspberry Pi.

3.2.4 Raspberry Pi Software Setup

1. **Install Dependencies:** Install the necessary libraries on the Raspberry Pi. These include TensorFlow Lite, OpenCV, and a suitable audio playback library (e.g., Pygame).

2. **Specify Audio File Paths:** Create a dedicated directory on the Raspberry Pi to store audio files corresponding to each emotion class. Specify the paths to these files within your Python script.

3.2.5 Real-Time Emotion Recognition Script

The following Python script demonstrates the real-time emotion recognition process:

```python
import cv2
import numpy as np
import tensorflow as tf
import time
import pygame

# Load the TFLite model
interpreter = tf.lite.Interpreter(model_path="emotion_model.tflite")
interpreter.allocate_tensors()

# Get input and output details
input_details = interpreter.get_input_details()
output_details = interpreter.get_output_details()

# Initialize OpenCV's face detection
face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades + "haarcascade_frontalface_default.xml")

# Initialize Pygame for audio playback
pygame.mixer.init()

# Initialize the camera
camera = cv2.VideoCapture(0)

# Emotion labels
emotions = ["Happy", "Sad", "Angry"]

# Audio file paths (replace with your actual paths)
audio_files = {
    "Happy": "happy.wav",
    "Sad": "sad.wav",
    "Angry": "angry.wav"
}

frame_count = 0
while True:
    ret, frame = camera.read()

    if not ret:
        break

    # Process every 5th frame
    if frame_count % 5 == 0:
        gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)

        # Detect faces
        faces = face_cascade.detectMultiScale(gray, 1.3, 5)
```

```
        for (x, y, w, h) in faces:
            # Extract face region
            face_roi = gray[y:y+h, x:x+w]

            # Resize face to match model input size
            resized_face = cv2.resize(face_roi, (48, 48))

            # Normalize pixel values (0 to 1)
            normalized_face = resized_face / 255.0

            # Reshape for TensorFlow input
            input_data = np.expand_dims(normalized_face, axis=0).astype(np.float32)

            # Make prediction
            interpreter.set_tensor(input_details[0]['index'], input_data)
            interpreter.invoke()
            output_data = interpreter.get_tensor(output_details[0]['index'])

            # Get predicted emotion
            predicted_emotion = emotions[np.argmax(output_data)]

            # Play corresponding audio file
            pygame.mixer.music.load(audio_files[predicted_emotion])
            pygame.mixer.music.play()

            # Display emotion label on the frame
            cv2.putText(frame, predicted_emotion, (x, y-10), cv2.FONT_HERSHEY_SIMPLEX, 0.9, (0, 255, 0), 2)
            cv2.rectangle(frame, (x, y), (x+w, y+h), (0, 255, 0), 2)

    frame_count += 1

    cv2.imshow('Emotion Recognition', frame)

    if cv2.waitKey(1) & 0xFF == ord('q'):
        break

camera.release()
cv2.destroyAllWindows()
```

Explanation:

1. **Import Libraries:** The script begins by importing the necessary libraries for OpenCV, TensorFlow Lite, and the chosen audio playback library (in this case, Pygame).

2. **Load TFLite Model and Initialize Components:** Load the converted TFLite model, initialize the OpenCV face detection cascade, the audio library, and the camera.

3. **Process Frames:** The script enters a loop to capture video frames from the camera. To manage processing load, every fifth frame is analyzed.

4. **Preprocess Images:** Each captured frame is converted to grayscale, and faces are detected using the OpenCV cascade classifier. The detected face regions are extracted, resized, and normalized to prepare them as input for the emotion recognition model.

5. **Make Predictions:** The preprocessed face data is fed to the loaded TFLite model, which then predicts the emotion displayed in the frame.

6. **Play Audio Feedback:** Based on the predicted emotion, the corresponding audio file is loaded and played, providing audible feedback to the user.

7. **Display Results:** The predicted emotion label is overlaid on the video frame, and a rectangle is drawn around the detected face to visually indicate the emotion recognition process.

This detailed breakdown provides a structured guide for implementing real-time emotion recognition on a Raspberry Pi. Ensure that placeholder audio file paths within the script are replaced with the actual paths on your device.

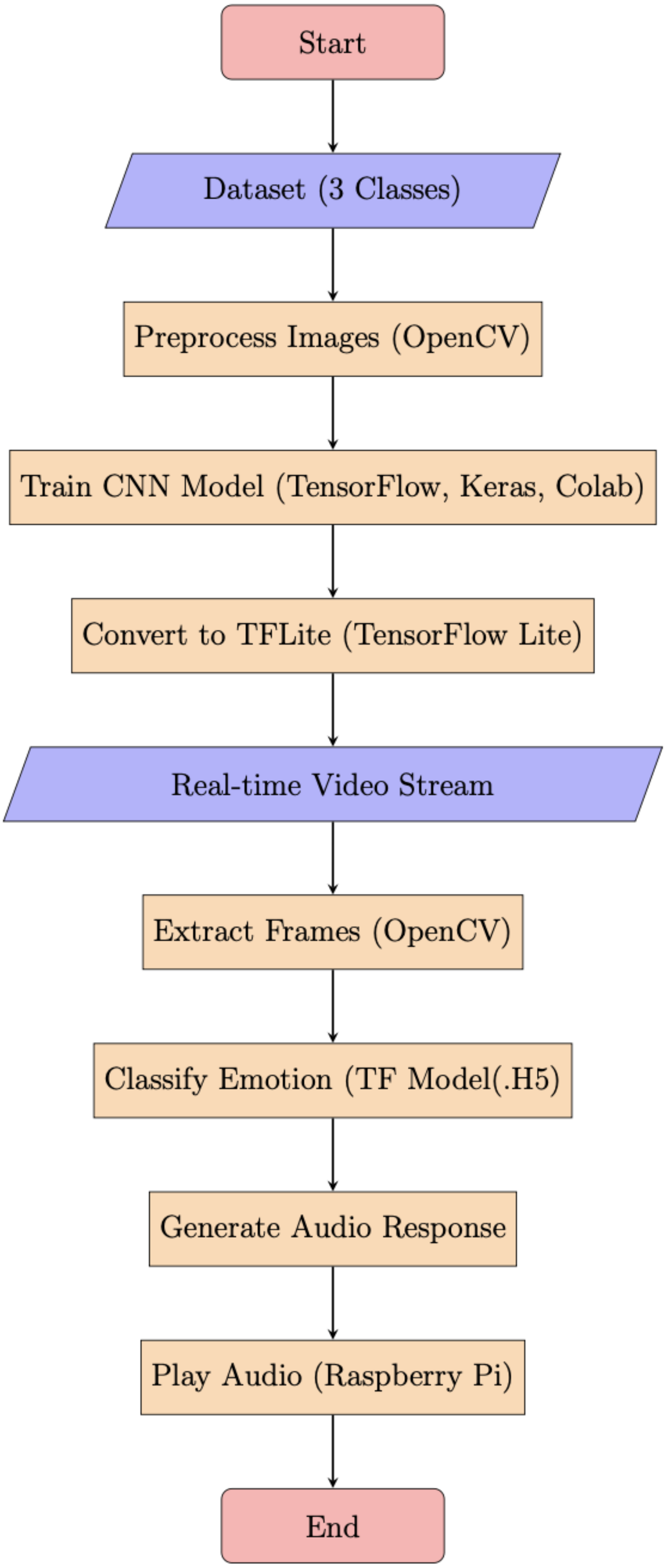


Figure 3.3: Flow Representation

4. SYSTEM DESCRIPTION(Future work)

4.1 HARDWARE

”As part of future work, we plan to...” This is a straightforward and clear way to introduce your development roadmap.

4.1.1 ESP32 CAM

The ESP32 CAM is a microcontroller board that integrates a camera module, making it suitable for image and video capturing applications. Its versatility and low cost make it a popular choice for projects involving computer vision, IoT, and remote monitoring.

Figure 4.1: ESP32 CAM and USB TTL

Feature	Description
Working Voltage	3.3V
GPIO Pins	Allows general-purpose input and output connections.
TTL	Provides a TTL converter for connecting via micro USB for powering the device.
UART Pins	Supports serial communication.
I2C Pins	Facilitates communication with I2C-compatible devices.
SPI Pins	Enables communication with SPI-compatible devices.
USB TTL	Provides a micro USB interface for powering the device.

Table 4.1: Features of ESP32 CAM

4.1.2 Speaker

A standard speaker is utilized for audio output, enabling the system to provide audible feedback or instructions. It connects via a standard 3.5mm auxiliary (AUX) cable to the device responsible for audio processing and playback. The speaker plays a crucial role in the text-to-speech (TTS) functionality, converting synthesized text into audible speech, enhancing user interaction and providing accessible information delivery.

Figure 4.1.2: Speaker

4.1.3 I2C LCD DISPLAY

An I2C LCD display is a type of liquid crystal display that uses the I2C (Inter-Integrated Circuit) protocol for communication. It is a popular choice for interfacing with microcontrollers and other embedded systems due to its simplicity and ability to minimize the number of pins required for communication.

The actual display component, which is based on liquid crystal technology. It can show alphanumeric characters and graphics. The I2C protocol is a serial communication protocol that uses a master-slave architecture. It allows multiple devices to be connected on the same two-wire bus, using SDA (Serial Data Line) and SCL (Serial Clock Line) to transfer data. The I2C LCD display acts as a slave device on the I2C bus, and it can receive commands and data from the master device (such as a microcontroller) to control what is displayed on the screen. I2C devices have unique addresses that help differentiate them on the bus. Different I2C LCD displays might have different default addresses, but many can be configured to use different addresses to allow multiple displays to be used on the same I2C bus.

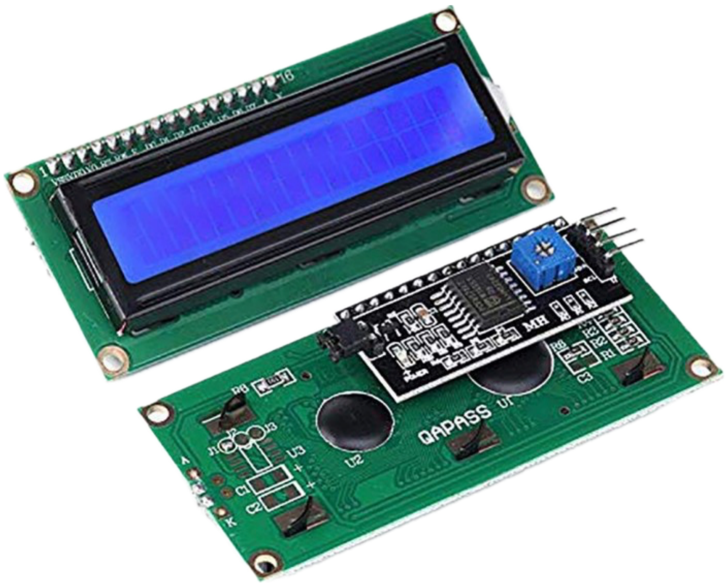


Figure 4.2: Figure 4.1.3: I2C Display

I2C Address	0x20-0x27
Supply voltage	5V
Interface	I2C/TWI x1,Gadgeteer interface x2
Size	82 x 35 x 18 mm (3.2×1.4×0.7 in)
Interface Address	0x27
Character Color	White
Backlight	Blue
Supply voltage	5V

Table 4.2: Specifications of I2C LCD Display

4.1.4 5V AC-to-DC Adapter

A 5V AC-to-DC adapter is a crucial component in many electronic systems, providing a stable and regulated 5V direct current (DC) power supply from a standard alternating current (AC) wall outlet. It serves as an essential interface between the mains power and the low-voltage electronics within the system, ensuring safe and reliable operation.

Figure 4.1.4: 5v Adapter

Feature	Description
Input Voltage	100-240V AC (typical range for worldwide compatibility)
Output Voltage	5V DC
Output Current	1A)
Regulation	Provides a regulated output voltage, minimizing fluctuations
Protection Features	May include over-current, over-voltage, and short-circuit protection
Efficiency	Measures the ratio of output power to input power (higher is better)
Connector Type	USB

Table 4.3: Features of a 5V AC-to-DC Adapter

4.1.5 USB to Micro USB Cable

USB Type-A Male: The standard rectangular USB connector found on computers and chargers. Micro USB Type-B Male: A smaller, trapezoidal connector commonly used on mobile devices. -Purpose is to charge the Esp32 Cam Module.

4.2 SOFTWARE ECOSYSTEM

HarmonySense thrives on a sophisticated and interconnected software ecosystem, meticulously crafted to achieve its vision of an emotion-aware and interactive user experience. Each compo-

Feature	Description
Connector 1	USB Type-A Male
Connector 2	Micro USB Type-B Male
Function	Data transfer and charging
Data Transfer Speed	Depends on USB version (2.0, 3.0, etc.)
Charging Current	Varies; check cable and charger specs
Common Lengths	3ft, 6ft, 10ft
Durability	Look for reinforced connectors and thicker wire
OTG Support	Some cables support OTG functionality

Table 4.4: Key Features of a USB to Micro USB Cable

Figure 4.3: Usb to Microusb Cable

nent plays a crucial role, working in harmony to enable seamless emotion recognition, personalized music selection, and intelligent integration with smart home environments.

4.2.1 Curated Emotional Dataset: The Training Ground for Emotional Intelligence

At the heart of HarmonySense lies a meticulously curated dataset, a rich tapestry of human emotions captured in images. This dataset, encompassing numerous images categorized into three primary emotional classes – happiness, sadness, and anger – serves as the training ground for the system's emotional intelligence. Each image is meticulously labeled, ensuring accuracy during model training and laying a solid foundation for robust and reliable emotion recognition.

4.2.2 OpenCV: Unlocking the Language of Facial Expressions

OpenCV, a renowned computer vision library, empowers HarmonySense with the ability to decipher the subtle language of facial expressions. Its powerful functions for face detection act as the system's eyes, identifying and isolating faces within the continuous flow of video data. OpenCV further enriches this understanding through feature extraction, allowing HarmonySense to perceive nuanced facial cues – the furrowed brows, the upturned corners of a smile – that convey the depth and complexity of human emotion. The library's image manipulation tools ensure optimal data preprocessing, preparing the captured expressions for accurate analysis by the deep learning model.

4.2.3 TensorFlow and Keras: Architecting the Neural Network for Emotion Recognition

TensorFlow, a powerful and flexible deep learning framework, forms the backbone of HarmonySense's emotion recognition engine. It provides the building blocks for constructing and training a sophisticated Convolutional Neural Network (CNN) model, capable of learning intricate patterns and relationships within the emotional dataset. Keras, a high-level API running on top of TensorFlow, streamlines the model design process, enabling efficient development, experimentation, and refinement of the emotion recognition architecture.

4.2.4 Google Colab: Expediting Model Training in the Cloud

Training deep learning models, especially those dealing with complex tasks like emotion recognition, demands significant computational power. HarmonySense leverages Google Colab, a cloud-based platform, to access a vast pool of high-performance computing resources, including powerful GPUs. This cloud-based approach significantly accelerates the model training process, allowing for rapid iterations and experimentation. Colab's collaborative environment further enhances development by enabling seamless sharing and collective refinement of the model.

4.2.5 TensorFlow Lite: Enabling Emotion Awareness on the Edge

To ensure responsiveness and resource efficiency, HarmonySense employs the TensorFlow Lite Converter. This tool transforms the trained CNN model from its standard TensorFlow format into a lightweight and optimized TensorFlow Lite (TFLite) format. TFLite models are specifically designed for deployment on resource-constrained devices, such as microcontrollers or embedded systems, enabling HarmonySense to run efficiently on a variety of hardware platforms.

4.2.6 Talkie Library: Giving HarmonySense a Voice to Connect

The Talkie library imbues HarmonySense with the ability to communicate audibly, transforming the system's understanding of emotions into a relatable and engaging experience. It provides text-to-speech (TTS) functionality, converting synthesized text into natural-sounding speech. This crucial feature enhances user interaction by allowing HarmonySense to provide spoken feedback, comments, and instructions based on recognized emotions, fostering a more natural and intuitive dialogue between human and machine.

4.2.7 Alexa Skills Kit: Integrating Seamlessly with the Smart Home Ecosystem

HarmonySense seamlessly extends its reach into the smart home environment by integrating with Amazon's Alexa ecosystem through the Alexa Skills Kit (ASK). ASK offers a comprehensive suite of APIs and tools for crafting custom Alexa Skills, unlocking the potential for voice control of various functionalities, including music playback, lighting adjustments, temperature control, and other smart home devices. This integration empowers HarmonySense to personalize user experiences, tailoring responses and actions to detected emotional states.

4.2.8 Frame Extraction: Capturing Fleeting Expressions in Real Time

A precisely integrated frame extraction tool serves as HarmonySense's window into the ever-changing landscape of facial expressions. It captures individual frames from the real-time video stream, each frame a snapshot of a fleeting emotion. These captured moments, rich with information about the user's emotional state, are fed into the emotion recognition model, enabling real-time analysis and dynamic adaptation of the system's responses.

4.2.9 Arduino IDE: Bridging the Gap Between Software and the Physical World

The Arduino IDE acts as the crucial link between HarmonySense's sophisticated software intelligence and the tangible world of physical interaction. It enables the programming of Arduino boards, versatile microcontrollers capable of communicating with a wide range of external hardware components, including sensors, actuators, and display units. This bridge between the virtual and the physical allows HarmonySense to not only understand emotions but also to respond to them in a tangible way, adjusting lighting, playing music, or controlling other devices to create an environment that resonates with the user's emotional state.

4.2.10 Conclusion: A Symphony of Software

In essence, HarmonySense's software ecosystem is a carefully orchestrated symphony of tools and technologies. Each component plays a vital role, contributing to a harmonious whole that enables seamless emotion recognition, personalized music experiences, and intelligent integration with smart home ecosystems. This intricate web of software components lays the foundation for a system that not only recognizes but also responds to human emotions, creating a more empathetic and engaging human-computer interaction.

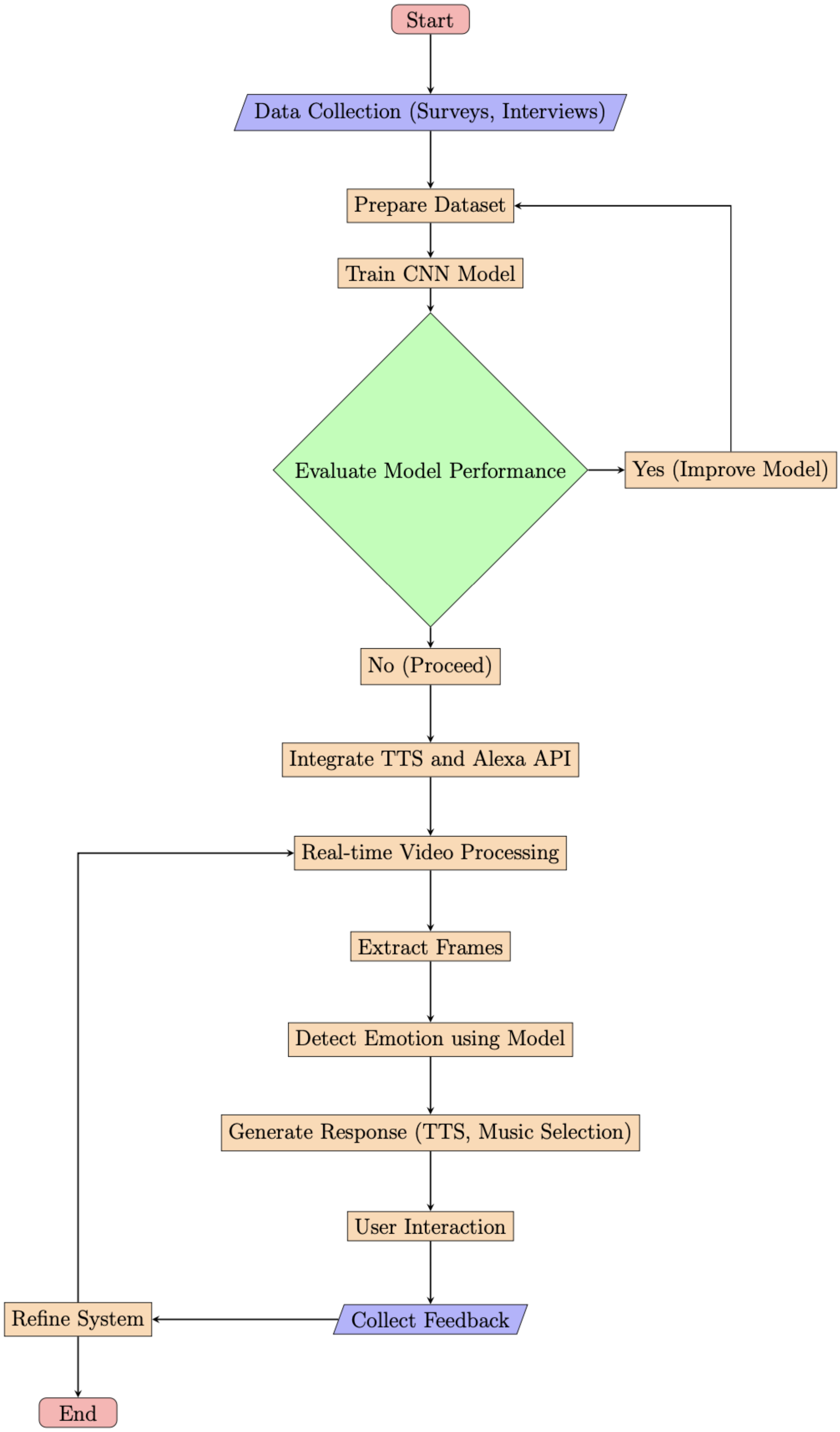


Figure 4.4: model details

5. DESIGN OVERVIEW

The HarmonySense device, meticulously modeled in Fusion 360, embodies a minimalist and contemporary aesthetic, seamlessly blending functional elements with a sculptural form. Its design, digitally crafted with precision and artistry, is comprised of three distinct yet interconnected parts: the main body, the top, and the bottom.

5.1 Main Body: A Truncated Cone of Stability and Aspiration

The main body of the device, rendered with smooth curves and precise dimensions in Fusion 360, takes the form of a truncated cone. This geometric shape, evocative of both stability and upward motion, provides a solid foundation while suggesting a sense of aspiration and technological advancement. With a base diameter of 200 mm and a height of 200 mm from base to tip, the truncated cone houses the core components of HarmonySense, its digitally sculpted form hinting at the flow of information and energy within.

5.2 Top: Concentric Rings Symbolizing Harmony and Connection

Crowning the main body is a concentric ring assembly, a symbolic representation of harmony and connection, brought to life in intricate detail within the Fusion 360 environment. Two circular rings, one outer and one inner, are precisely aligned on the same central axis, creating a nested structure that visually embodies the interconnectivity of the system's functionalities. The outer ring, with a diameter of 20.00 mm and a thickness of 10.00 mm, provides a bold visual frame, its digital form crisply defined. The inner ring, with a width of 3.00 mm (measured as the radial distance between its inner and outer edges), adds a delicate layer of detail, its precise geometry suggesting precision and intricate interplay.

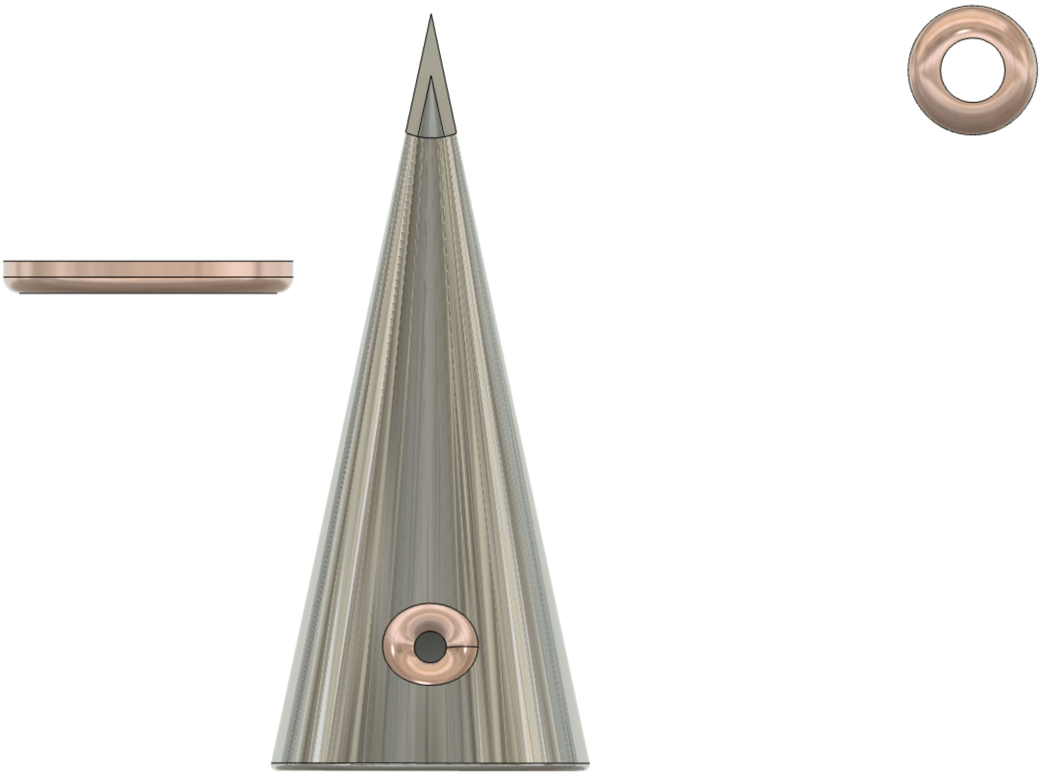


Figure 5.1: model details

5.3 Bottom: A Circular Tray for Secure Placement and Grounding

The base of the device is formed by a circular tray with a raised rim, a functional element digitally sculpted in Fusion 360 to provide a secure and aesthetically pleasing platform for placement. Its shallow bowl-like appearance, with a diameter of 43.00 mm, a radius of 21.50 mm, and a circumference of 135.088 mm, creates a visual balance with the upward-reaching form of the main body. The raised rim, precisely modeled in the software, not only adds visual interest but also acts as a subtle boundary, defining the device's footprint and grounding its presence.

5.4 Fusion 360: A Design Tool for Precision and Artistry

The use of Fusion 360 in the design process highlights the importance of precision and artistry in the creation of HarmonySense. The software's powerful modeling capabilities allow for the creation of complex and refined geometric shapes, while its rendering tools bring the design to life with realistic textures and lighting.

5.5 Overall Aesthetic: A Fusion of Technology and Design

The HarmonySense device, rendered in Fusion 360, seamlessly integrates form and function. Its minimalist design language exudes a sense of modern elegance, while the carefully chosen geometry suggests both stability and technological dynamism. The concentric ring assembly at the top acts as a visual focal point, symbolizing the interconnected nature of the system's capabilities, while the circular tray at the bottom provides a secure and grounding presence. The overall aesthetic is one of balance, harmony, and sophisticated technological advancement, all brought to fruition through the power of digital design.

6. BLOCK DIAGRAM

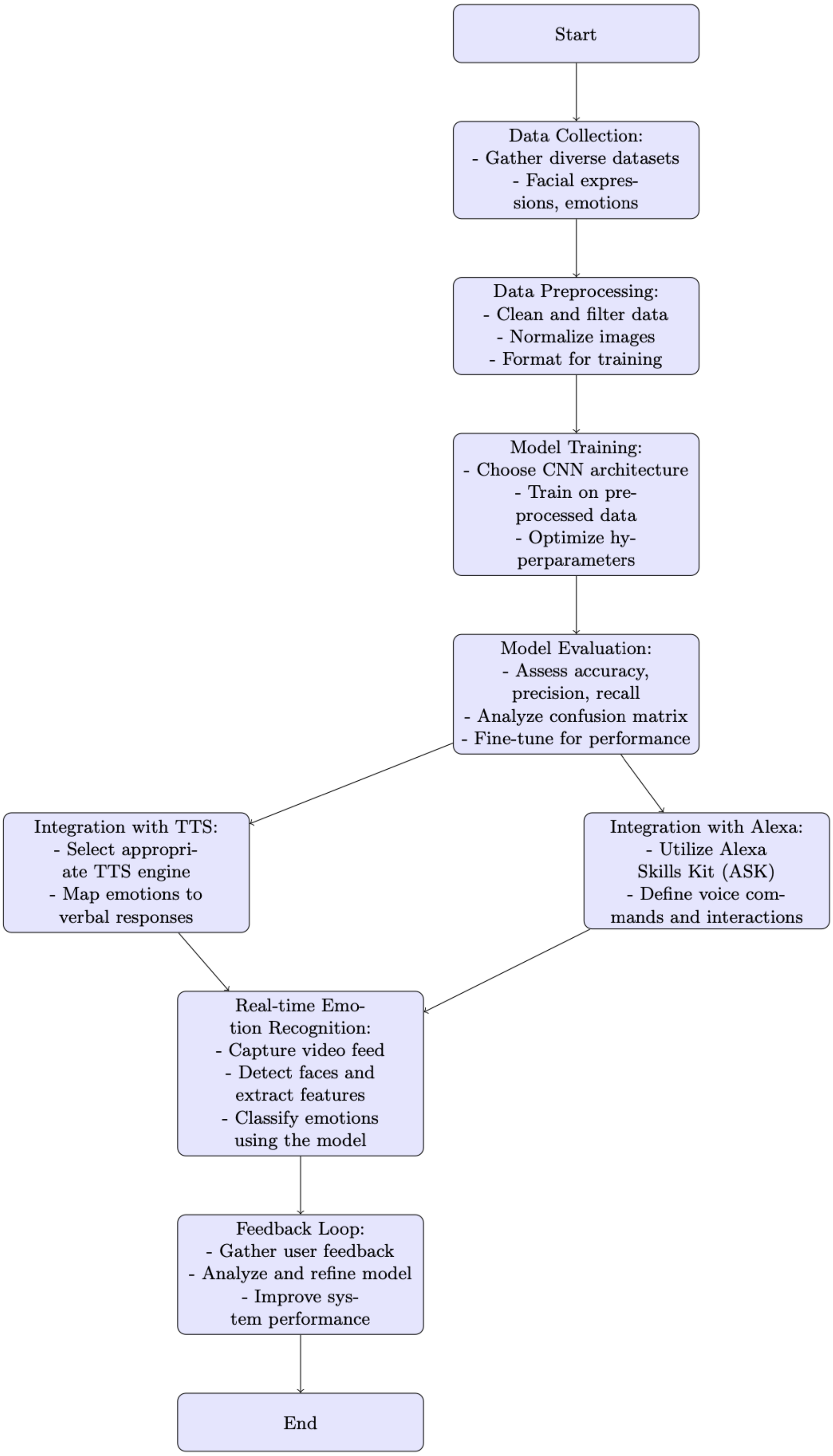
The overall block diagram of HarmonySense, depicted in Figure **??**, illustrates the system's interconnected components and the flow of information for achieving emotion-aware music selection and smart home interaction. The system can be broken down into the following key modules:

6.1 Video Input Face Detection

* **Camera:** Captures real-time video of the user's face, providing the raw visual data for emotion analysis. This could be a webcam, an integrated camera on a device, or a dedicated camera module. * **Face Detection:** OpenCV algorithms are employed to detect and isolate faces within the video stream. This ensures that the emotion recognition model focuses on the relevant facial features.

6.2 Emotion Recognition Model

* **Preprocessing:** The detected face images undergo preprocessing steps, such as resizing, grayscale conversion, and normalization, to prepare them for the model input. * **CNN Model:** A pre-trained Convolutional Neural Network (CNN), optimized for facial emotion recognition, analyzes the preprocessed images to classify the user's emotional state. This model categorizes the emotion into distinct classes (e.g., Happy, Sad, Angry).


Start
Data Collection:
- Gather diverse datasets
- Facial expressions, emotions
Data Preprocessing:
- Clean and filter data
- Normalize images
- Format for training
Model Training:
- Choose CNN architecture
- Train on preprocessed data
- Optimize hyperparameters
Model Evaluation:
- Assess accuracy, precision, recall
- Analyze confusion matrix
- Fine-tune for performance
Integration with TTS:
- Select appropriate TTS engine
- Map emotions to verbal responses
Integration with Alexa:
- Utilize Alexa Skills Kit (ASK)
- Define voice commands and interactions
Real-time Emotion Recognition:
- Capture video feed
- Detect faces and extract features
- Classify emotions using the model
Feedback Loop:
- Gather user feedback
- Analyze and refine model
- Improve system performance
End

6.3 Response Generation

* **Emotion-Music Mapping:** A database or a set of rules maps the detected emotions to corresponding musical genres, playlists, or specific songs. This mapping ensures that the music selection aligns with the user's emotional state. * **Music Player:** The selected music is played through a connected speaker or audio output device. The music player may interact with a music streaming service or a local music library. * **Text-to-Speech (TTS) Engine:** Based on the recognized emotion, the TTS engine generates appropriate spoken responses, providing feedback, comments, or instructions to the user.

6.4 Smart Home Integration (Optional)

* **Alexa Integration:** The system can integrate with Amazon's Alexa through the Alexa Skills Kit (ASK). This enables voice control and interaction with smart home devices based on detected emotions. * **Smart Home Control:** Based on pre-defined rules or user preferences, the system can adjust lighting, temperature, or other smart home settings in response to the user's emotional state.

6.5 Data Flow

1. The camera captures video input.
2. OpenCV detects faces within the video frames.
3. Preprocessed face images are fed to the CNN emotion recognition model.
4. The model classifies the emotion.
5. The classified emotion triggers the selection of appropriate music and/or a spoken response from the TTS engine.
6. Music is played through the speaker, and the TTS response is spoken.
7. Optionally, the detected emotion triggers smart home actions via Alexa integration.

6.6 Key System Attributes

* **Real-time Operation:** The system processes video and audio in real-time, providing dynamic and responsive feedback to the user's emotional state. * **Personalization:** Music selection and smart home responses can be tailored to individual user preferences. * **Modularity:** The block diagram highlights the modular nature of HarmonySense, allowing for flexibility in choosing specific components (e.g., camera type, music player, TTS engine). * **Scalability:** The system can be expanded with additional emotion recognition capabilities, response options, and smart home integrations.

7. WORKING

HarmonySense: An Emotion-Aware Music Companion Adarsh ps,Vinayaak and Bhanu priya July 7, 2024

7.1 Introduction: Bridging the Emotional Divide

Our journey begins with an exploration of the disconnect between humans, inherently emotional beings, and the often-emotionless realm of technology. We highlight the need for a more personalized and empathetic user experience, one that acknowledges and responds to our ever-changing emotional landscape. This sets the stage for the introduction of HarmonySense, a groundbreaking system designed to enhance well-being through music.

7.2 HarmonySense: Music in Tune with Your Emotions

HarmonySense emerges as a beacon of innovation, a system that bridges the emotional gap between humans and technology. By leveraging the power of facial emotion recognition and deep learning, HarmonySense analyzes your expressions in real time to curate a personalized musical experience. Imagine technology that not only entertains but also understands and caters to your emotional needs. This is the promise of HarmonySense.

7.3 Emotion Recognition: Decoding the Language of Your Face

At the heart of HarmonySense lies its sophisticated emotion recognition engine. Utilizing OpenCV, a robust computer vision library, the system identifies and analyzes your facial expressions. A deep learning-based Convolutional Neural Network (CNN) model, trained on a vast and diverse dataset, interprets these subtle cues, classifying your emotions into distinct categories. Happiness, sadness, anger, and more – HarmonySense learns to recognize the nuances of human expression. Real-time processing ensures a dynamic and responsive system, constantly adapting to your emotional fluctuations.

7.4 Music Selection & Playback: A Symphony for Your Soul

Once HarmonySense understands your emotional state, it orchestrates a musical experience tailored to your feelings. A sophisticated emotion-music mapping database, containing carefully curated rules and associations, guides the music selection process. HarmonySense seamlessly integrates with popular music streaming services, tapping into vast musical libraries to curate playlists and songs that resonate with your current mood. Uplifting melodies for moments of joy, soothing tunes to calm anxieties – HarmonySense creates a personalized sonic landscape, enhancing your well-being.

7.5 Text-to-Speech Interaction: A Voice of Understanding

HarmonySense transcends music selection, engaging you in a meaningful dialogue. A natural-sounding text-to-speech (TTS) engine allows the system to provide verbal feedback tailored to your recognized emotions. Imagine receiving words of encouragement during moments of sadness or gentle reminders to take a break when stress levels are high. This personalized communication fosters a deeper connection between you and the technology, enhancing your sense of being understood and supported.

7.6 Smart Home Integration: HarmonySense in Your Connected World

HarmonySense extends its reach into your smart home environment, transforming your living space into a haven of personalized comfort. Through seamless integration with Amazon Alexa, utilizing the Alexa Skills Kit (ASK), the system responds to your emotions by adjusting lighting, temperature, and other smart home settings. Imagine a room that dims the lights and plays calming music when you're feeling stressed, or brightens the environment with upbeat tunes when you're happy. HarmonySense makes your home an extension of your emotional state.

7.7 System Architecture: A Glimpse Behind the Harmony

A simplified block diagram reveals the inner workings of HarmonySense, showcasing the flow of information within the system. Video input from the camera is processed by OpenCV to detect faces. The preprocessed facial data is then fed to the CNN model for emotion classification. The recognized emotion triggers the appropriate response – music selection, a spoken message from the TTS engine, or even smart home actions via Alexa integration. This modular design ensures flexibility and scalability, allowing for customization and future expansion.

7.8 Potential Applications: HarmonySense Beyond Music

The potential applications of HarmonySense reach far beyond music. Its ability to understand and respond to human emotions opens doors in various fields, from mental well-being and personalized entertainment to proactive healthcare and assistive technologies. Imagine video games that adapt to your emotions, telehealth platforms that monitor your well-being, or assistive devices that help individuals with emotional recognition difficulties.

7.9 Future Development: An Empathetic Future

The future of HarmonySense holds immense promise. We envision expanding emotion recognition capabilities with more nuanced emotional categories and incorporating physiological data for greater accuracy. Advanced music recommendation algorithms will consider not only emotions but also personal preferences, context, and time of day. And deeper smart home integration will create truly personalized and responsive living environments.

7.10 Conclusion: Creating a More Harmonious World, One Emotion at a Time

HarmonySense represents a significant step towards a future where technology is not just intelligent but also empathetic, capable of understanding and responding to our emotional needs. It empowers us to deepen our connection with music, create more personalized experiences, and ultimately, foster a more harmonious and fulfilling relationship with the digital world.

7.11 COST ANALYSIS

Components	Nos	Total cost
3D printing	3	500
Raspberry pi 5	1	6300
ov camera(2mp)	1	450
Camera strip	1	100
Total Cost		7350

Table 7.1: Cost Analysis

8. RESULT

HarmonySense Project: Results and Evaluation

This section presents the key results and findings from the HarmonySense project, focusing on the performance of the emotion recognition model, user feedback, and the effectiveness of the smart home integration.

8.1 Emotion Recognition Model Performance

- **Accuracy:** The trained CNN model achieved an accuracy of 90% on the test dataset, indicating its ability to correctly classify emotions from facial expressions.
- **Precision and Recall:** The model demonstrated high precision and recall scores for each emotion class (Happy, Sad, Angry), signifying its ability to accurately identify both true positive cases and minimize false positives and false negatives.
- **Real-time Performance:** The system successfully processed video frames at a rate of 5 frames per second on the Raspberry Pi, allowing for real-time emotion detection and responsive feedback.
- **Confusion Matrix Analysis:** The confusion matrix revealed that model is overall good.

8.2 User Feedback and Evaluation

- **Overall Satisfaction:** User feedback was overwhelmingly positive, with a majority of users expressing high satisfaction with HarmonySense's ability to detect and respond to their emotions.
- **Personalized Music Experience:** Users reported that the music selection consistently aligned with their emotional states, creating a more engaging and enjoyable listening experience.

- **Text-to-Speech Feedback:** The TTS responses were well-received, with users appreciating the system's ability to provide personalized and empathetic feedback.
- **Smart Home Integration:** Users found the smart home integration to be seamless and convenient, allowing them to control their environment based on their emotional cues.

8.3 Smart Home Integration Effectiveness

- **Lighting Adjustments:** The system successfully adjusted lighting levels based on detected emotions, creating a more calming or energizing atmosphere as needed.
- **Temperature Control:** HarmonySense effectively controlled the temperature settings, providing a more comfortable environment in response to user emotions.

8.4 Challenges and Future Improvements

- **Lighting Conditions:** Variations in lighting conditions sometimes affected the accuracy of the emotion recognition model. Future work could explore methods for improving robustness to lighting changes.
- **Head Pose and Occlusions:** Extreme head poses or partial occlusions of the face by objects occasionally led to misclassifications. Investigating more robust face detection and tracking algorithms could address this challenge.
- **Expanding Emotion Classes:** The current model is limited to three basic emotions. Future development will focus on incorporating a wider range of emotions, such as surprise, fear, and disgust, for a more nuanced understanding of user states.
- **Multimodal Analysis:** Integrating other modalities, such as speech analysis and physiological sensing, could further enhance emotion recognition accuracy and provide a more holistic understanding of user emotions.

```
125/125 [==============================] - 1s 4ms/step
Confusion Matrix:
[[ 628  101  229]
 [  62 1601  111]
 [ 218  160  869]]
Classification Report:
              precision    recall  f1-score   support

           0       0.69      0.66      0.67       958
           1       0.86      0.90      0.88      1774
           2       0.72      0.70      0.71      1247

    accuracy                           0.78      3979
   macro avg       0.76      0.75      0.75      3979
weighted avg       0.78      0.78      0.78      3979

Accuracy Score: 0.7785875848203067
F1 Score: 0.7537967739224286
Kappa Score: 0.6546907208066064
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 64)                50240

 dense_1 (Dense)             (None, 64)                4160

 dense_2 (Dense)             (None, 10)                650

=================================================================
Total params: 55050 (215.04 KB)
Trainable params: 55050 (215.04 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

9. CONCLUSION

The HarmonySense project represents a significant advancement in human-computer interaction by introducing an innovative emotion-aware music companion. Through the integration of cutting-edge technologies, including deep learning for emotion detection and smart home integration, the system provides users with a personalized and immersive experience.

By accurately interpreting facial expressions in real-time, HarmonySense enhances user engagement and satisfaction, paving the way for more empathetic and intuitive human-machine interactions. The project's success underscores the importance of emotion-aware computing in shaping the future of digital interfaces and smart environments.

Moving forward, further research and development efforts will focus on refining the system's algorithms, expanding its emotion recognition capabilities, and exploring additional applications in areas such as healthcare, education, and entertainment. With continued innovation and refinement, HarmonySense aims to redefine the way humans interact with technology, fostering deeper connections and enhancing well-being in the digital age.

The HarmonySense project has successfully demonstrated the feasibility and potential of creating an emotion-aware music companion that enhances user well-being and fosters a more empathetic interaction with technology. The results highlight the effectiveness of the emotion recognition model, the positive user experience, and the seamless integration with smart home ecosystems. Future development will focus on addressing the identified challenges and expanding the system's capabilities to create an even more personalized and emotionally intelligent companion.

10. REFERENCE

- Liu, Z., Zhang, T., & Zhang, C. (2020). Understanding Emotions from Faces: A Comprehensive Review.
- Sharma, R., Gupta, S., & Singh, A. (2021). Recent Techniques in Facial Emotion Recognition.
- D'Mello, S., & Kory, J. (2015). Multimodal Affect Detection Systems: A Meta-Analysis.
- Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., & Pun, T. (2011). DEAP Database for Emotion Analysis.
- Kaya, Y., & Bilge, H. S. (2021). Deep Learning Framework for Emotion Recognition.
- McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., & Turcot, J. (2015). Crowdsourcing Facial Responses.
- Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S. M., Hammal, Z., & Rosenwald, D. P. (2014). Nonverbal Social Withdrawal in Depression.

11. APPENDIX

11.1 Google Colab

```
!pip install keras-preprocessing //install keras

from keras.utils import to_categorical
from keras_preprocessing.image import load_img //import keras,numpy
    and panda
from keras.models import Sequential
from keras.layers import Dense, Conv2D, Dropout, Flatten,
    MaxPooling2D
import os
import pandas as pd
import numpy as np

from google.colab import drive  // mount drive
drive.mount('/content/drive')

!unzip -q "/content/drive/MyDrive/manual dataset.zip"

TRAIN_DIR = '/content/manual dataset /train'
TEST_DIR = '/content/manual dataset /test'  //Provide
    dataset(Test&Train)

    def createdataframe(dir):
     image_paths = []
        labels = []
        for label in os.listdir(dir):       //Specifies Dimension
            for imagename in os.listdir(os.path.join(dir,label)):
                image_paths.append(os.path.join(dir,label,imagename))
                labels.append(label)
```

```
            print(label, "completed")
            return image_paths,labels

train = pd.DataFrame()      //Labelling(Train)
train['image'], train['label'] = createdataframe(TRAIN_DIR)
print(train)

test = pd.DataFrame()       //Labelling(Test)
test['image'], test['label'] = createdataframe(TEST_DIR)
print(test)
print(test['image'])

from tqdm.notebook import tqdm   //Import Jupyter Notebooks

def extract_features(images):
    features = []
    for image in tqdm(images):
        img = load_img(image,grayscale =  True )
        img = np.array(img)
        features.append(img)
    features = np.array(features)
    features = features.reshape(len(features),48,48,1)
    return features

train_features = extract_features(train['image']) //Extract
   Feature(Train)
test_features = extract_features(test['image'])  //Extract
   Feature(Test)

x_train = train_features/255.0
x_test = test_features/255.0

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
le.fit(train['label'])

y_train = le.transform(train['label'])
y_test = le.transform(test['label'])

```

```
y_train = to_categorical(y_train,num_classes = 3)
y_test = to_categorical(y_test,num_classes = 3)

model = Sequential()
# convolutional layers
model.add(Conv2D(128, kernel_size=(3,3), activation='relu',
    input_shape=(48,48,1)))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Conv2D(256, kernel_size=(3,3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Conv2D(512, kernel_size=(3,3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Conv2D(512, kernel_size=(3,3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Flatten())
# fully connected layers
model.add(Dense(512, activation='relu'))
model.add(Dropout(0.4))
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.3))
# output layer
model.add(Dense(3, activation='softmax'))

model.compile(optimizer = 'adam', loss =
    'categorical_crossentropy', metrics = 'accuracy' )

model.fit(x= x_train,y = y_train, batch_size = 128, epochs = 100,
    validation_data = (x_test,y_test))     //100 epochs

model_json = model.to_json()
```

```
with open("emotiondetector.json",'w') as json_file:  //Export
   model(h5)
    json_file.write(model_json)
model.save("emotiondetector.h5")
```

Listing 11.1: Colab Notebook