

Facial Emotion Recognition for Visually Impaired People

CEN 493 - Graduation project

Shadi Al Minawi (252001554)

Third Trimester 2025

Advisor:

Dr. Ashraf El-Sayed

Table of Contents

Abstract.....	3
Chapter 1 The Problem Statement	4
Section 1.1 Need Statement.....	4
Section 1.2 Objective Statement.....	4
Section 1.3 Literature Review	5
Chapter 2 Requirements Specification	9
Section 2.1 User Needs	9
Section 2.2 Requirement Specification	9
Chapter 3 Overview of The System.....	12
Section 3.1 Concept Design.....	12
3.1.1 Machine/Deep learning	12
3.1.2 Hardware Platform	14
3.1.3 Camera	16
Section 3.2 Preliminary Work	17
Section 3.3 System Design Overview	23
Design Level 0	23
Design Level 1	24
Chapter 4 Design Implementation.....	26
Section 4.1 Prototype Implementation.....	26
Section 4.2 Prototype Verification	29
Chapter 5 Conclusion	32
References	33
Appendix A The Project Plan	35
Section A.1 Work Breakdown Structure.....	35
Section A.2 Gantt Chart	37
Section A.3 Cost Estimation.....	37

Abstract

This project focuses on developing a deep learning-based facial emotion recognition system designed to assist visually impaired individuals in social interactions. By accurately identifying and interpreting facial expressions, the system aims to bridge the communication gap faced by the visually impaired community, often hindered by an inability to perceive emotional and social cues visually. The system will operate on an embedded platform, ensuring portability and accessibility. It will analyse facial expressions in real time, leveraging advanced deep learning algorithms tailored for high accuracy and low latency. The identified emotions will be conveyed to the user through audible reports, enabling an intuitive understanding of the social and emotional context during interactions. This innovative approach not only enhances the social engagement of visually impaired individuals but also fosters inclusivity and empathy within the broader community. By providing a technologically advanced solution, this project seeks to empower visually impaired individuals with greater independence in social settings and contribute significantly to improving their quality of life.

Chapter 1

The Problem Statement

Section 1.1 Need Statement

Visually impaired individuals encounter significant challenges in their social interactions, primarily due to their inability to perceive non-verbal cues, especially facial expressions. This limitation critically impedes their ability to fully comprehend the emotions and intentions of others, which can lead to misunderstandings and an increased risk of social isolation. The World Health Organization reports that approximately 285 million people worldwide are visually impaired, with 39 million classified as blind and 246 million having low vision. Moreover, around 2.2 billion people globally live with some form of vision impairment or blindness, underscoring the widespread prevalence and profound impact of visual impairment. This vast number of affected individuals highlights the urgency of addressing this issue [1] [2].

The difficulty in interpreting non-verbal cues has a profound effect on the social well-being and mental health of visually impaired individuals. Research indicates that this challenge contributes to increased feelings of loneliness and depression among those who are visually impaired [3]. Consequently, it is imperative to recognize the significance of this issue to improve the quality of life for those with visual impairments and promote social inclusion and mental well-being throughout society. Addressing this need is crucial to mitigate the adverse effects on the mental health and social interactions of visually impaired individuals, emphasizing the importance of societal efforts to foster a more inclusive environment for all.

Section 1.2 Objective Statement

The facial emotion recognition system will employ computer vision algorithms to analyze facial expressions. Utilizing a device's camera as input, the system will capture facial features, extract relevant emotional cues, and classify expressions into distinct emotions. The output will consist of audible feedback to convey the recognized emotions to the user, maintaining a non-intrusive and user-friendly interaction. The system's adaptability to varying environmental conditions and compatibility with standard devices aim to provide an inclusive and seamless user experience for individuals with visual impairment. Detailed design and implementation specifications will be developed in subsequent project phases.

Section 1.3 Literature Review

Facial Emotion Recognition (FER) technology analyzes facial expressions from images and videos to infer emotional states. It's pivotal in non-verbal communication, offering insights into human emotions and being crucial for social interaction and communication. However, challenges such as variability in facial expressions, contextual and cultural differences, and technological limitations like lighting and occlusions affect accuracy. FER has vast applications in healthcare, detecting disorders, education, gauging student engagement, security, monitoring public spaces, and customer service to personalize experiences. This technology, while promising, faces ethical and privacy concerns, necessitating careful consideration in deployment. Yet traditional methods exist that can be utilized in FER efforts.

The application of various techniques in FER, including neural networks and traditional machine learning methods. Specifically, it mentions the use of Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) as traditional feature descriptors combined with data classifiers like Support Vector Machine (SVM), k-nearest Neighbors (KNN), or Random Forest for emotion detection and recognition. These methods were compared to Artificial Neural Networks (ANNs), highlighting the shift towards using ANNs due to their superior performance in detecting and recognizing emotions in a subject-independent manner. The document suggests that while traditional methods have their merits, the integration of ANNs, especially deep learning techniques, has opened new possibilities for practical applications across various fields such as healthcare, security, business, education, and manufacturing, offering improved accuracy and efficiency in emotion recognition tasks. Emotion Detection using Image Processing in Python outlines a methodology for detecting human emotions from facial expressions using Python, OpenCV, and NumPy. It details creating a dataset, processing images to extract emotions, and classifying emotions using Fisher Face classifiers. The study involves converting images to grayscale, cropping, and organizing them into directories based on emotion types. It emphasizes the importance of using consistent image size and properties for accurate emotion recognition, suggesting an 80/20 split for training and classification to enhance accuracy and efficiency [4][5].

In [6], the authors outline a facial expression recognition project employing Convolutional Neural Networks (CCN) specifically to identify human emotions from facial images. It covers the essential stages of FER systems, including preprocessing, face detection, feature extraction, and expression classification. However, the document primarily focuses on using CNN to recognize seven key human emotions and does not explicitly detail the use of Local Binary Patterns, Histogram of Oriented Gradients, facial landmarks, or local feature descriptors as requested. These techniques are integral to traditional FER methods and contribute to the feature extraction phase by analyzing texture, shape, and changes in facial attributes, which are critical for accurate emotion classification.

In [7], the authors discuss an innovative Facial Expression Recognition method that enhances FER performance through facial action units-based image retrieval. This technique leverages Multi-layer Perceptrons (MLPs) to assess the reliability of FER results. Unreliable outcomes trigger a search for similar facial expressions using facial action units, aiding in the reclassification process. It integrates traditional feature extraction methods like LBP and HOG, alongside advanced neural network models, to refine emotion classification. This approach signifies a melding of classical feature analysis and modern deep learning, aiming to improve accuracy in emotion detection.

In [8], the authors present a system for automatic facial expression recognition, utilizing image processing techniques and neural networks. It focuses on extracting features from the contours of eyebrows, eyes, and mouth using a scalable rectangle, aiming to improve over manual point selection and complex face models. The approach reduces the number of features to quicken recognition time while maintaining accuracy. This method involves preprocessing to adjust lighting, detecting facial features, normalizing these features to account for scale differences, and classifying expressions using SimNet, a combination of Fuzzy Logic and ANNs.

In [9], the authors provide a comprehensive overview of recent advancements in FER using deep learning techniques. It details the challenges in the field, such as variations in facial expressions and environmental conditions, and discusses the effectiveness of various deep learning architectures like CNN and Recurrent Neural Networks (RNNs) in overcoming these challenges. The paper highlights the importance of large and diverse datasets for training models and reviews several databases used in FER research. A comparative analysis of different studies offers insights into improving FER systems' accuracy and efficiency. "FER Using Deep Learning" [10], explores FER using deep learning techniques. It includes a comprehensive examination of CNN and SVM for FER, detailing the methodology, dataset preparation (FER2013 and CK+ datasets), and implementation of these models. The thesis aims to improve the accuracy and efficiency of FER systems, addressing challenges such as occlusion, contrast variations, class imbalance, and intra-class variations.

In [11], the authors present a study focusing on improving facial expression recognition using CNN. The research utilizes the FER2013 dataset, comprising samples of seven universal facial expressions, to train the model. The results demonstrate that the CNN-based technique enhances facial expression recognition accuracy without necessitating the encoding of multiple CNN layers, which typically results in a computationally intensive model. This approach offers a solution to high computational costs associated with facial expression recognition, achieving near state-of-the-art accuracy while ensuring efficiency and prediction capabilities in the health sector.

It focuses on the critical role of facial expression recognition in conveying human emotions. It provides a comprehensive review of automated FER systems, emphasizing the challenges and progress in the field. The paper highlights the efficiency of deep learning architectures, particularly CNN, as a novel solution for FER issues. It investigates recent and highly cited works, detailing

various deep learning methods applied to facial expression recognition, the databases used, and the success of these techniques in improving recognition accuracy [12].

In [13], the authors explore the application of deep learning and transfer learning to FER, focusing on evaluating model performance and the impact of transfer learning on traditional sequential models. The thesis uses the FER-2013 database, containing 35,887 images across seven emotion categories, and employs a k-fold split method for training and testing. It compares the performance of a sequential CNN model, a pre-trained ResNet50 model, and a pre-trained VGG16 model, using metrics like F1 score and confusion matrices. The study demonstrates the effectiveness of transfer learning in reducing computational time and achieving comparable accuracy, making it a recommended approach for FER tasks.

In [14], the author delves into a sophisticated system leveraging deep learning to empower visually impaired people by enabling them to perceive facial expressions through auditory feedback. This technology harnesses the power of CNN, a class of deep neural networks highly effective in analyzing visual imagery, to identify a range of facial expressions. By training these networks with extensive datasets, the system can accurately classify different emotions conveyed through facial expressions. Once identified, these emotions are translated into audio signals, providing real-time, intuitive feedback to the user. Otherwise, each not only enhances social interactions for the visually impaired by giving them access to non-verbal cues that are otherwise inaccessible but also contributes to their emotional understanding and connectivity in interpersonal communications. Also presents a deep learning-based facial expression recognition system for assisting visually impaired persons, implemented on a wearable device. It utilizes MobileNetV2 for its efficiency and speed, which makes it suitable for real-time applications on edge devices like Raspberry Pi. The system combines datasets including CK+, ANN, and JAFFE for training, achieving notable accuracy. Experiments in indoor and outdoor settings with visually impaired participants showed success rates of 80% and 78%, respectively, demonstrating the system's effectiveness in real-world conditions for enhancing social interactions for the visually impaired through auditory feedback of facial expressions.

In [15], the authors outline a comprehensive study on developing a lightweight FER system tailored for visually impaired individuals. This system leverages a novel approach using partial transfer learning with a custom-trained Convolutional Neural Network, aiming to improve accuracy in emotion recognition while ensuring the system remains efficient for real-time applications on edge devices. The document highlights the challenges faced by visually impaired people in social interactions due to the inability to perceive visual cues and proposes this technology as a solution. With an emphasis on portability and low computational requirements, the system demonstrated promising results, achieving high recognition accuracy, particularly for emotions like happiness, sadness, and surprise, making it a significant advancement in assistive technologies for enhancing the quality of life and social integration of visually impaired individuals. The designed system showcased an impressive achievement in terms of accuracy and efficiency. The system achieved a notable accuracy rate of 82.1% on the enhanced Facial Expression Recognition 2013 dataset

such as EFER2013 and CK+, an improvement over the previous model. This high level of performance is particularly significant considering the system's design for real-time application on edge devices, which are characterized by their limited processing power. The system's architecture, optimized for efficiency, comprises only 1.49 million parameters, demonstrating its suitability for low-resource environments and its potential to significantly aid visually impaired individuals in social interactions by accurately interpreting facial emotions.

Actual Label	Angry	Disgust	Sad	Surprise	Happy
Face Image					
Predicted Label	Sad	Disgust	Sad	Sad	Happy

Figure 1.1 Samples from the EFER2013 dataset with the actual and predicted labels. [15]

Actual Label	Anger	Sad	Anger	Fear	Happy	Disgust	Surprise
Face Image							
Predicted Label	Sad	Sad	Anger	Sad	Happy	Disgust	Surprise

Figure 1.2 Samples from the CK+ dataset with the actual and predicted labels. [15]

The studies on FER highlight the pivotal role of deep learning techniques in advancing the field by addressing traditional challenges such as variability in facial expressions, contextual and cultural differences, and technological limitations. Deep learning methods, particularly CNN, have demonstrated superior performance in detecting and recognizing emotions compared to traditional methods like Histograms of Oriented Gradients and Local Binary Patterns. These studies reveal that while traditional methods have their merits, the integration of deep learning techniques has significantly improved the accuracy and efficiency of emotion recognition systems. Applications span across various fields including healthcare, security, education, and customer service, showcasing the practical benefits of deep learning. Furthermore, innovations such as using facial action units for reclassification and employing deep learning on edge devices for real-time applications, particularly for assisting visually impaired individuals, underscore the transformative potential of deep learning in FER. The comprehensive examination of deep learning architectures and their comparative analysis against traditional methods affirm the efficacy and adaptability of deep learning, making it the cornerstone technique for advancing FER technology.

Chapter 2 Requirements Specification

Section 2.1 User Needs

User needs in projects refer to the specific requirements, preferences, and expectations of the end users that must be met by the project deliverables or outcomes.

- The system should be based on a general off-the-shelf embedded platform.
- The system should be capable of recognizing and classifying facial expressions with high accuracy.
- The system should be easy to operate, where the user captures a picture of a person's face to analyze their facial expression.
- The system should return the recognized facial emotion audibly through a speaker.
- The system should be flexible enough to recognize a wide range of facial expressions accurately.
- The system should be reliable in identifying emotions across different individuals and lighting conditions.
- The system should be cost-effective, considering both hardware and software components.
- The system should have a high scanning speed to analyze facial expressions quickly.
- The system should be designed to minimize power consumption.

Section 2.2 Requirement Specification

- The system should use an off-the-shelf embedded platform, due to its superior performance with deep learning models compared to other options.
- The system should accurately recognize and classify facial expressions with at least 90% accuracy.
- The system's training data should include a diverse range of facial expressions to ensure high accuracy in recognition.
- The system should perform a quick scan of facial images within 3 seconds or less.
- The system shall classify and output recognized emotion within 3 seconds.
- The system should include (ON/OFF) to optimize energy consumption.
- The system should audibly output recognition results within 5 seconds.
- The project budget is limited to 1000 SAR and must not exceed this limit.

Table 2.1 matching Marketing requirements to Engineering requirements.

Marketing Requirements	Engineering Requirements	Justification
2,6	The system should use an off-the-shelf embedded platform, due to its superior performance with deep learning models compared to other options.	Using an off-the-shelf embedded platform simplifies the system's development and integration process while delivering the performance needed for deep learning models.
1	The system's training data should include a diverse range of facial emotion datasets (CK+, FER2013) to ensure high accuracy in recognition.	Diverse facial expression training data reduces bias and ensures high recognition accuracy across different demographics.
8	The system should accurately recognize and classify facial expressions with at least 90% accuracy.	High accuracy ensures reliable recognition, providing visually impaired users with dependable emotional cues, which improves their social interactions.
2,5	The system should perform a quick scan of facial images within 3 seconds or less.	A quick scan minimizes latency, allowing users to receive results promptly and interact with others without delays.
2,3,8	The system shall classify and give an output of recognized emotion within 3 seconds	Fast classification and output ensure that users receive timely results, enabling them to understand emotions in real time.
4	The system should audibly output recognition results within 5 seconds.	Delivering the recognized emotion audibly within seconds ensures users can respond promptly to the emotional cues.
7,9	The system should include (ON/OFF) to optimize energy consumption.	Energy-saving modes optimize power consumption and extend battery life, ensuring reliable performance for mobile or wearable systems.
7	The project budget is limited to 1000 SAR and must not exceed this limit.	Limiting the budget ensures cost-effectiveness and accessibility, while still providing high-quality performance.

Marketing Requirements:

1. The system should be based on a general off-the-shelf embedded platform.
2. The system should be capable of recognizing and classifying Facial emotions with high accuracy.
3. The system should be easy to operate where the user takes a picture of a facial emotion.
4. The system should return the recognized facial emotion audibly.
5. The system should be flexible enough to recognize a wide range of facial emotions.
6. The system should be reliable.
7. The system should be cost-effective.
8. The system should have a high scanning speed.
9. The system should not consume too much power.

Chapter 3 Overview of The System

Our FER design should be based on general off-the-shelf embedded platforms with a suitable programming language. It should be capable of recognizing facial expressions with high accuracy and return the recognized emotions audibly, ensuring that visually impaired individuals receive prompt, clear, and accessible feedback.

Section 3.1 Concept Design

Table 3.1 Concept overview.

Machine Learning	Hardware Platform	Camera	Speaker
Deep Learning	Raspberry Pi	Raspberry Pi camera V1	Speaker
Image Processing	Arduino	Raspberry Pi camera V2	

3.1.1 Machine/Deep learning:

Machine/Deep Learning is a branch of Artificial Intelligence that enables systems to learn and improve from experience automatically, without the need for explicit programming.

Table 3.2 Technique comparison.

Method	Strengths	Weaknesses
Deep Learning	- Excellent accuracy, especially with large datasets. - Can handle complex data and recognize patterns with high precision. - Capable of improving its accuracy with additional data.	- Requires significant training time. - Requires large datasets. - May require specialized hardware or software.
Image Processing	- Excellent accuracy, especially with large datasets. - Can handle complex data and recognize patterns with high precision. - Capable of improving its accuracy with additional data.	- May not be as accurate as deep learning-based models. - May not handle complex recognition tasks.

Our project on FER for visually impaired people can be approached using two methods: image processing and deep learning. Image processing techniques, such as Histograms of Oriented Gradients and Local Binary Patterns, involve extracting features manually and using classifiers like Support Vector Machines or k-nearest Neighbors for emotion detection. While these methods are computationally less intensive and easier to implement, they often struggle with variations in lighting, facial expressions, and occlusions, leading to lower accuracy. In contrast, deep learning, particularly CNN, automatically learns features from data, making it more robust to such variations and significantly improving recognition accuracy. Deep learning models can handle large, diverse datasets and capture complex patterns in facial expressions, which is crucial for accurate emotion detection. Given the need for high accuracy and real-time processing in assistive technologies for visually impaired individuals, deep learning offers superior performance and adaptability, making it the preferred choice for our project.

So, we chose: **Deep Learning.**

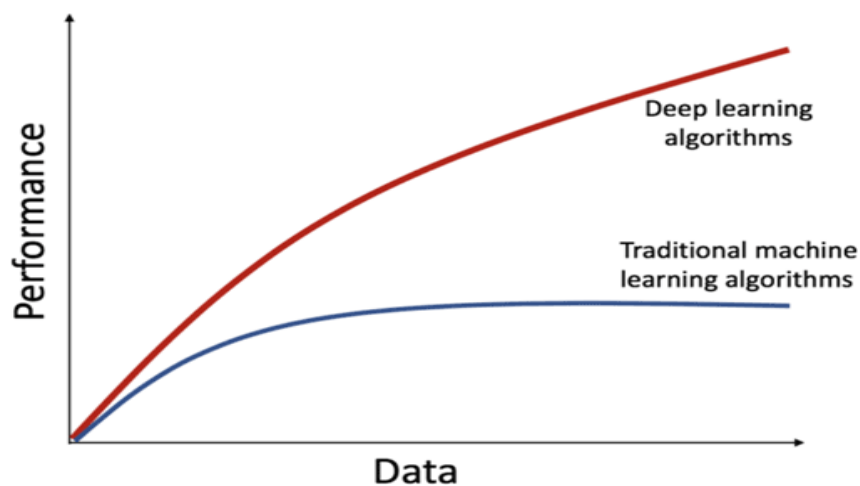


Figure 3.1 Deep Learning: Comparative Study

3.1.2 Hardware Platform

A hardware platform consists of a group of compatible hardware components that can run software applications. Programs must be designed for a specific platform, which includes a standardized type of processor and associated hardware components, as each platform has its machine language.

Table 3.3 Hardware comparison.

Method	Strengths	Weaknesses
Raspberry Pi	- Low power consumption - Fast memory - great with Python - High Speed - Have his own Linux OS	- High cost - Non-upgradable memory. - Long setup .
Arduino	- Lower cost. - No need for setup. - Lower power.	- Single-tasking. - Lower speed - Doesn't have an OS. - It doesn't deal with Python or deep learning.

Raspberry Pi and Arduino are both highly popular hardware platforms used by computer engineers, whether they are beginners or professionals. Although Raspberry Pi is mainly intended for learning computer programming, Arduino is focused on rapid programming and circuit prototyping. In this discussion, we will explore the similarities and differences between these two platforms and explain the reasons behind choosing Raspberry Pi over Arduino.

We chose **Raspberry Pi** for its multitasking processor and support for Python programming language, which was essential for developing the Deep Learning module for the Arabic handwritten character/digit recognition dataset.

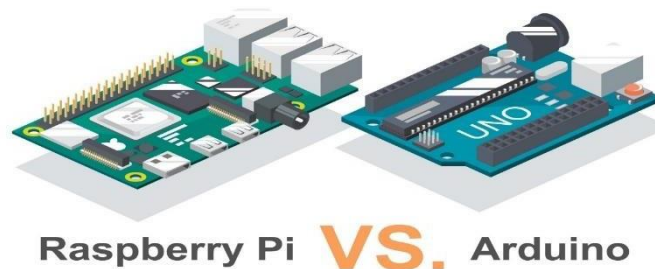


Figure 3.2 Comparison of hardware.

Table 3.4 Hardware comparison.

Raspberry Pi		Arduino
- A mini-computer with its own Raspberry OS capable of handling multiple programs at a time.	Definition	- It is a microcontroller which is just a normal part of the computer. It can be used to run a single program code again and again.
- It has 4 USB ports for establishing a connection with different devices.	Number of USB ports	- It has just 1 USB port to connect to the computer.
- Python, Scratch, Ruby, C, and C++ are the primary programming languages for developing applications on the Raspberry Pi.	Languages Supported	- C or C++ programming languages can be used to program Arduino.
- 32 bits	CPU Architecture	- 8 bits
- The Raspberry Pi processor speed- 700MHz- 1.5GHz	Processor speed.	- The average speed of most of the Arduinos is 16 MHz

3.1.3 Camera

The camera is a crucial hardware component that collects the required data for the project application.

Table 3.5 Camera (Strengths & Weaknesses).

Camera type	Strengths	Weaknesses
Raspberry Pi Camera V1	- 5 Megapixels. - Easy to install. -Lower cost than the V2.	- Lower Resolution. - requires configuration
Raspberry Pi Camera V2	- 8 Megapixels. - Easy to install.	-Thermal issues -Limited frame rate

For our project on FER for visually impaired people, we chose the Raspberry Pi Camera Module v2 over the v1 due to its superior features. The v2 camera has an 8-megapixel sensor, offering higher-resolution images compared to the 5-megapixel sensor in the v1. This higher resolution is crucial for accurately detecting facial expressions. Additionally, the v2 camera provides better image quality and color fidelity, enhancing the performance of our deep-learning models. Despite some minor limitations, the v2 camera's improvements in resolution and clarity make it the better choice for our project's needs.

So, we chose: **Raspberry Pi Camera V2**

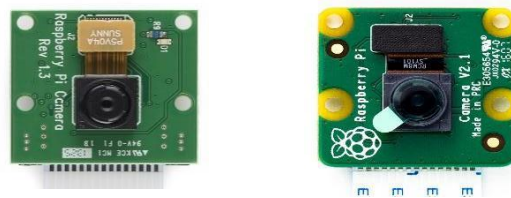


Figure 3.3 Camera V1 and Camera V2 respectively.

Section 3.2 Preliminary Work

Datasets are an essential part of any computer vision system. A benchmark dataset allows researchers to compare different machine learning approaches and techniques quickly and fairly. In this project, we utilize two of the most used datasets, namely **FER2013** and **CK+ (Extended Cohn-Kanade Dataset)**.

FER2013Dataset:

FER2013 is a publicly available dataset created for facial expression recognition tasks. It was introduced during the ICML 2013 Challenges in Representation Learning. The dataset consists of 35,887 grayscale images of 48x48 pixels, categorized into seven different facial expressions: anger, disgust, fear, happiness, sadness, surprise, and neutral. These images were collected from various sources, ensuring diversity in age, gender, and ethnicity, making it a valuable benchmark for training and evaluating facial emotion recognition models.



Figure 3.4 A sample from FER2013

CK+ Dataset (Extended Cohn-Kanade Dataset):

CK+ is a widely used dataset for facial expression recognition, containing sequences of images that capture the progression of facial expressions. It includes 593 sequences from 123 participants, with a focus on seven universally recognized expressions: anger, contempt, disgust, fear, happiness, sadness, and surprise. The images are annotated with both Action Units (AUs) and emotion labels, making the dataset ideal for training and evaluating facial expression recognition models. The data were collected under controlled laboratory conditions to ensure consistency and high-quality annotations.



Figure 3.5 A sample from CK+ dataset

Why We Chose Convolutional Neural Networks (CNNs)

For our preliminary work, we utilized **CNNs**, a cornerstone technique in deep learning, particularly effective for image-related tasks. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from images, making them highly suited for tasks like facial emotion recognition.

In this project, we focused on developing a **facial emotion recognition system for visually impaired people**, we trained our CNN model specifically for this task. This allowed us to tailor the architecture and parameters of the model to better suit the FER2013 dataset and the unique requirements of facial emotion recognition.

This approach was chosen over transfer learning to ensure the model was optimized specifically for the dataset and task at hand, without relying on pre-existing features from unrelated datasets. The trained CNN model was further integrated into our system, providing analysis of facial expressions and delivering auditory feedback to visually impaired users.

After deciding to use Convolutional Neural Networks (CNNs) for their ability to learn spatial hierarchies of features and their effectiveness in emotion recognition tasks, the next step was to select the most suitable CNN architecture for our project. To ensure optimal performance, we compared two widely recognized models, VGGNet and DenseNet121. This comparison allowed us to evaluate their strengths and weaknesses, particularly in terms of accuracy, efficiency, and suitability for real-time applications in facial emotion recognition. By analyzing their performance, we aimed to identify the model that best aligns with the requirements of our system.

The table compares the performance of VGGnet, ResNet152 on the FER2013 dataset using various evaluation metrics including precision, recall, F1-score, and accuracy.

Table 3.6 Comparisons of models on FER2013

Model	Dataset	Precision	Recall	F1-Score	Accuracy
VGGnet	FER2013	0.683	0.651	0.633	0.701
Densenet121		0.579	0.576	0.575	0.576

True positives (TP): refers to cases where the system correctly recognizes an emotion,

True negatives (TN): refers to cases where the system correctly identifies no face or no emotion.

False positives (FP): refers to cases where the system incorrectly identifies a no emotion as an emotion.

False negatives (FN): refers to cases where the system fails to recognize an emotion.

Table 3.7 tp,tn,fn and fp of densenet121

Class	Emotions	TP	TN	FP	FN
0	Angry	403	4494	438	407
1	Fear	330	4486	443	483
2	Happy	1192	3919	360	271
3	Sad	467	4162	657	456
4	Surprise	436	4947	155	204
5	Neutral	448	4390	349	555

		Confusion Matrix					
True Label	Angry	403	80	54	150	23	87
	Fear	112	330	56	178	65	66
	Happy	61	55	1192	68	30	52
	Sad	115	130	80	467	13	114
	Surprise	19	94	40	22	436	28
	Neutral	108	75	126	222	22	448
		Angry	Fear	Happy	Sad	Surprise	Neutral
		Predicted Label					

Figure 3.6 Confusion matrix of densenet121

DenseNet121 demonstrates strong capabilities in facial emotion recognition, excelling in identifying certain emotions like "Happy," which achieves a high true positive rate of 1,192. Its dense connectivity allows for efficient feature reuse and improved gradient flow, which enhances training speed and overall performance. Additionally, the model's parameter efficiency reduces the computational load compared to other deep architectures, making it suitable for tasks with limited memory. However, it struggles to distinguish between similar emotions, such as "Sad" and "Fear," as reflected by higher false positive rates and confusion in the matrix. The model's computational demands, due to its depth and dense connections, can hinder real-time applications, particularly on devices with limited processing power. Furthermore, its performance heavily depends on the quality and balance of the training dataset, with biases leading to lower accuracy for certain classes like "Fear" (TP = 330). Finally, the high complexity of DenseNet121 increases the risk of overfitting on small or imbalanced datasets, emphasizing the need for careful dataset preparation and model optimization to maximize its potential.

The Final Model (VGGnet):

Following a thorough evaluation and comparison, it was evident that VGGNet outperformed ResNet152 in terms of both accuracy and performance. This model exhibited superior results and proved to be the most suitable choice for our needs. By leveraging the power of Kaggle cloud for training and carefully selecting the VGGNet model, we successfully obtained a highly accurate and efficient solution for our requirements.

Table 3.8 Comparison of tp,tn,fn and fp on VGGnet

Class	Emotions	TP	TN	FP	FN
0	Angry	407	4740	253	342
1	Fear	323	4681	234	504
2	Happy	1251	4048	221	222
3	Sad	510	4378	432	422
4	Surprise	479	4995	85	183
5	Neutral	792	4024	706	220

The confusion matrix for its performance in this project shows its strong ability to accurately detect emotions like "Happy" and "Surprise," with high True Positive rates, demonstrating its reliability in identifying distinct facial expressions. While the model achieves impressive results overall, there is room to refine its recognition of more subtle emotions to further enhance its performance.

true label	Angry	407	46	52	99	8	131
	Fear	117	323	36	163	54	133
	Happy	22	23	1251	36	16	125
	Sad	51	74	36	510	4	256
	Surprise	18	56	35	16	479	58
	Neutral	23	27	57	111	2	792
		predicted label					
		Angry	Fear	Happy	Sad	Surprise	Neutral

Figure 3.7 Confusion matrix of VGGnet

Moreover, VGGNet's structured approach enables it to consistently process and classify complex visual data, maintaining reliability across most emotion categories. While the outcomes suggest slight limitations in recognizing subtler or overlapping emotions, the model's overall performance underscores its robustness and suitability as a foundation for building assistive technologies, especially those aimed at improving social interaction for visually impaired individuals.

Section 3.3 System Design Overview

Design Level 0:

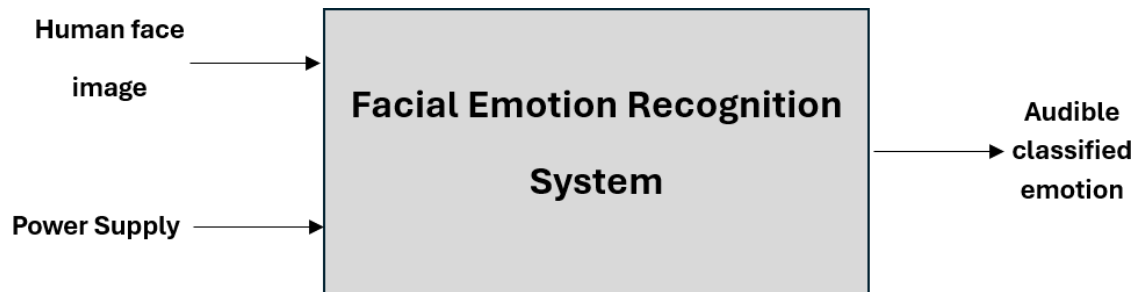


Figure 3.8 System design level 0.

Module	FER System
Input	Human face image: 8Mpixels Power Supply: 220 volts AC rms, 60Hz.
Output	Audio output signal
Functionality	Capturing an image of a person's face using a camera and detecting the face, the system extracts facial features and classifies the emotion using a trained deep learning model. Finally, it outputs the detected emotion through an audio message.

Table 1.9 Explaining the functionality of the system.

Design Level 1:

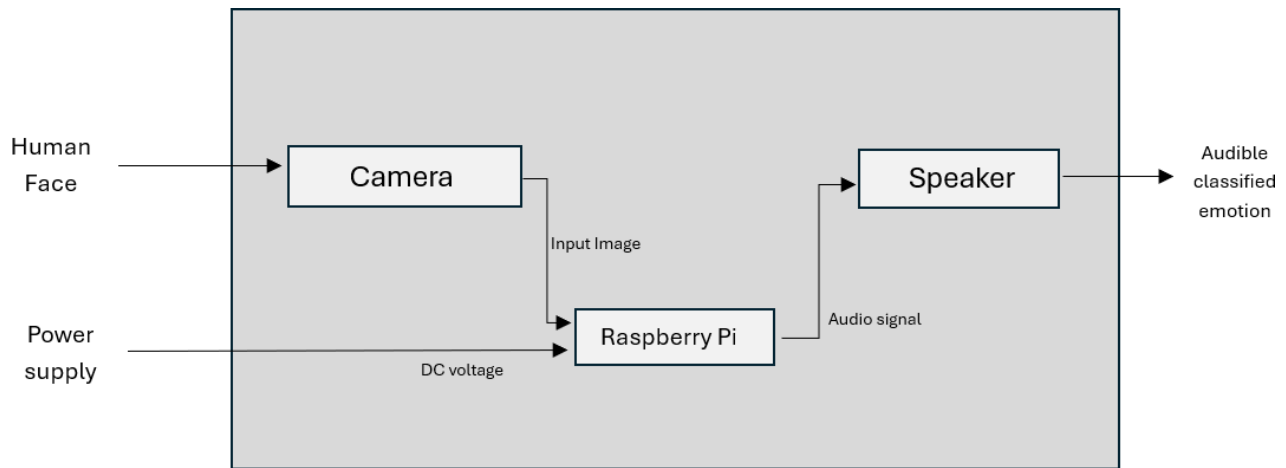


Figure 3.9 Detailed system design.

Module	Raspberry Pi Camera V2
Input	Human Face
Output	Image of the human face
Functionality	The camera captures an image of a human face and then outputs it to Raspberry Pi.

Table 3.10 Functionality of Camera.

Module	Raspberry Pi 4 model b
Input	Image of the human face DC voltage
Output	Audio signal
Functionality	The Raspberry Pi processes the input image then performs tasks like extracting facial features, identifying emotions expressed on the human face and emotion classification then outputs the classified emotion as an audio signal.

Table 3.11 Functionality of Raspberry Pi.

Module	Speaker
Input	Audio signal
Output	Audible classified emotions
Functionality	The speaker receives this audio signal and produces an audible output, allowing the user to hear the classified emotion

Table 3.12 Functionality of speaker.

Chapter 4 Design Implementation

The implementation phase is where all the components of the project come together, and we provide the necessary details to clarify our project idea. This section aims to provide a clear understanding of the project's feasibility and applicability to the intended objectives. We will also discuss any challenges encountered during the implementation process, providing a comprehensive overview of our journey.



Figure 4.1 Component of the system

Section 4.1 Prototype Implementation

Our implementation process began by installing the Raspbian operating system on a MicroSD card using the Raspberry Pi Imager software. However, compatibility issues arose with certain versions of Raspbian, as not all supported the Tensorflow library. To address this, we identified and utilized a specific 64-bit version of the Raspberry Pi's operating system that was compatible with Tensorflow.



Figure 4.2 MicroSD card and Raspberry Pi Imager

Following the successful installation of the operating system, we set up the required environment and dependencies, including Tensorflow, PiCamera2 and OpenCV libraries. Tensorflow served as the backbone for building and training our CNN, which replaced the initial plan to use Fastai and transfer learning. PiCamera2 facilitated interaction with the Raspberry Pi camera module. These tools collectively enhanced the system's functionality, enabling the implementation of a robust deep learning application.

After that, we utilized Thonny, a Python IDE, to write code for testing the model using images stored on the desktop. We initially tested the model, which was built based on a VGGNet-inspired architecture, with the FER2013 dataset and specifically took a test image of a happy face. At this stage, we had not yet connected the camera and speaker to the Raspberry Pi. By executing the code, the output was printed in the Thonny shell, allowing us to observe the model's predictions and verify its performance. This step was crucial in ensuring that the model, designed to leverage the strengths of VGGNet's structured architecture, was functioning correctly and accurately identifying emotions based on the provided test image.

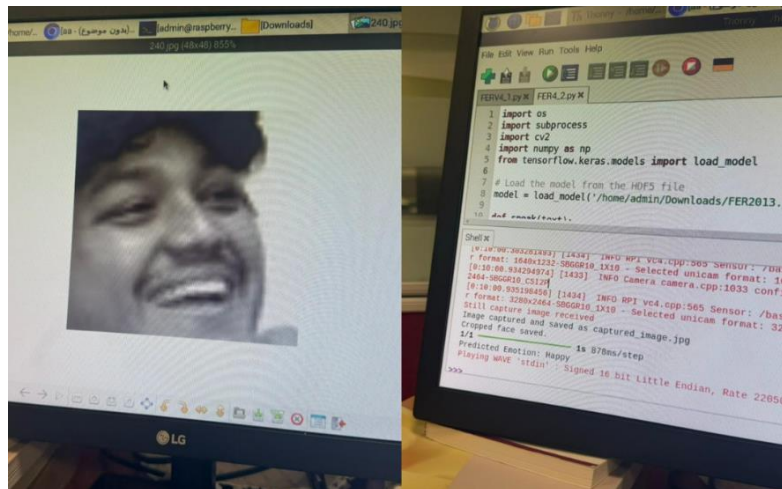


Figure 4.3 A sample from dataset and the predicted output

After assembling all the basic components of the project and setting up the system, we tested the camera on real-time images and observed the crucial importance of having a stable setup for capturing images. Ensuring that the images were properly aligned and centered was essential for accurate emotion recognition. Initially, we used a setup created with a 3D-printed Raspberry Pi 4 case that securely held all the components except the audio output as in figure 4.4. This setup provided a stable base for the camera and other essential components, preventing unwanted movement during image capture. The design of the case allowed for precise positioning and easy integration of the Raspberry Pi, camera module, and other hardware. This ensured a reliable and consistent system for testing and fine-tuning the emotion recognition model to meet our project's requirements.

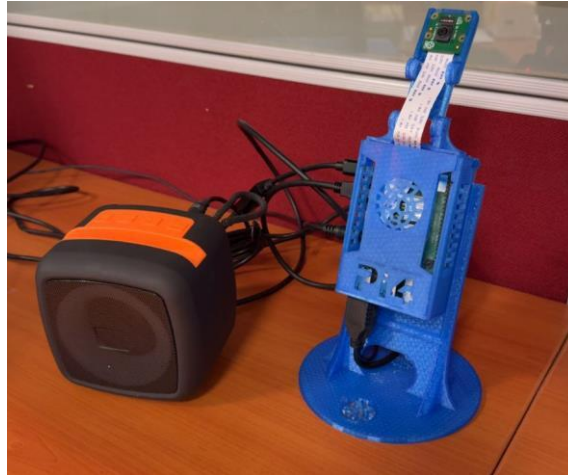


Figure 4.4 The final prototype

The user interface, shown in Figure 4.5, was developed to be simple and easy to use, ensuring visually impaired individuals can interact with the system without difficulty. It was created using Python and Tkinter, which provides a basic and reliable framework for building user-friendly interfaces. The interface includes essential features such as capturing images, analyzing facial expressions, and providing spoken feedback. With clear buttons and sound prompts, it guides users through each step. This design focuses on making the system straightforward while still being fast and effective, helping users easily understand and respond to emotions in real time.

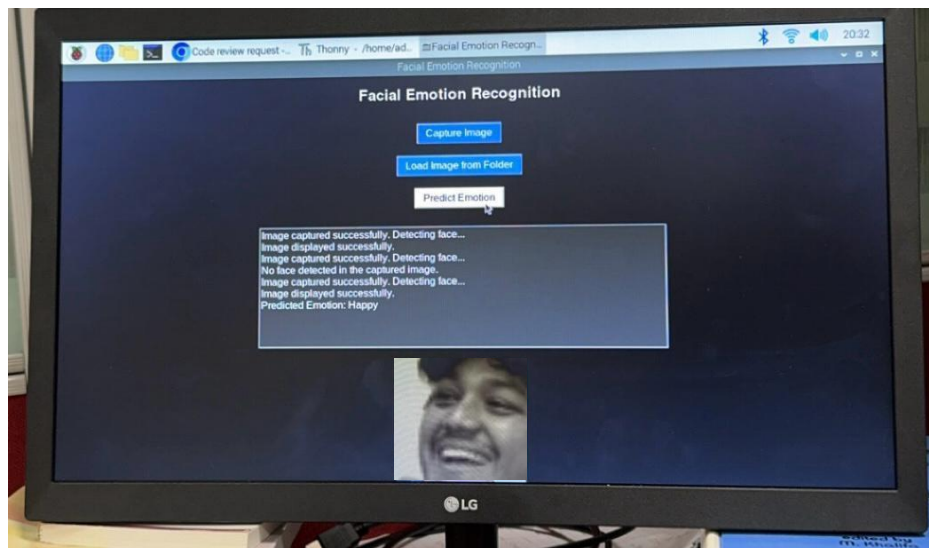


Figure 4.5 Interface of the system

Section 4.2 Prototype Verification

This section showcases the prototype testing results, which were conducted to verify the design requirements. The testing process included acceptance tests and integration tests.

Table 4.1 Acceptance Test

Test Case Name:		Facial emotion recognition System – Camera module test	Test ID #:		FERS-01	
Description:		capture the image of emotion and return the exact value of the emotion using a speaker.	Type:		■ white box	
Setup:		Connecting camera V2, and integrating space key and with speaker to the raspberry pi4 with good lighting and clear image				
Step	Action	Expected Results	Pass	Fail	N/A	Comments
1	Connect a speaker or headphone	Sound is clear	✓			
2	Capture the image	When capturing there should be a sound of camera click	✓			
3	Listening to value note	Listening to the value of the note captured by the camera.	✓			
Overall test result:			✓			

Table 4.2 Integration Test

Test Case Name:		Facial Emotion Recognition system – Camera with deep learning models	Test ID #:			FERS-02
Description:		The system's main requirement is for a Deep Learning model to recognize the emotions of an individual from the camera image.	Type:			■ black box
Setup:		Select an accurate Deep Learning model for recognizing Emotions from camera images under optimal conditions				
Step	Action	Expected Results	Pass	Fail	N/A	Comments
1	Load the deep learning model using python code	The VGGNet model should load effortlessly using Python and TensorFlow libraries on the Raspberry Pi system.	✓			Try again if its not recognized
2	Take a picture of happy	The model should accurately detect each emotion, particularly the emotion “happy” with high precision.	✓			Try again if its not recognized
3	Take a picture of sad	The model should accurately detect each emotion, particularly the emotion “sad” with high precision.	✓			Try again if its not recognized

4	Take a picture of angry	The model should accurately detect each emotion, particularly the emotion "angry" with high precision.	✓			Try again if its not recognized
5	Take a picture of neutral	The model should accurately detect each emotion, particularly the emotion "neutral" with high precision.	✓			Try again if its not recognized
6	Take a picture of fear	The model should accurately detect each emotion, particularly the emotion "fear" with high precision.	✓			Try again if its not recognized
7	Take a picture of surprised	The model should accurately detect each emotion, particularly the emotion "surprised" with high precision.	✓			Try again if its not recognized
Overall test result:			✓			

Chapter 5 Conclusion

In conclusion, this project successfully developed a facial emotion recognition system using deep learning techniques, specifically designed to assist visually impaired individuals in understanding emotions around them. The system demonstrates the value of integrating technology to enhance quality of life by providing accurate and fast emotion recognition, converting facial expressions into easy-to-understand audio outputs. The results highlight the model's efficiency in classifying primary emotions, emphasizing its role as a supportive tool for improving social interaction for this group. This project marks a significant step towards fostering inclusivity through technological innovation.

References

- [1] W. S. Liste, "Blindness and vision impairment," World Health Organization, 10 August 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>. [Accessed 4 February 2024].
- [2] Alarcos Cieza, Stuart Keel, Ivo Kocur, "World report on vision," World Health Organization, Switzerland, 2019.
- [3] Hawton A, Green C, Dickens AP, Richards SH, Taylor RS, Edwards R, Greaves CJ, Campbell JL, "The impact of social isolation on the health status and health-related quality of life of older people," National Library of Medicine, 20 February 2011. [Online]. Available: <https://link.springer.com/article/10.1007/s11136-010-9717-2#citeas>. [Accessed 4 February 2024].
- [4] Andrada-Livia Cîrneanu, Dan Popescu and Dragos Iordache, "New Trends in Emotion Recognition Using Image Analysis by Neural Networks, A Systematic Review," *Sensor*, 23 August 2023. [Online]. Available: <https://doi.org/10.3390/s23167092>.
- [5] Manas Sikri, Archit Gupta, Raghav Puri, ShivendraGoel, Nitish Pathak and Mohit Tiwari, "Emotion Detection using Image Processing in Python," in *Computing for Sustainable Global Development*, New Delhi, 2018.
- [6] ANGANA MITRA , SOUVIK CHOUDHURY and SUSMITA MOITRA, "IMAGE PROCESSING FACIAL EXPRESSION RECOGNITION," RCC Institute of Information Technology, West Bengal, 2018.
- [7] S.-W. Jung, "Facial Action Units-Based Image Retrieval for Facial Expression Recognition," IEEE, South Korea, 2019.
- [8] Hsi-Chieh Lee, Chia-Ying Wu and Tzu-Miao Lin, "Facial Expression Recognition Using Image Processing Techniques and Neural Networks," National Quemoy University, Kinmen, 2013.
- [9] Wafa Mellouka and Wahida Handouzia, "FER using deep learning: review and insights," 12 August 2020. [Online]. Available: <http://www.sciencedirect.com/>.
- [10] D. D. Patil, "FER USING DEEP LEARNING," California State University, Northridge, 2023.
- [11] Edeh Michael Onyema, Piyush Kumar Shukla, Surjeet Dalal, Mayuri Neeraj Mathur, Mohammed Zakariah and Basant Tiwari, "Enhancement of Patient Facial Recognition through Deep Learning Algorithm: ConvNet," *Journal of Healthcare Engineering*, p. 8, 2021.

- [12] Meriem Sari, Abdelouahab Moussaoui and Abdenour Hadid, "Automated Facial Expression Recognition Using Deep Learning Techniques: An Overview," *International Journal of Informatics and Applied Mathematics*, p. 15, 2019.
- [13] E. B. Kirkvik, "FER using Deep Learning," Oslo Metropolitan University, Oslo, 2022.
- [14] Hendra Kusuma, Muhammad Attamimi and Hasby Fahrudin, "Deep learning based facial expressions recognition system for assisting visually impaired persons," *Bulletin of Electrical Engineering and Informatics*, vol. 9, p. 12, 2020.
- [15] DINA SHEHADA, AYAD TURKY, WASIQ KHAN, BILAL KHAN and ABIR HUSSAIN, "A Lightweight FER System Using Partial Transfer Learning for Visually Impaired People," *IEEE Access*, vol. 11, p. 9, 2023.

Appendix A The Project Plan

Section A.1 Work Breakdown Structure

Table A.1 WBS

ID	Activity	Description	Deliverables / Checkpoints	Duration (days)	People	Resources	Predecessors
	FER			265			
	Research	Searching over the internet and research papers	Have some knowledge about the project	20	Bassam Mshari Hatem	IEEE Research Gate	
	Gather Information	Collect related information from different sources	Sort out the gathered information	20	Bassam Mshari Hatem	IEEE Research Gate	
	Update With Supervisor			3	Bassam Mshari Hatem		
	Report 1			47			
1	Problem Statement	Define the issue and how to solve it	Have a need statement, objective statement, and user needs	7	Bassam Mshari Hatem		1.1,1.2,1.3
	Update With Supervisor			1	Bassam Mshari Hatem		
2	Research Survey	Gathering data through participant responses.		10	Bassam Mshari Hatem		
3	Full Requirement Specification	Translate user needs to engineering requirements	Extract the engineering requirements	10	Bassam Mshari Hatem		3.1,3.2

4	Design	Choose the suitable components and define how to implement it	A final chosen design	20	Bassam Mshari Hatem	IEEE Research Gate Raspberry pi	4.1,4.2
	Update With Supervisor			1	Bassam Mshari Hatem		
	Buy Components	Identify parts, Place order	have the components needed to the system	3	Bassam Mshari Hatem	Amazon	
	Hardware integration	Combine all hardware components	All hardware components are combined	7	Bassam Mshari Hatem		
	Software And Coding	Writing codes needed for the system	Meet all the requirement and translate it to code	70	Bassam Mshari Hatem		
	Testing	Testing codes	Check each function and verify	14	Bassam Mshari Hatem		
	Report 2	Writing report 2		4	Bassam Mshari Hatem		
	Implementation and verification	Implement and verify the whole system	Testing each module on the field	7	Bassam Mshari Hatem		
	Summary and conclusion	Writing summary and conclusion		1	Bassam Mshari Hatem		
	Testing project	Test the whole system	Have the system work on the field	14	Bassam Mshari Hatem		

Section A.2 Gantt Chart

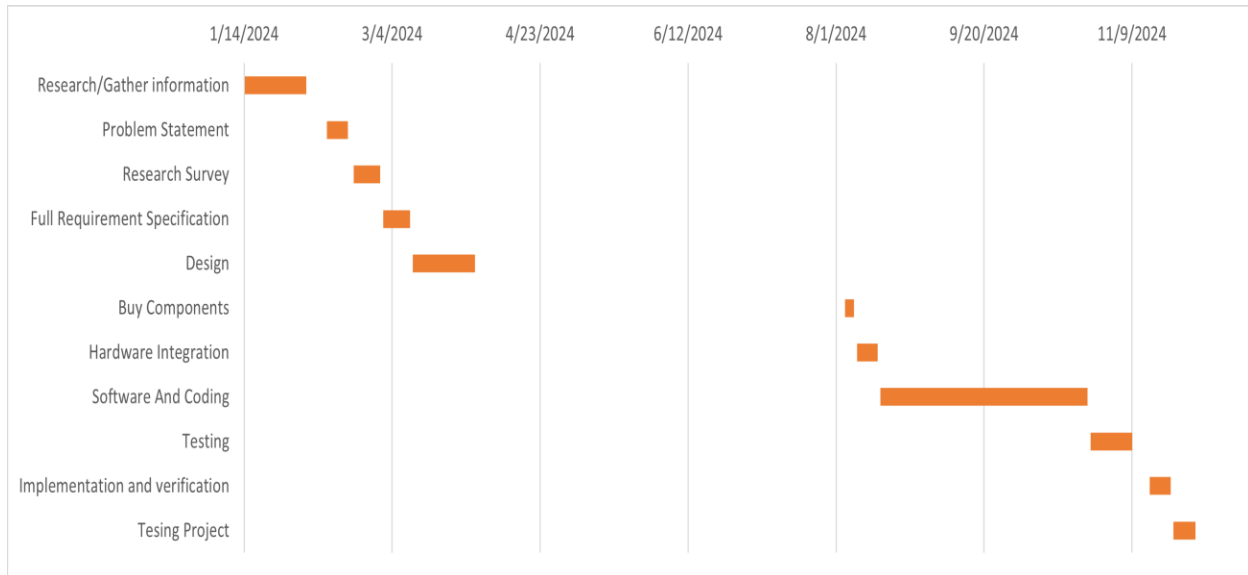


Figure A.1 Gantt Chart

Section A.3 Cost Estimation

Table A.2 Cost Estimation

Equipment	Cost
Raspberry Pi 4 + components	550 SR
Camera	115 SR
Speaker	60 SR
Total Price	725 SR