**Audible Sight**



**B.S. (SE) 2017**

**Final Year Project Report**

**Submitted by:**

**Fareena Imran 2017/Comp/BS(SE)/21593** **1720669**

**Hina Alla-ud-din 2017/Comp/BS(SE)/21600**  **1720676**

**Javeria Kamran 2017/Comp/BS(SE)/21605 1720618**

**Mariam Safdar 2017/Comp/BS(SE)/21618 1720694**

**February 24, 2021**

**Department of Computer Science and Software Engineering**

**Jinnah University for Women**

5-C Nazimabad, Karachi 74600

**Department of Computer Science and Software Engineering**

**Jinnah University For Women**

**Project Approval**

Project Title: Audible Sight

By:

Fareena Imran 2017/Comp/BS(SE) /21593 1720669

Hina Alla-ud-din 2017/Comp/BS(SE)/21600 1720676

Javeria Kamran 2017/Comp/BS(SE)/21605 1720618

Mariam Safdar 2017/Comp/BS(SE)/21618 1720694

Approval Committee:

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

Name: Tehreem Qamar

Designation: Lecturer

(Internal Advisor)

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

(Head of the Department)

# Abstract

The progressive advancement of mobile technology significantly impacted the accessibility of software functionalities to increase digital inclusion for visually impaired persons. Recent studies focused on technological advances for visually impaired in terms of accessibility and usability. Blind lifestyle has been improved a lot with the help of assistive mobile applications and the results will be more productive when these applications also have native language support. This vigorous objective is adopted in Audible Sight. Moreover, the accessibility is not much supported in many applications as there is absence of real time processing and voice input because as these features play a vital role in improving user experience for blind. The main objective of this report is to propose an assistive mobile application that works in real-time and can be customized according to the locality.

Audible sight is an assistive mobile application which is designed for Pakistan’s visually impaired people. Audible Sight is providing guidance in Urdu Language because it is the national language of Pakistan. Main features of this application are object detection, facial expression recognition, identification of known person with real time processing, text reading, color recognition and voice input support. Object detection feature will recognize the object in real time also the application will read the expression of people and narrate the results in Urdu language. In Person recognition module, if the user’s camera captures the image of a person who is relative of the blind then application will vocally inform about person existence. Text reading and color recognition features will help them to read short text and recognize colors. This application is aimed to help the blind to be more aware of their surroundings and regain their independence.

# Table of Content

[Abstract i](#_Toc65101355)

[Table of Content ii](#_Toc65101356)

[List of Figures v](#_Toc65101357)

[List of Tables vi](#_Toc65101358)

[Acknowledgment vii](#_Toc65101359)

[Chapter 1 1](#_Toc65101360)

[Introduction 1](#_Toc65101361)

[1.1 Project Overview 1](#_Toc65101362)

[1.2 Project Deliverables 3](#_Toc65101363)

[Chapter 2 4](#_Toc65101364)

[Theoretical Background 4](#_Toc65101365)

[2.1 Introduction to Object Detection and Facial Expression Recognition 4](#_Toc65101366)

[2.2 Techniques For Implementation 5](#_Toc65101367)

[2.2.1 Object Detection 5](#_Toc65101368)

[2.2.2 Facial Expression Recognition 5](#_Toc65101369)

[2.3 Literature Review 5](#_Toc65101370)

[Chapter 3 8](#_Toc65101371)

[Software Requirements Specifications 8](#_Toc65101372)

[3.1 Purpose 8](#_Toc65101373)

[3.2 Product Scope 8](#_Toc65101374)

[3.3 User Classes and Characteristics 8](#_Toc65101375)

[3.4 Operating Environment 9](#_Toc65101376)

[3.5 Design And Implementation Constraints 9](#_Toc65101377)

[3.6 Assumption and Dependencies 9](#_Toc65101378)

[3.6.1 Technical Assumption 9](#_Toc65101379)

[3.6.2 Business Assumption 9](#_Toc65101380)

[3.6.3 Assumption About User 9](#_Toc65101381)

[3.7 functional Requirements 10](#_Toc65101382)

[3.8 Other Nonfunctional Requirements 11](#_Toc65101383)

[3.8.2 Usability 11](#_Toc65101384)

[3.8.3 Reliability 11](#_Toc65101385)

[3.8.4 Learnability 11](#_Toc65101386)

[3.8.5 Maintainability 11](#_Toc65101387)

[3.9 Security Requirements 11](#_Toc65101388)

[3.10 Software Quality Attributes 11](#_Toc65101389)

[Chapter 4 13](#_Toc65101390)

[System Analysis and Design 13](#_Toc65101391)

[4.1 Relational Diagram 13](#_Toc65101392)

[4.2 Entity Relationship Diagram 14](#_Toc65101393)

[4.3 UML Diagrams 15](#_Toc65101394)

[4.3.1 Use Case Diagram 15](#_Toc65101395)

[4.3.2 Activity Diagram 17](#_Toc65101396)

[4.3.3 Class Diagram 18](#_Toc65101397)

[4.3.4 Data Flow Diagram 19](#_Toc65101398)

[4.3.5 Deployment Diagram 21](#_Toc65101399)

[Chapter 5 22](#_Toc65101400)

[Tools and Technologies 22](#_Toc65101401)

[5.1 Software Tools & Technology 22](#_Toc65101402)

[5.1.1 Android Studio 22](#_Toc65101403)

[5.1.2 TensorFlow 22](#_Toc65101404)

[5.1.3 Google Vision API 22](#_Toc65101405)

[5.2 Designing Tools 22](#_Toc65101406)

[5.2.1 Just in Mind 22](#_Toc65101407)

[5.2.2 Photoshop 23](#_Toc65101408)

[5.3 Documentation Tools 23](#_Toc65101409)

[5.3.1 MS Word 23](#_Toc65101410)

[5.3.2 MS Visio 23](#_Toc65101411)

[Chapter 6 24](#_Toc65101412)

[User Interface Design 24](#_Toc65101413)

[6.1 Web User Interfaces 24](#_Toc65101414)

[6.1.1 Home Page 24](#_Toc65101415)

[6.2 Software Interfaces 26](#_Toc65101416)

[6.3 Mobile Application Interfaces 26](#_Toc65101417)

[Chapter 7 30](#_Toc65101418)

[Methodology 30](#_Toc65101419)

[7.1 Project Plan 30](#_Toc65101420)

[7.1.1 Develop overall domain model 30](#_Toc65101421)

[7.1.2 Build List of Features 30](#_Toc65101422)

[7.1.3 User Stories 30](#_Toc65101423)

[7.1.4 Plan by Feature 31](#_Toc65101424)

[Chapter 8 34](#_Toc65101425)

[Implementation 34](#_Toc65101426)

[8.1 Database Design 34](#_Toc65101427)

[8.2 implementation technique 34](#_Toc65101428)

[8.3 Website Design 35](#_Toc65101429)

[8.4 Mobile Application Design 35](#_Toc65101430)

[8.5 Deployment 36](#_Toc65101431)

[Chapter 9 37](#_Toc65101432)

[Testing 37](#_Toc65101433)

[9.1 Test Plan 37](#_Toc65101434)

[9.2 Unit Testing 37](#_Toc65101435)

[9.3 Integration Testing 37](#_Toc65101436)

[9.4 Functional Testing 37](#_Toc65101437)

[9.5 System Testing 38](#_Toc65101438)

[9.6 Stress Testing 38](#_Toc65101439)

[9.7 Performance Testing 38](#_Toc65101440)

[9.8 Acceptance Testing 38](#_Toc65101441)

[9.8 Test Cases 39](#_Toc65101442)

[9.8.1 GUI Testing 39](#_Toc65101443)

[9.8.2 Object Detection Module Testing 39](#_Toc65101444)

[9.8.3 Emotion Recognition Module Testing 40](#_Toc65101445)

[9.8.4 Color Recognition Module Testing 41](#_Toc65101446)

[9.8.5 Text Reading Module Testing 41](#_Toc65101447)

[9.8.6 Person Identification Module Testing 42](#_Toc65101448)

[9.8.7 Stress Testing 43](#_Toc65101449)

[9.8.8 Voice Testing 43](#_Toc65101450)

[Conclusion 44](#_Toc65101451)

[References 45](#_Toc65101452)

# List of Figures

[Figure 4.1: Relational Diagram 13](#_Toc65101701)

[Figure 4.2: Entity Relationship Diagram 14](#_Toc65101702)

[Figure 4.3: Use case Diagram 15](#_Toc65101703)

[Figure 4.4: Activity Diagram 17](#_Toc65101704)

[Figure 4.5: Class Diagram 18](#_Toc65101705)

[Figure 4.6: Data Flow Diagram - Level 0 19](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101706)

[Figure 4.7: Data Flow Diagram - Level 1 19](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101707)

[Figure 4.8: Data Flow Diagram - Level 2 20](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101708)

[Figure 4.9: Deployment Diagram 21](#_Toc65101709)

[Figure 6.1: Splash Screen 26](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101710)

[Figure 6.2: Object Detection Screen 26](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101711)

[Figure 6.3: Person Detection Screen 27](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101712)

[Figure 6.4: Input name Dialog Box 27](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101713)

[Figure 6.5: Person Recognition Screen 28](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101714)

[Figure 6.6: Color Recognition Screen 28](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101715)

[Figure 6.7: Text Reading Screen 29](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101716)

[Figure 6.8: Emotion Recognition Screen 29](file:///D:\Hina\FYP\FYP%20-%20Project%20Report\FYP%20Project%20Final%20Report%20-%20Audible%20Sight.docx#_Toc65101717)

[Figure 9.1: App on Play Store 36](#_Toc65101718)

# List of Tables

[Table 4.1: Use case Description 16](#_Toc65101826)

[Table 7.1: Project Plan 31](#_Toc65101827)

[Table 7.2: Sprint Plan 31](#_Toc65101828)

[Table 9.1: GUI Testing 39](#_Toc65101829)

[Table 9.2: Object Detection Module Testing 39](#_Toc65101830)

[Table 9.3: Emotion Recognition Module Testing 40](#_Toc65101831)

[Table 9.4: Color Recognition Module Testing 41](#_Toc65101832)

[Table 9.5: Text Reading Module Testing 41](#_Toc65101833)

[Table 9.6: Person Identification Module Testing 42](#_Toc65101834)

[Table 9.7: Stress Testing 43](#_Toc65101835)

[Table 9.8: Voice Testing 43](#_Toc65101836)

# Acknowledgment

A victorious story is incomplete without remembering those who have inspired it. We would like to express our gratitude towards all those individuals who extended their support in preparing this project which was indeed a great learning process for us.

During the process of this project, it was a wonderful experience to be a part of Computer Science and Software Engineering, JUW, where we have opportunity to work under the brilliant and supportive staff. We express our sincere gratitude to the supporting staff authorities who have been very helpful in the successful completion of our project.

We are heartily indebted to Ms. Tehreem Qamar who guided us throughout the project and shared valuable suggestions and extended great encouragement to us. She is always ready to help with a smile and we are very much obliged to her for her advice, guideline and support during our project.

# Chapter 1

# Introduction

# 1.1 Project Overview

Audible sight is an assistive smart mobile application which is a highly innovative solution for Pakistan’s visually impaired people. Over 2 million people are blind in Pakistan, there are many applications specifically designed for blind but we have observed that these applications are not feasible for such people who are not familiar with English Language. Audible Sight is providing guidance in Urdu Language because it is most understandable language in Pakistan. Main features of this application are object detection, facial expression recognition, text reading, identification of known person and color recognition. Object detection feature will recognize the object in real time also the application will read the expression of people and narrate the results in Urdu language. In Person recognition module, first the users are required to save image of the person (either friend or family) along with his/her name. If the user’s camera captures the image of a person whose image is saved then application will vocally inform about person existence. Color recognition feature will recognize colors even their different shades. Text reading module will read small text clearly. This project is aimed to help the blind to be more aware of their surroundings and regain their independence. In this project, we will use mobile camera to collect environmental data. Data obtained will be forwarded for further processing to the application. Using text-to- speech library, processed data will be converted to the sound.

This project is based on **Machine Learning** in which we have used TensorFlow API for object detection, person identification and emotion recognition. For text reading Google vision API is used. Camera API with JSON is used for the implementation of color recognition feature. We have combined the functionality of these APIs into single Application to achieve our objective. Machine learning is the subset of AI in which we train our systems on vast amount of data so that they learn from data and carry out tasks based on their learning.

**Artificial Intelligence** is the intelligence of machines. It’s making machine intelligent so that they can perform tasks like humans. AI is a broader view of making machines capable to carryout tasks smartly as humans do. AI is very fast field which holds many other research areas within it like Machine Learning, Neural Networks, Robotics, Evolutionary computing, Vision, Natural language processing. Neuron nodes connected together like a web to form artificial neural network similar as human brains. The series of neural network algorithms are constructed to recognize patterns and these patterns are trained on the data collected from the real world which comprises of text, images, sounds or time series. These algorithms are used to cluster the unlabeled data according to the similarities and then classify it when it is trained on labelled datasets. In order to learn the correlation between data and label, neural network have to learn from the knowledge conveyed by humans that will be help to classify labelled datasets. This is known as **supervised learning**. The few concrete example of Supervised Learning is Detection of faces, identification of people in images, recognition of facial expressions (angry, joyful) and Identification of objects in images (stop signs, pedestrians etc.) which are the main focus of Audible Sight Application.

CNN are commonly used for implementation of AI in image processing and solving problems through image classification and image recognition. CNN an artificial neural network is a kind of deep neural network which were designed from the biologically driven models so the researchers found how a mammal, how a human perceives an image into the brain is in different layers and that’s how this CNN was designed and hence this is true and very efficient for all the image processing pattern recognition kind of application. CNN is most popularly used for analyzing images. It can also be used for other data analysis or classification problem as well. Faster R-CNN is updated version of CNN.

In CNN, an input image goes through different layers. First the input layer goes through a bunch of layers, called convolutional layer. This layer takes a patch from input and apply a set of filters to the given input then this activated data goes into another layer which called “POOLIM” layer. Then goes to another convolution layer then POOLIM layer and so on. At the end of the network generally there are layers called fully connected layers. In fully connected layers each and every node is connected to the next node in the coefficients that is a heavy data-driven load where there are lot of coefficients which are loaded to support each and every node in the POOLIM data. The Fully connected layer at the end there are multiple sets of output from POOLIM layer then this layer draw top 3 or top 5 best cases for the object into consideration. At the end fully connected layer is picks the highest probability.

# 1.2 Project Deliverables

Audible Sight is a mobile application for assistance. The camera of the smart phone is used to capture the image and provide information in real time.

This project brings the power of Artificial Intelligence to deliver an intelligent Mobile app, designed to help visually impaired people. Here, we are using the TensorFlow’s object detection API to enable real time detection and google translation API to narrate the results in Urdu language. The result is that user will be able to know about front objects and recognizes friends and people around them, including their emotions, easily read text without the need to have any brail version of it along with the recognition of colors.

# Chapter 2

# Theoretical Background

# 2.1 Introduction to Object Detection and Facial Expression Recognition

Many researches have been done aiming to provide facilities to disabled. Focusing specifically on blind a lot of work has been done and still researches are under working that are addressing to the solutions to the problems of the blind and visually impaired [1]. There are several researches that had worked on facial recognition [2] and object detection [3] feature. They have used different techniques and several algorithms on the basis on their perspective. Like some main focused on providing efficiency while few focused-on accuracies along with the efficiency.

Recently, face recognition has gained much importance in IT world [4]. As it can be solved for many purposes like for biometric verification, identifying in identification of person on social networking site. Despite all these different usage, facial recognition came under the consideration of experts for using I for the well-being of people with the help of technology. At initial stage, it was used with minimum data set, but with its popularity developers and researchers aimed to work on it more willingly to produce fruitful results. Several different algorithms were used to increase the efficiency and accuracy of facial recognition. Many imaging techniques are also used along with the data-set libraries and algorithms.

At initial stage, object detection was manly focused for security and surveillance purposes [5], but many researchers that work for the development on the social welfare applications consider it of worth importance. Deep see and neural network combinedly were first into consideration for detecting objects at real time [6]. After its testing with the perspective if its use for visually impaired and blinds, this combination didn’t satisfy many technology experts with their result. So, with time many other techniques and algorithms are now in used for this purpose.

# 2.2 Techniques For Implementation

### **2.2.1 Object Detection**

* Object detection is defined as a set of related tasks for recognizing objects in digital images.
* Image segmentation and blob analysis, which uses simple object properties such as size, shape, or color are mainly used techniques.
* Feature-based object detection, which uses feature extraction, matching, and RANSAC to estimate the location of an object
* Region-Based Convolutional Neural Networks, or R-CNNs, are a family of techniques for addressing object localization and recognition tasks, designed for model performance.
* For fast and real time object detection, YOLO (You Only Look Once) is also the most used technique to build intelligent applications.

### **2.2.2 Facial Expression Recognition**

* Histogram Oriented Gradient, its accuracy is 98% with ORL database
* Fusion algorithm, its accuracy is 99.07% with UGC-JU database
* Modified local directional pattern, its accuracy is 96.25% with Deep Belief Network database
* LBP technique, shape model, its accuracy is 88.76% with PHPID database
* Principal Component Analysis PCA.

# 2.3 Literature Review

This section covers the related work of different areas of research related to techniques and algorithm use for object recognition and facial recognition. First, we studied different research paper then focus on our specific goal of supporting efficient algorithm on the basis of accuracy, frame per second and efficiency. We concluded some efficient algorithm that are YOLO, SSD, Faster R-CNN and PCA. These algorithms are mostly used for object recognition, person identification and facial recognition. Summarized details of research papers are described as follows.

In paper [5], Pawan Mishra, Gyanendra Saroha concluded that there are many technologies built for blind users to make their life easy but some have some limitations regarding to technologies or devices usage. By observing all these limitations, a system is proposing in which blind can use camera to recognize object that is android smartphone. The technologies discussed in this paper using OpenCV library and Google cloud vision API for image processing. The dataset used to store image is COCO dataset.

In paper [7], Prof. Hassan Soliman, Ph.D, Ahmed Saleh,Ph.D, Eman Fathi proposed face recognition task in which the traditional Principle Component Analysis (PCA) is used for performing recognition task and Fuzzy based Pixel wise Information Extraction (FPIE) and PCA algorithms are used. Its System accuracy is 92% in which the time taken to recognize face is 0.35 sec. The images used in this are grey scale. For experiments, the testing set cover 25 images, Images include persons that belong to the training set.

In paper [8], Aesha Shah, Kavin Shah, Vidhi Shah, Chintan Shah presented the deployment of face recognition in mobile devices using Viola-jones Algorithm and PCA algorithm. Viola-Jones is used for face detection and PCA is used for face recognition. Face detection algorithm takes at least 2 frames per second for practical applications and the goal is to distinguishes faces from non-faces and detection rate is very high. They have used PCA because of its speed and straightforwardness.

In research paper [9], Chucai Yi, Yingli Tian, Aries Arditi aimed at developing a system that will detect objects in real time in high speed and then convey the results to blind user through audio or voice assistance. For higher efficiency and accuracy Faster RCNN (Region-based Convolutional Neural Network) algorithm has been used. Implementation was tested on two different models SSDMobileNetV2 and MobileNetV1 bur the SSDModelV2 yielded more efficient result with 35% increase in accuracy as compared to the MobileNetV1.

In paper [10], Ruxandra Tapu, Bogdan Mocanu, Titus Zaharia demonstrated that an efficient framework name “DEEP-SEE" is integrated in assistive device for visually impaired. This framework utilizes CNN (Convolutional Neural Network) and computer vision algorithms. This framework work for monitoring and recognizing objects in real time. Computer vision algorithm (YOLO) works with tracking procedure to detect object that make the system fast that is it detect objects more than 100fps and CNN is used for visual patterns and motions. The accuracy of this assistive device is greater than 90%.

# Chapter 3

# Software Requirements Specifications

# 3.1 Purpose

The purpose to develop this project is to solve the problems of blind and visually impaired people. Audible sight is focusing on Pakistan’s visually impaired. Main purpose is aimed to help blind by giving them opportunities as others, to fully participate in society.

Provide them assistance verbally to understand what’s coming in front, awareness of facial expression of people in Urdu language. Audible Sight is aimed to give a platform to make blind and visually impaired more independent, provide equal opportunities as sighted people in society, confidence to enhance social circle. Team is determined to make the environment more accessible and remove the feeling of discomfort from blind and visually impaired during interactions with people.

# 3.2 Product Scope

Audible Sight is an Android application which helps blind and visually impaired people to detect object, identify known person, recognize facial expressions and colors and text reading. The main focus of our project is to facilitate the blind community of Pakistan which faces language barrier in using technology by providing voice assistance in Urdu language throughout the application. This application will be free to download for android phones from play store.

Furthermore, the application does not require internet connection for any of its feature. All the information regarding the application is provided on Audible sight website. There is also a contact us option in website to concern for any help form us.

# 3.3 User Classes and Characteristics

Blind and visually impaired people are the users of our application. Each of these two-user use application for separate purpose so their requirements are different from other one.

Ever thought how the life of a blind person, their life is full of risk. They can’t even walk alone through a busy street or through a park. They shall need some assistance from others. They are also curious about the beauty of the world, they want to explore the world like others, and to be aware of what is happening in front of them. Even though they can find their own things without anyone’s need. The blind people can use application for Face Detection, Facial expression recognition, Object detection, Color Recognition and Text Reading.

# 3.4 Operating Environment

Platform: Android

# 3.5 Design And Implementation Constraints

* The project will be mobile based application.
* Two languages English and Urdu will be supported for voice commands.
* The application must be implemented in Java.
* Users have to download the application on their phone in order to use it.
* The smartphone shall have at least 5 MP camera resolutions.
* Cost of the smartphone should be in a reasonable margin to be affordable.
* Battery life and recharging time of camera are also important. Recharging should not be more than 10 hours and also battery life should not be less than 6 hours.

# 3.6 Assumption and Dependencies

Assumption is something that you [accept](https://dictionary.cambridge.org/dictionary/english/accept) as [true](https://dictionary.cambridge.org/dictionary/english/true) without [question](https://dictionary.cambridge.org/dictionary/english/question) or [proof](https://dictionary.cambridge.org/dictionary/english/proof). Assumption can basically depend on the knowledge, current information and developer experience. That’s include

### **3.6.1 Technical Assumption**

We work on latest version of software, tools and technology (for example we are working on latest version of Android Studio).

### **3.6.2 Business Assumption**

Product develops in the determined budget and cover all the project expenses.

### **3.6.3 Assumption About User**

Blind people easily use our all module without any difficulty (for example that blind user of our application use the app frequently to get guidance).

# 3.7 functional Requirements

Audible Sight is a mobile application that provide a feasible solution for blind people. Audible Sight will provide guidance in Urdu Language. In our project we will use AI for operation. Here, we are using the TensorFlow’s object detection API to enable real time detection and Firebase natural language API to narrate the results in Urdu language. The user will be able to detect object and recognize friends and people around them, including their emotions, also can recognize colors and read short text.

**3.7.1 Object Detection**

* The system will have a good response time for real time object detection.
* The system will narrate the object name in Urdu language.
* This module of system will also detect the person.

**3.7.2 Facial recognition**

* The system will recognize the facial expression in real time.
* The system will allow user to save a particular person image.
* The system will narrate the person’s name in Urdu language when identified.

**3.7.3 Emotion Recognition**

* The system will recognize the happy expression in real time.
* The system will recognize the sad expression in real time.
* The system will recognize the angry expression in real time.
* The system will recognize the scared expression in real time.
* The system will recognize the neutral expression in real time.
* The system will recognize the surprised expression in real time.
* The system will narrate the expressions in Urdu language.

**3.7.4 Color Recognition**

* The system will recognize the several different colors.
* The system should narrate the color name when user will tap on the screen.

**3.7.5 Text Reading**

* The system will detect and extract text in real time.
* The system will read computer generated text only.

# 3.8 Other Nonfunctional Requirements

**3.8.1 Efficiency**

The system must be efficient to perform tasks and returns the result in minimum amount of time.

### **3.8.2 Usability**

The interface of mobile application must be accessible for blinds user rather than intimidating and frustrating; it must be easy to use.

### **3.8.3 Reliability**

Audible sight system must be reliable under load. It should not stop working when used in real environment.

### **3.8.4 Learnability**

The system must be easy to learn for user by its consistent and simple user interface.

### **3.8.5 Maintainability**

The data used at the back of the system must be maintained and updated time to time.

# 3.9 Security Requirements

Security systems need database storage just like many other applications. User authorization and data encryption are important security requirements of the project. Users are providing their data like their images. System should store user data on database securely and set access permissions to these data carefully.

# 3.10 Software Quality Attributes

**3.10.1 Correctness**

the system must be according to its specifications for correctness to satisfy the customers need and must be appropriate to work independently.

**3.10.2 Extensibility**

The system must be extensible so that the changes should be made without causing any side effects and disturbance in its smooth working.

**3.11.3 Testability**

The system must have a modularize structure so that testing and debugging could easily be done to monitor system behavior at different conditions.

**3.11.4 Integrity**

The system integrity or security must be followed. System must be free from any sort of viruses, privacy and data loss threats.

# Chapter 4

# System Analysis and Design

# 4.1 Relational Diagram

This diagram is demonstrating the representation of relational database of Audible Sight comprises of the following tables as shown in figure 4.1. There are five entities which are camera, TF\_Model, PersonFace, Emotion, Object. One camera can use many TF\_Model but only one TF\_Model can operate camera at a time. Many person faces can be detected by one TF\_Model. These TF\_Model can also detect number of emotions and objects but for these detections one TF\_Model can be used at a time.

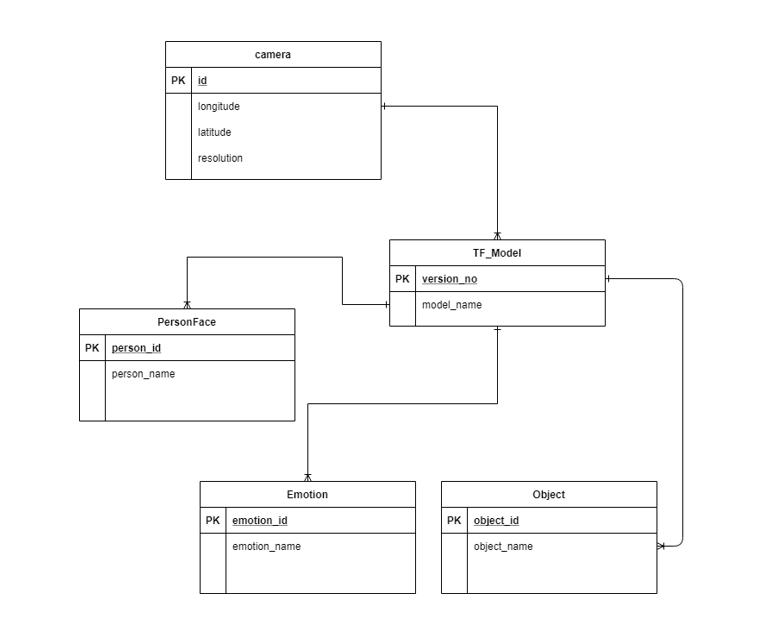


Figure 4.1: Relational Diagram

# 4.2 Entity Relationship Diagram

This diagram is demonstrating a visual representation of the relationships between different entities. TF\_Model is recognizing image captured by the camera. TF\_Model can recognize person, emotion and object with respect to the captured image. One camera can use many TF\_Model but only one TF\_Model can operate camera at a time. Many person faces can be detected by one TF\_Model. These TF\_Model can also detect number of emotions and objects but for these detections one TF\_Model can be used at a time.

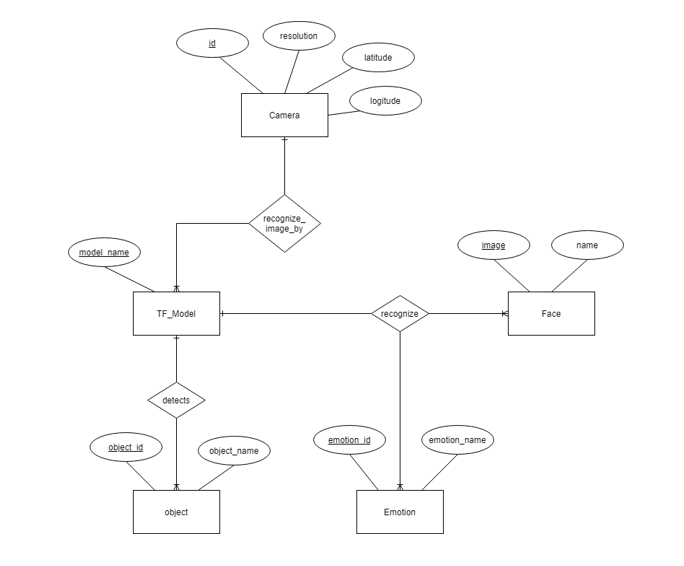


Figure 4.2: Entity Relationship Diagram

# 4.3 UML Diagrams

### **4.3.1 Use Case Diagram**

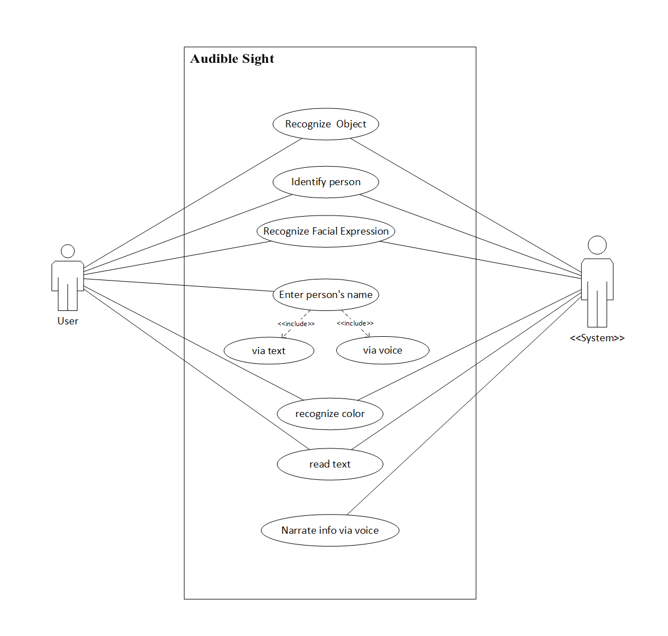


Figure 4.3: Use case Diagram

Table 4.1: Use case Description

|  |  |  |
| --- | --- | --- |
| **Use case** | **Actor** | **Description** |
| Facial expression recognition | System, User | This use case describes the process of recognizing facial expression of a person who comes in front of camera. |
| Identify Person | System, User | This use case describes the process of person identification. The system will recognize persons and narrate their names when they will be subjected towards the camera. |
| Input person’s name | User | This use case describes that the user needs to input person’s name to save his/her image and name in the database. User can enter name by typing or voice input. So that the system will be able to recognize person when he/she comes in front of camera. |
| Recognize object | System, User | This use case describes the process of recognize object that comes in front of camera. So that blind user will be able to walk independently. |
| Narrate info. Via voice | System | This use case describes the process of narrating object, person and emotions in native language. So that user will be able to easily understands about their surroundings. |
| Recognize Color | System, User | This use case describes the process of color recognition. The system can recognize 1600+ different colors. |
| Read Text | System, User | This use case describes the process of Text Reading. When the camera is subjected to any computer-generated-text, this will be converted to speech. |

### **4.3.2 Activity Diagram**

The sequence of activities involved in mobile application is demonstrated in this diagram. When user will open mobile application, the user will be able to detect objects in indoor environment then user can switch between the activities by clicking on respective button to use all the available features.

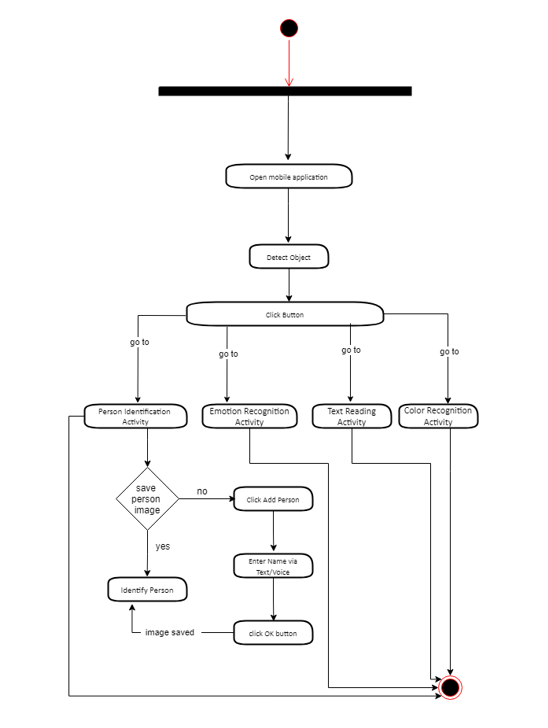


Figure 4.4: Activity Diagram

### **4.3.3 Class Diagram**

This diagram depicts the static structure of project comprises of five classes. Camera class having attribute of Id, longitude, latitude, resolution and is responsible for capturing images. TF\_Model comprises of attributes named as version\_no, model\_name and has compare accuracy and create bounding boxes. Object class contain attributes object\_id, object\_name and this can detect, count and recognize objects. PersonEmotion class consist of attributes named as emotion\_id, emotion\_name and is responsible for recognizing emotions. PersonFace class having attributes of person\_id, person\_name and is responsible for recognizing person faces and can save person’s name.

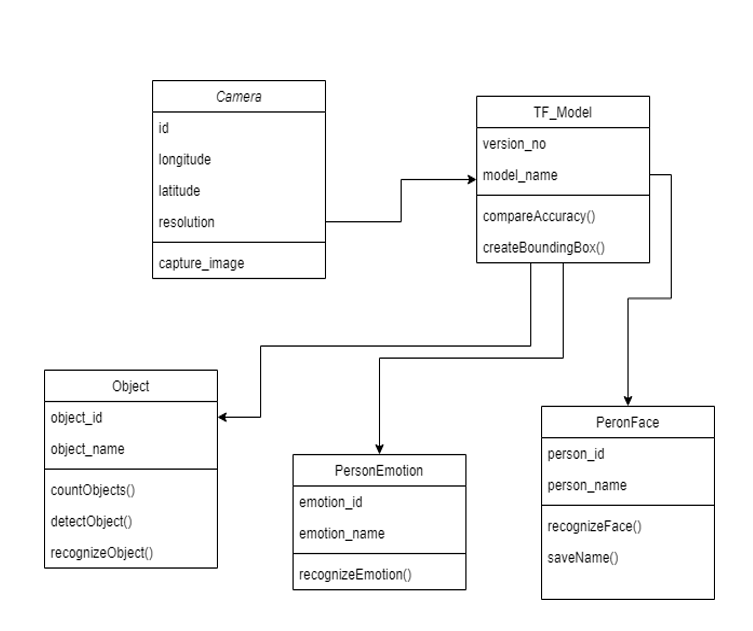


Figure 4.5: Class Diagram

### **4.3.4 Data Flow Diagram**

This **DFD level 0** is showing the interaction of user with the system which comprises of different models implemented in the application.

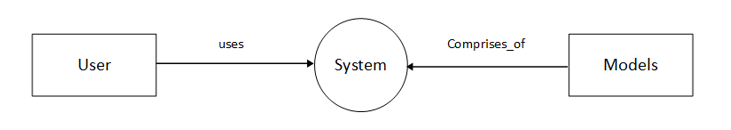


Figure 4.6: Data Flow Diagram - Level 0

This **DFD level 1** is briefly describing the interaction of user with system that user can perform object detection through camera, can recognize emotions and colors, read text and identify persons by using application.

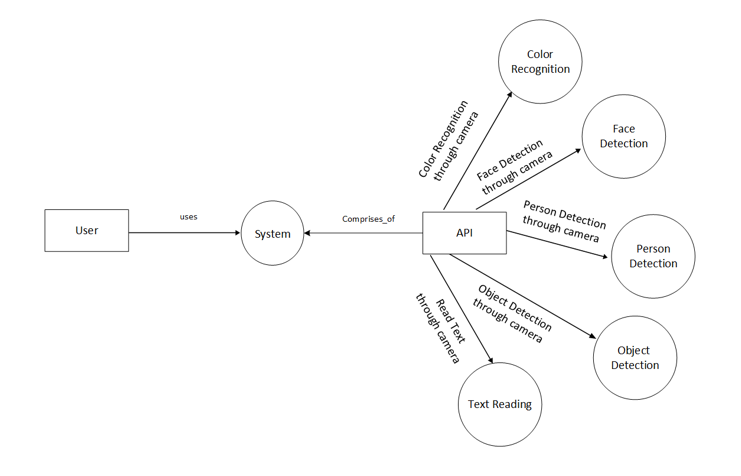


Figure 4.7: Data Flow Diagram - Level 1

This **DFD level 2** is briefly describing the processes involved in system. For identification of person, user will save person’s name and image. Audible sight database contains trained data for object detection and emotion recognition. System will then provide information on the basis of saved data by comparing screen’s object and trained data. With the help of Google Vision API, the system will detect and extract text from electronic documents. Also, the system will recognize the focused colors from saved hexadecimal color code having 1600 different colors.

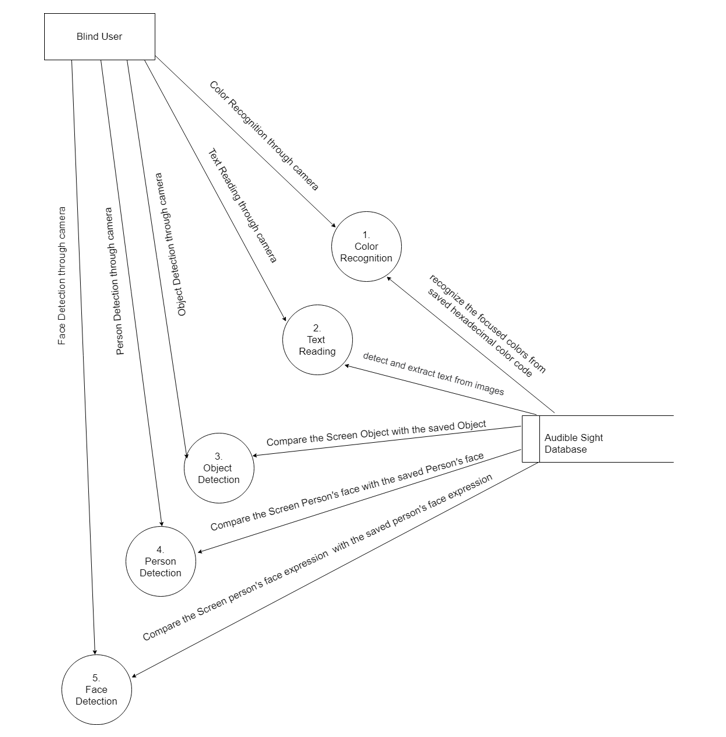


Figure 4.8: Data Flow Diagram - Level 2

### **4.3.5 Deployment Diagram**

This diagram is demonstrating the static deployment of the system. The application is executable in android environment, the android device version must be 8 or above. The application is deployed with a set of different APIs which are responsible for working of system’s features.

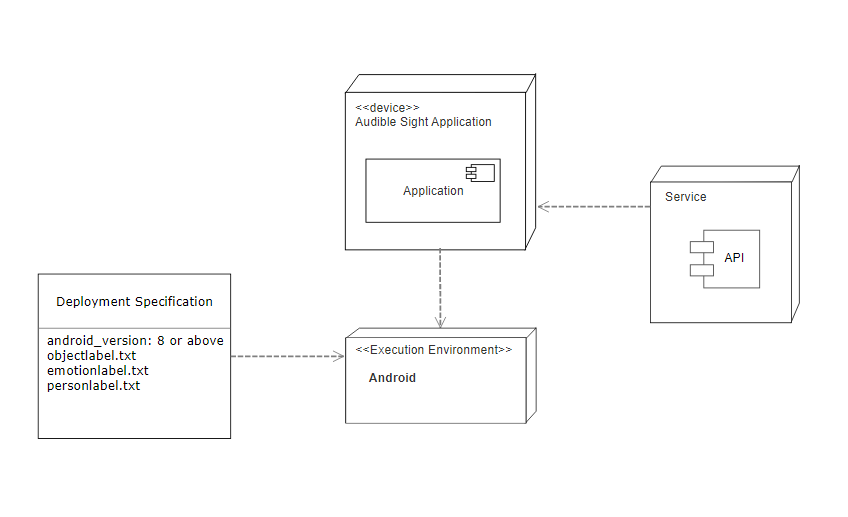


Figure 4.9: Deployment Diagram

# Chapter 5

# Tools and Technologies

# 5.1 Software Tools & Technology

### **5.1.1 Android Studio**

Android Studio is Android's official IDE. It is designed for Android to accelerate development and help build the highest-quality apps for every Android device. Audible Sight is an android based mobile application, So, we have used it because it offers tools for Android developers, including rich code editing, debugging, testing, and profiling tools.

### **5.1.2 TensorFlow**

TensorFlow is an open-source machine learning framework. It is a symbolic math library and is also used for machine learning applications. It uses to build and train models by using the high-level API, which makes getting started with TensorFlow and machine learning easy. We have used TensorFlow object detection API to implement object detection and facial recognition in our android application by applying their respective models and label classes.

### **5.1.3 Google Vision API**

Google Vision API is used for text recognition that helps blind to read text without the help of braille documents. Google vision API uses OCR technique to extract text from images and recognize text on real time.

# 5.2 Designing Tools

### **5.2.1 Just in Mind**

We used JUST IN MIND software for make the prototypes of our mobile application and websites, it helped us a lot to make a rough draft and design of our website and application. Through this we made designs of application and website which we were thinking in our mind, we plot our ideas on to this. Making prototypes of them so we decided how it will look alike, is it sufficient or not. Understandable or not. Is the interface being good for user’s perspective or not. Or then we decided what we should do to make our system efficient and effective for users.

### **5.2.2 Photoshop**

We used PHOTOSHOP for making the poster of our project and also for the object and facial expression’s images. We create our project logo through Photoshop and making pamphlet for publicity of our project which defines the purpose of our system and where mention the features and functionalities of our project.

# 5.3 Documentation Tools

### **5.3.1 MS Word**

Microsoft Word allows you to create professional-quality documents. We have used Microsoft word for documentation because of its features including spell check, grammar check, text and font formatting, image support, advanced page layout, and more. It allows us to type words, sentences and paragraphs much like a typewriter. However, Microsoft Word allows to edit words after typing them, i.e., correct spelling, grammar, change words, delete or add words or even move entire blocks of text to other parts of the document.

### **5.3.2 MS Visio**

Microsoft Visio can be used to create simple or complicated diagrams. It offers a wide variety of built-in shapes, objects, and stencils to work with. The graphics that are used in Visio are standard images utilized by flowcharts, decision diagrams, playbooks, and even network diagramming. We have used MS Visio to define the overall flow of our projects using different UML diagrams.

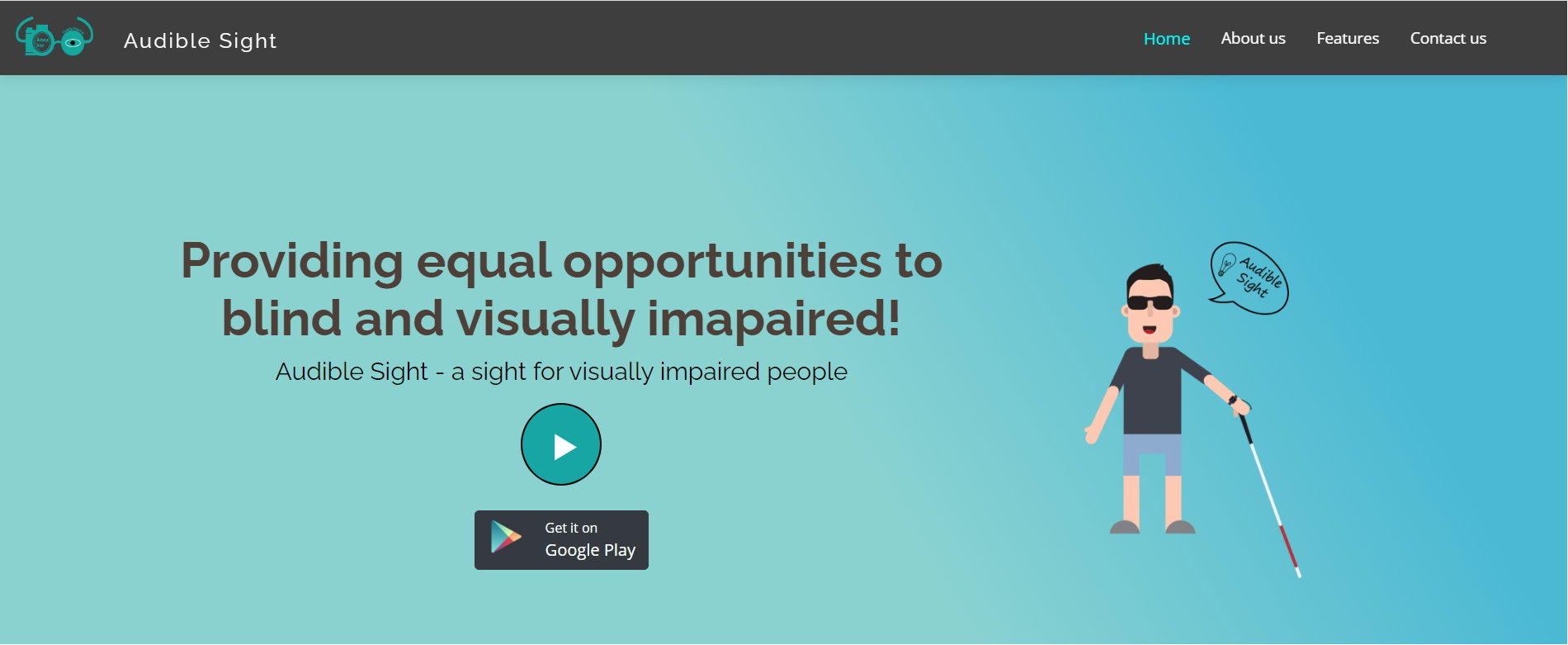
# Chapter 6

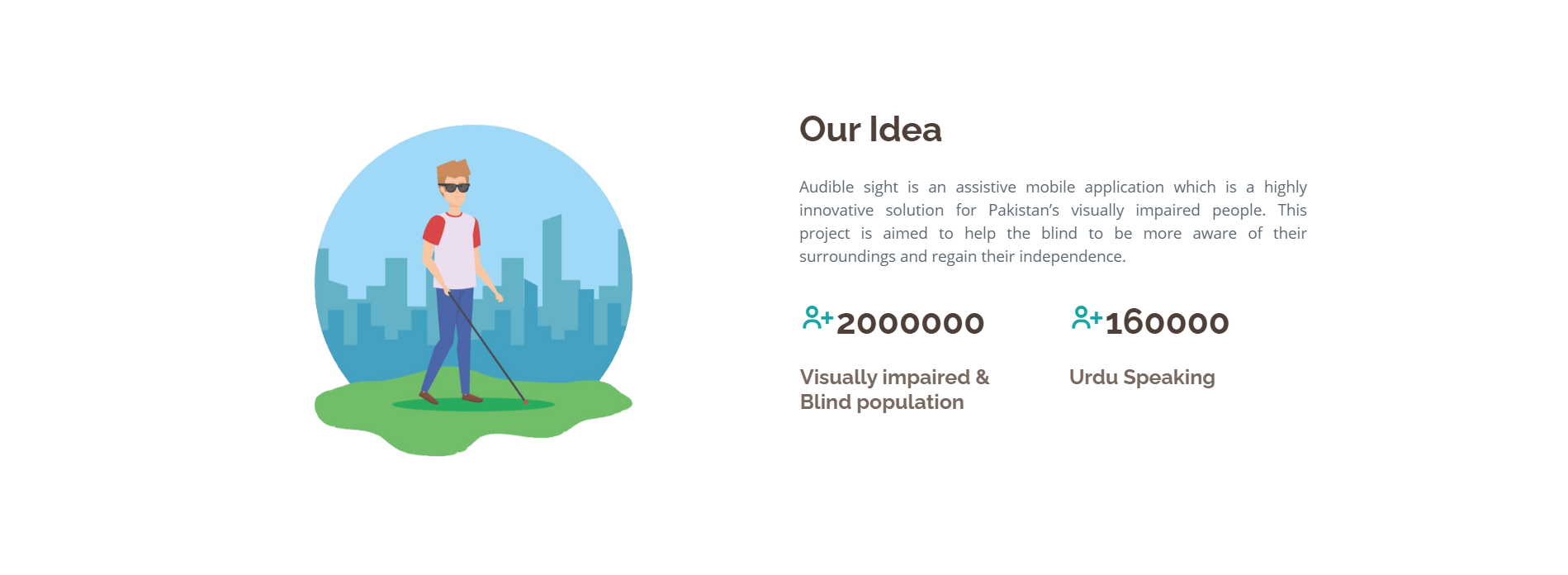
# User Interface Design

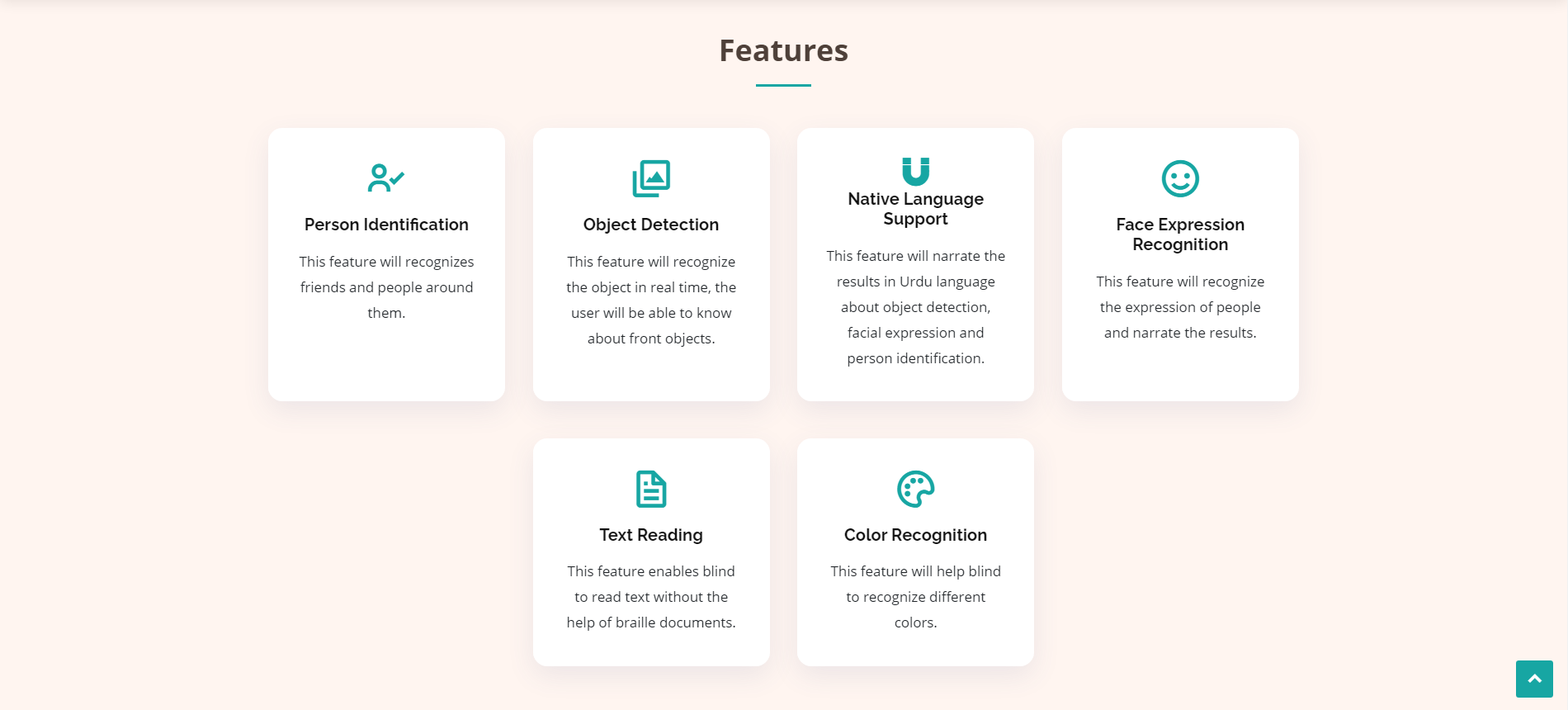
# 6.1 Web User Interfaces

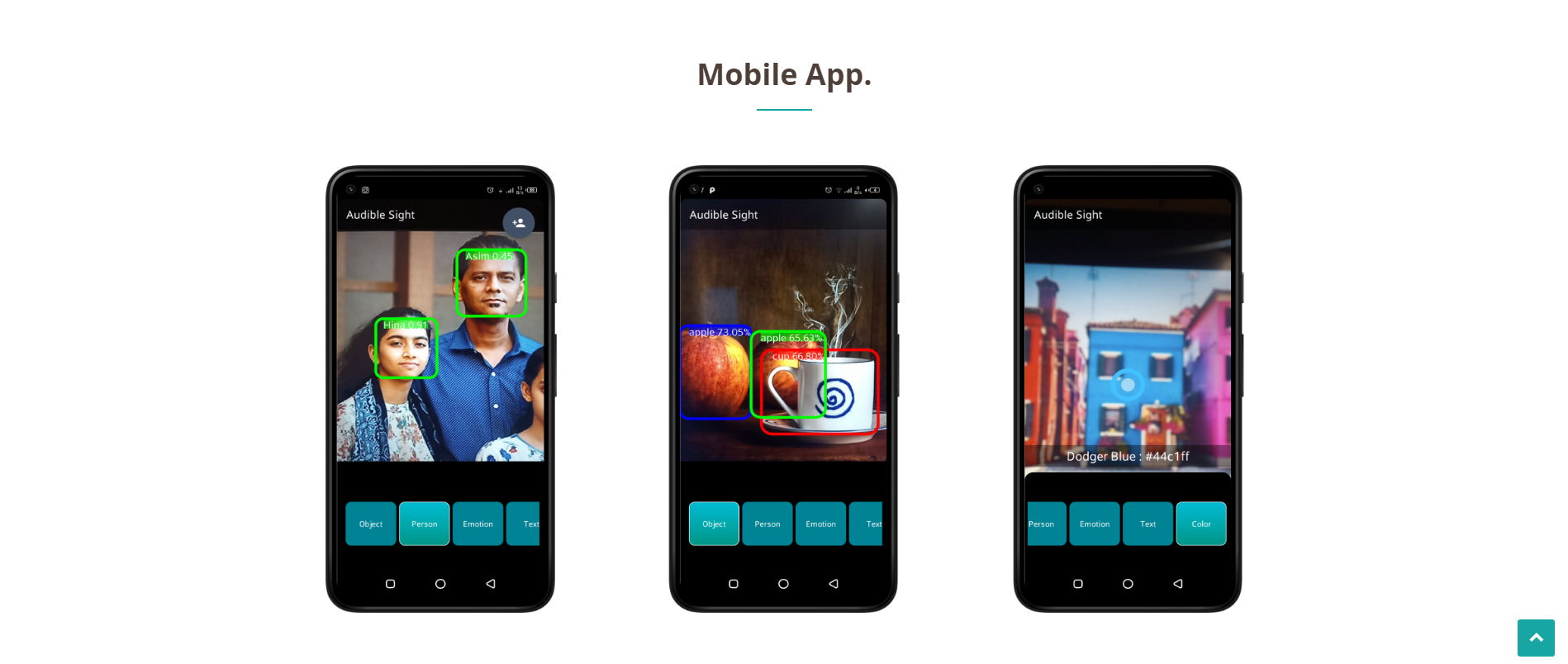
### **6.1.1 H****ome Page**

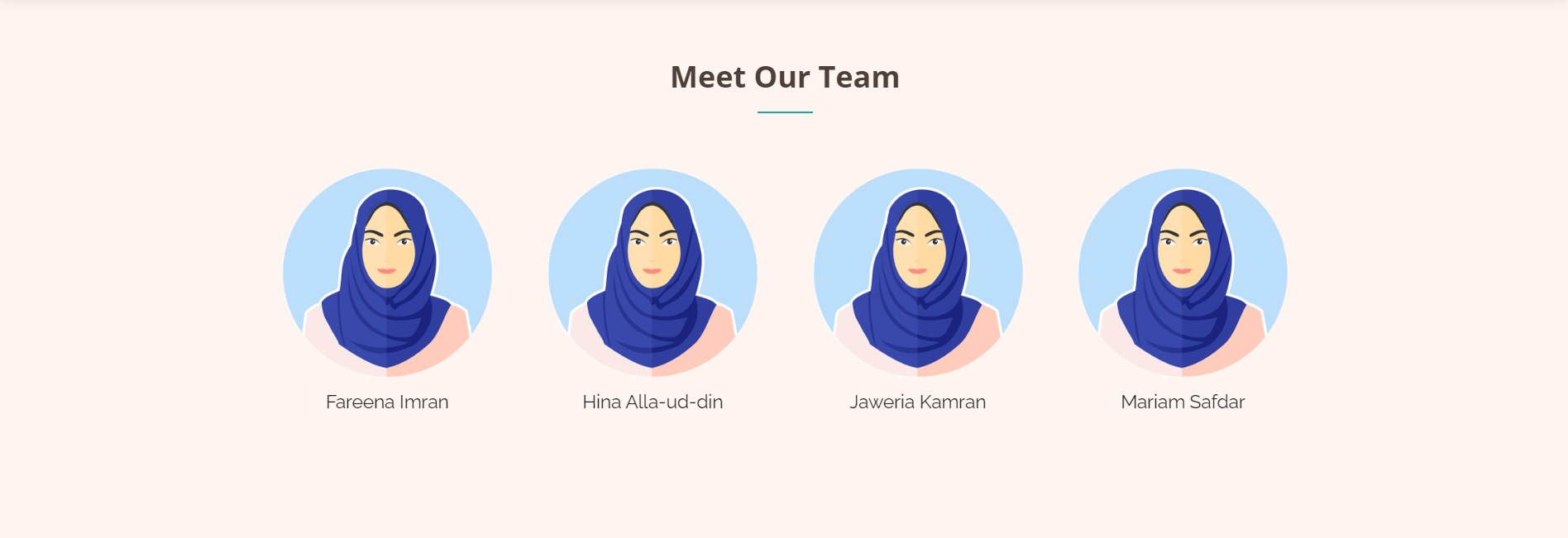
Home page display options to get login or register. There is a short demo video on top that describe the feature of Audible Sight. Also, user can contact by submitting contact form.

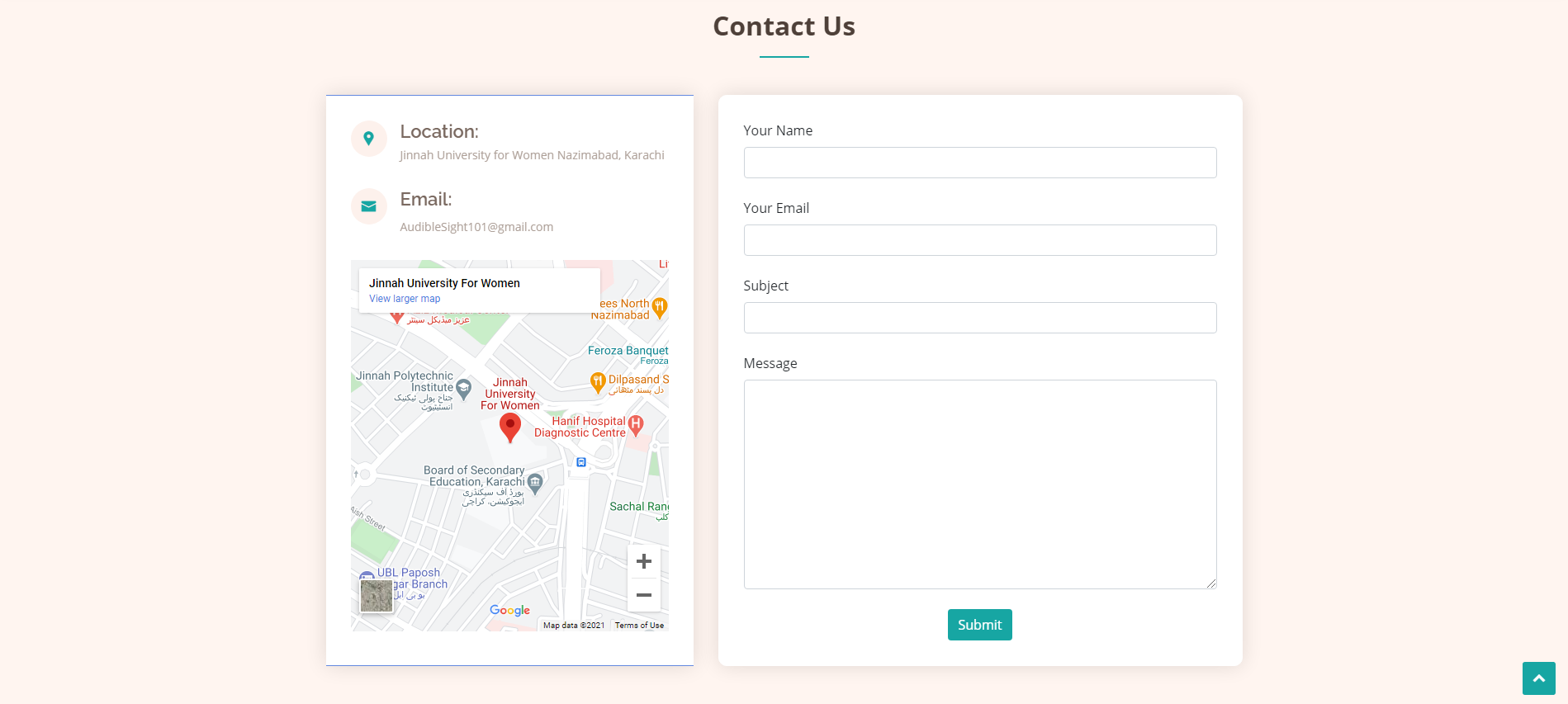


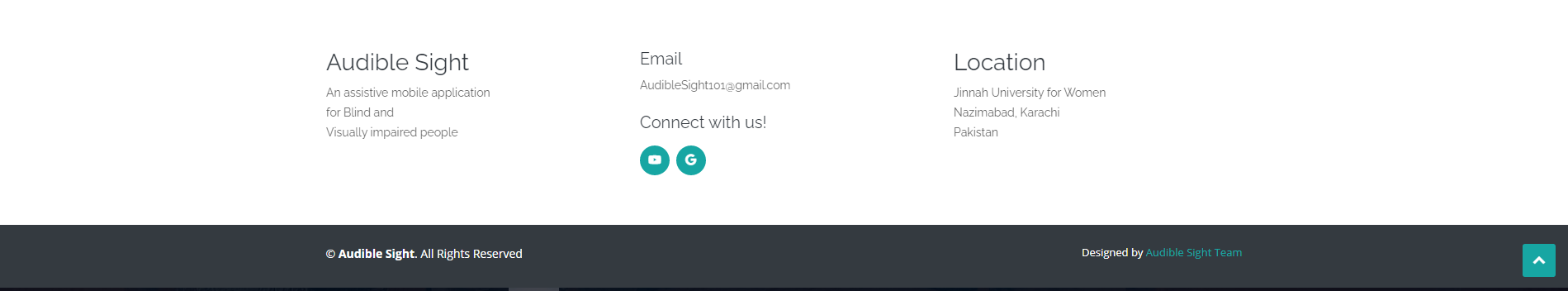












# 6.2 Software Interfaces

Application is a mobile application, will need an Android version 4.0 or higher in order to perform. Windows operating systems will be used during development process. The application will be implemented by using Android SDK. For image processing and conversion of processed data into sound, Text-to-Speech libraries will be used respectively.

# 6.3 Mobile Application Interfaces

The splash screen of Audible Sight is shown in figure 6.1. The first screen opens directly after the splash screen that recognize objects with voice over as shown in figure 6.2.

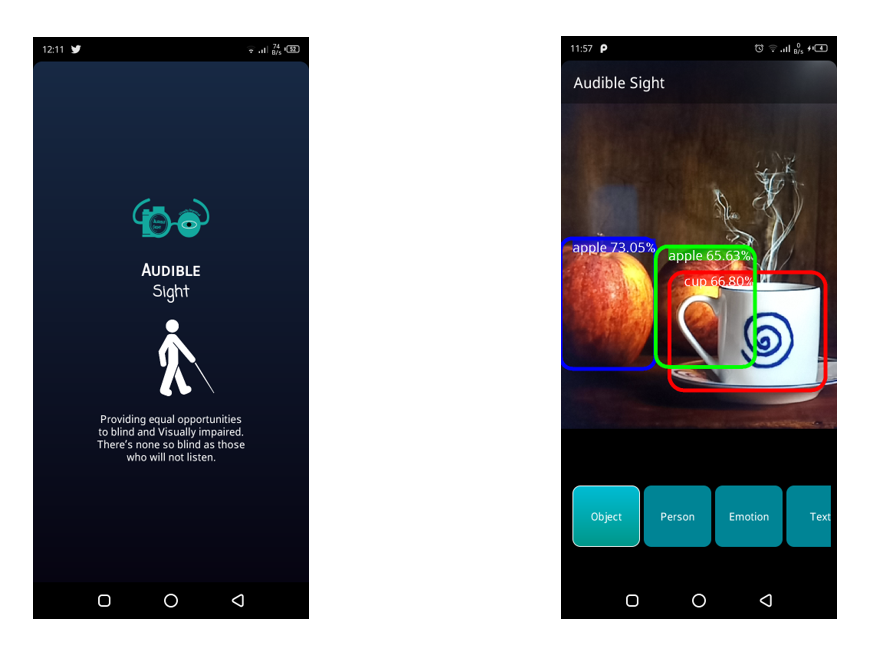
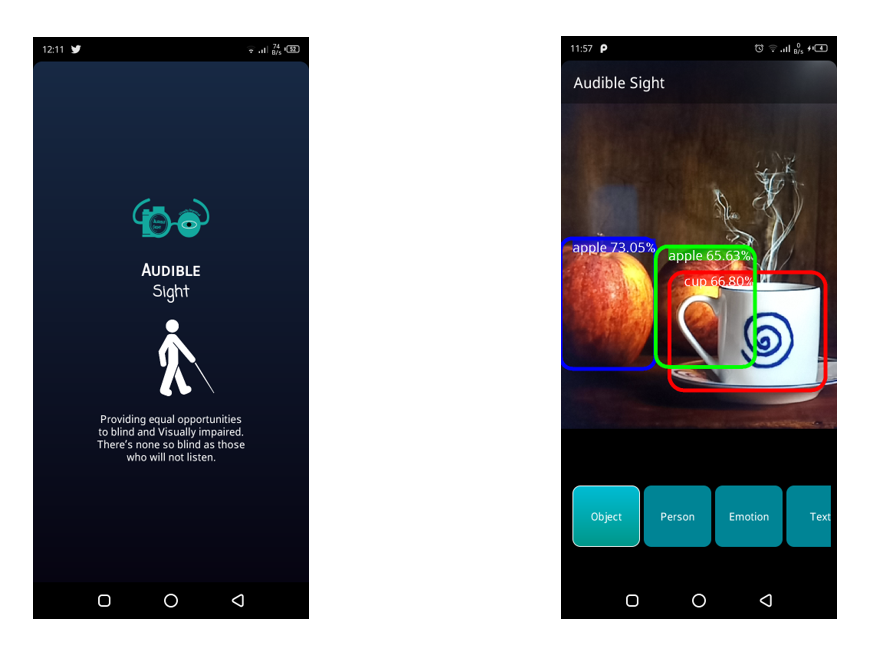


Figure 6.2: Object Detection Screen

Figure 6.1: Splash Screen

Person detection screen allow user to add person name in database to recognize person in future and display a bounding box for detected faces in frame as illustrated in figure 6.3. The add person button allow user to input person’s name through voice in the dialog box as shown in figure 6.4.

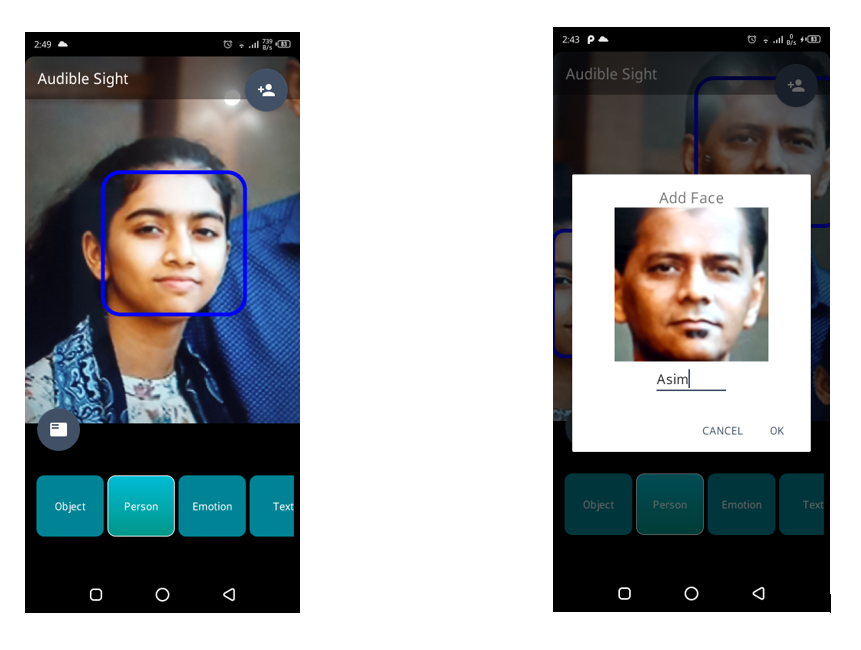
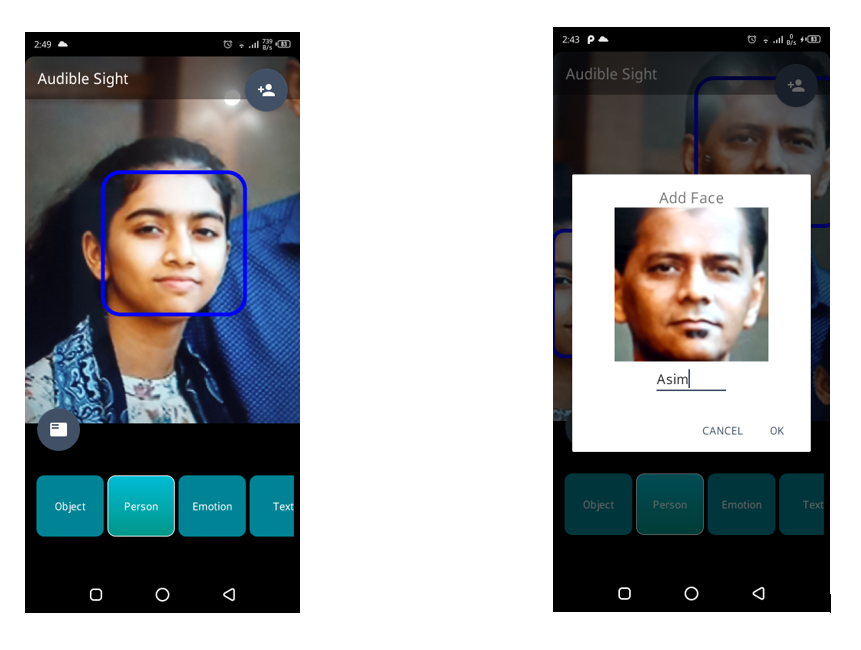


Figure 6.4: Input name Dialog Box

Figure 6.3: Person Detection Screen

Person Recognition screen recognize person with the saved name that is visible in the bounding box which can be seen in figure 6.5. This screen recognizes the color in real time and narrate that color name on tap screen, shown in figure 6.6.

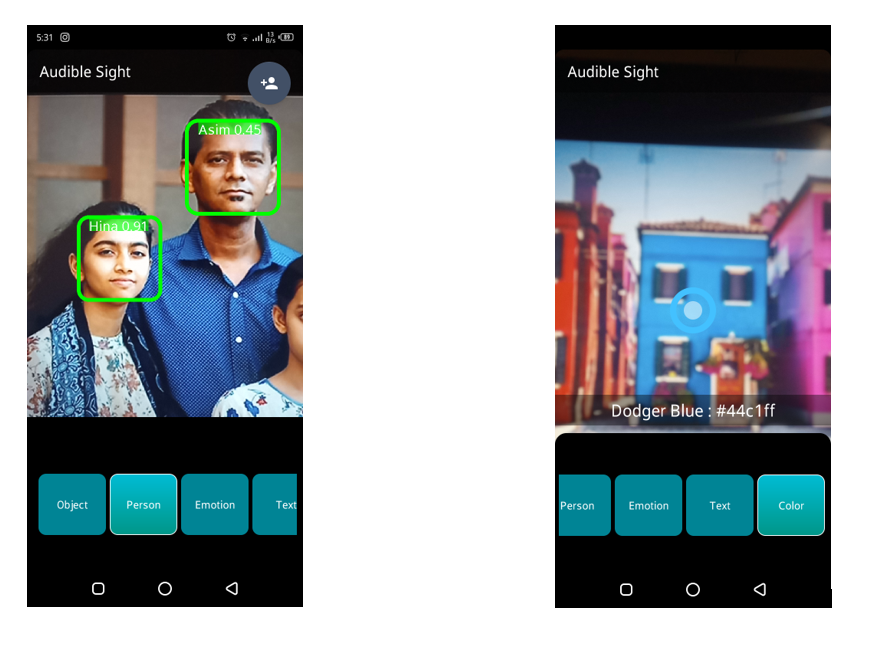
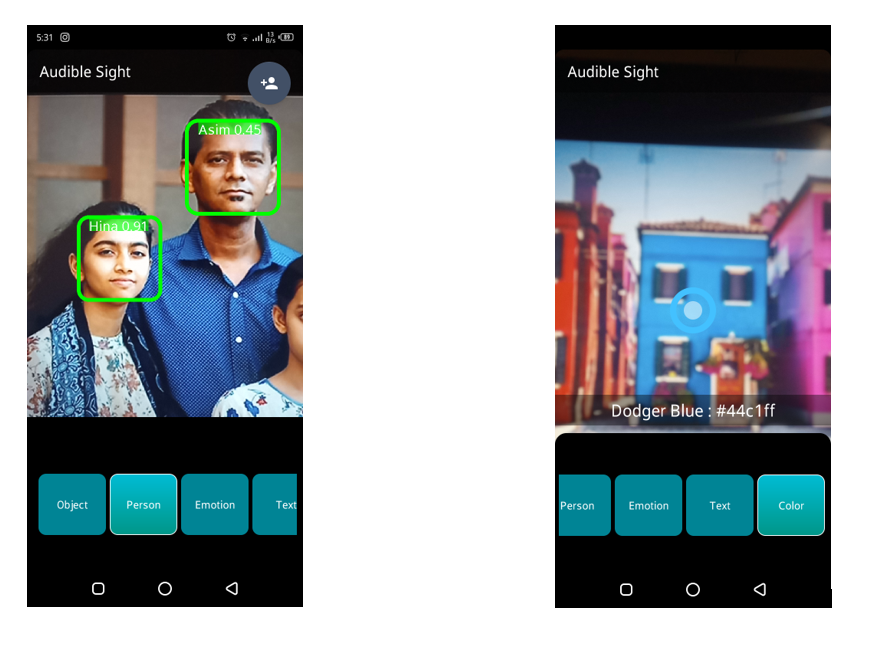


Figure 6.6: Color Recognition Screen

Figure 6.5: Person Recognition Screen

Text Reading screen detects and extract text in real time and read that text as demonstrated in figure 6.7. Emotion Recognition screen recognizes person’s emotions and narrate the detected emotion name, indicated in figure 6.8.

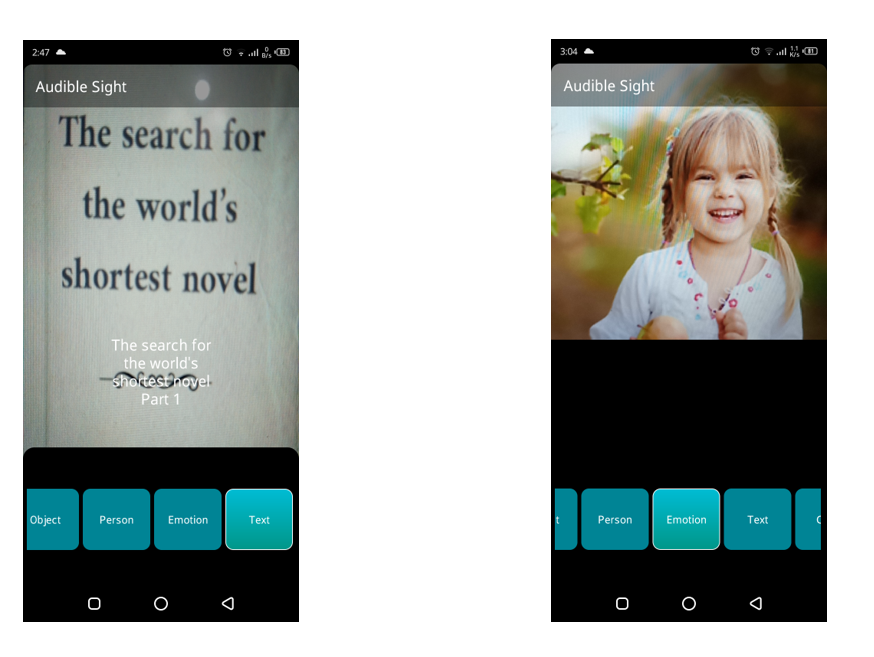
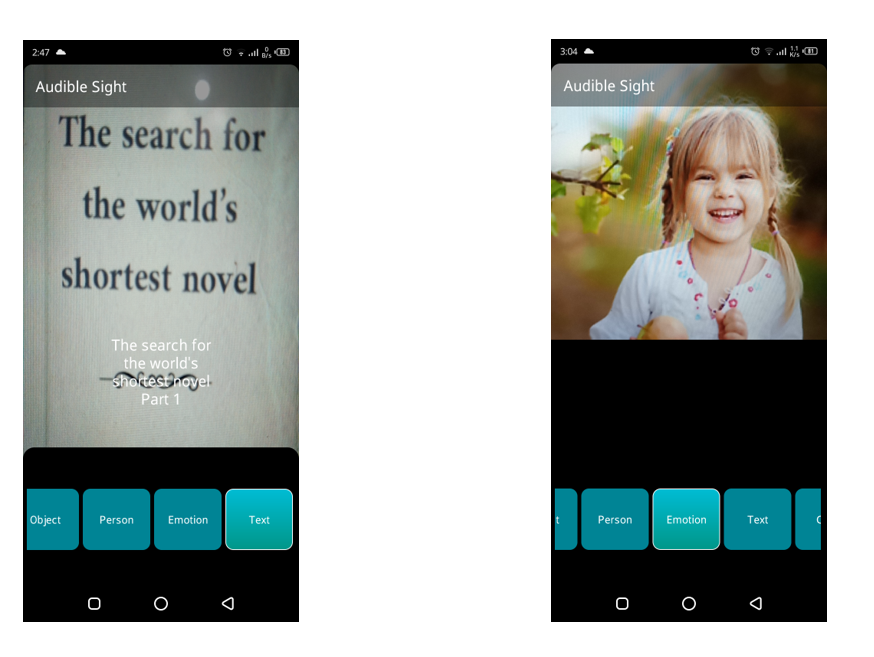


Figure 6.8: Emotion Recognition Screen

Figure 6.7: Text Reading Screen

# Chapter 7

# Methodology

To develop a project, it is feasible to follow a process model because it enables developers to specify, design, implement, test and modify software systems. These process models define different phases of the system and the order in which they are conducted. In Audible Sight project, the Agile process model is used with Scrum framework for managing process.

# 7.1 Project Plan

For project development, Feature Driven Development methodology is used and for management Scrum framework is used. FDD consists of five high level activities:

### **7.1.1 Develop overall domain model**

Mockups of mobile application have been designed to understand the overall flow of our project.

.

### **7.1.2 Build List of Features**

Following is the features list built for the project:

* Recognition of objects
* Recognition of facial expression
* Identification of persons
* Reading of computer-generated text
* Recognition of color
* Support of native language (Urdu)

### **7.1.3 User Stories**

* As a blind user I want to detect object in front of me so that I can go anywhere independently.
* As a blind user, I want to identify known person so that I can know if any-body related to me is there to talk or help me.
* As a blind user I want to know the face expression of a person near to me so that I can know the person's mood.
* As a blind user I want the system to be more efficient enough to tell me about the things in my surrounding so that I can know about the environment in real time.
* As a blind user I want all the information in my national language (Urdu) so that I can easily understand the information.
* As a blind user, I want to read the short text so that I can read the menus and sign boards easily.
* As a blind user, I want to recognize different colors so that I can remember things with their colors.

### **7.1.4 Plan by Feature**

Table 7.1: Project Plan

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PROJECT NAME | **Audible Sight** | | | PROJECT MANAGER | **Mariam Safdar** |
| PROJECT DELIVERABLE | Android Application | | | | |
| SCOPE STATEMENT | This project is for visually imapaired people to guide them in their day-to-day activities. The project is based on mobile application that detect object, recognize facial expressions, identify persons, recognize colors, read text and narrate the results in urdu language. | | | | |
| START DATE | 1st January, 2020 | END DATE | 1st January, 2021 | OVERALL PROGRESS | 90% |

Table 7.2: Sprint Plan

|  |  |
| --- | --- |
| **TASK NAME** | **ASSIGNED TO** |
| **SPRINT 1 - Month 1 - January, 2020** | **Requirement Gathering** |
| Gather information for object detection algorithms and color recognition techniques | Hina Allauddin |
| Gather information for Face recognition algorithms | Jaweria Kamran |
| **SPRINT 2 - Month 2 - February, 2020** | **Requirement Gathering** |
| Gather information for tools and technologies related to project | Hina Alla-ud-din |
| Gather requirements for text read techniques. | Fareena Imran |
| **SPRINT 3 - Month 3 - January, 2020** | **Designing** |
| Home page – website | Hina Allauddin |
| Contact Form – website | Fareena Imran |
| Make website responsive – website | Hina Allauddin |
| Logo design | Fareena Imran |
| Mobile Application Mockups | Hina Allauddin, Fareena Imran, Jaweria Kamran, Mariam Safdar |
| **SPRINT 4 - Month 4 - April, 2020** | **Object Detection implementation** |
| Search object detection for mobile application. | Hina Allauddin, Fareena Imran |
| Implement object detection for mobile application | Hina Allauddin, Jaweria Kamran, Fareena Imran, Mariam Safdar |
| **SPRINT 5 - Month 5 – May ,2020** | **Person Identification implementation** |
| Search algorithm for face recognition for mobile application | Hina Allauddin, Jaweria Kamran, Mariam Safdar |
| Implement code to save captured face in database. | Fareena Imran |
| **SPRINT 6 - Month 6 – June ,2020** | **Translate into Urdu language** |
| Search for Implementation to translate information into Urdu language | Hina Allauddin |
| Implement code to translate information into Urdu language | Hina Allauddin |
| **SPRINT 7 - Month 7 – July, 2020** | **Emotion Recognition implementation** |
| Search for Implementation of emotion recognition | Hina Allauddin |
| Implement emotion recognition for mobile application | Hina Allauddin |
| **SPRINT 8 - Month 8 – August, 2020** | **Color Recognition implementation** |
| Search for Implementation of color recognition | Hina Allauddin |
| Implement color recognition for mobile application | Hina Allauddin |
| **SPRINT 9 - Month 9 – September, 2020** | **Text Reading implementation** |
| Search for Implementation of text reading feature | Fareena Imran |
| Implement text reading for mobile application | Fareena Imran |
| **SPRINT 10 - Month 10 – October, 2020** | **Integration** |
| Integrate object detection module with voice input feature | Hina Allauddin |
| Advance person identification module by saving multiple person images | Hina Allauddin |
| **SPRINT 11 - Month 11 – November, 2020** | **Integration** |
| Integrate object, person and emotion recognition module in Audible Sight application | Hina Allauddin |
| Integrate color recognition and text reading module in Audible Sight application | Hina Allauddin |
| **SPRINT 12 - Month 12 – December, 2020** | **Testing** |
| Unit testing | Hina Allauddin, Fareena Imran, Jaweria Kamran, Mariam Safdar |
| Integration Testing | Jaweria Kamran, Mariam Safdar, Fareena Imran, Hina Allauddin |
| **SPRINT 13 - Month 13 – January, 2021 (1st – 3rd Week)** | **Documentation** |
| Complete documentation of Audible Sight | Jaweria Kamran, Mariam Safdar, Fareena Imran, Hina Allauddin |
| **SPRINT 14 - Month 13 – January, 2021 (4th Week)** | **Deployment** |
| Deploy application on Google Play Store | Hina Allauddin |

# Chapter 8

# Implementation

# 8.1 Database Design

All the data to detect indoor objects is provided in COCO (Common Objects in Context) Dataset. In our project there are 83 class for object detection. For Implementation of object detection, we used pretrained model that is Mobile\_Net\_SSD\_Quantized\_V1 and this model is trained on COCO Data Set. SSD is a framework that is used to realize the multi box detector model and MobileNet is a high-level feature classification and detection. COCO dataset is mainly used for object detection and person key points. This dataset contains 330K images including 80 object categories. All the data is processing using tensor flow API and there is no internet connectivity require for the application to run. The model takes an input image of 300x300 pixels, with 3 channels per pixel (red, blue, and green). The model delivers four arrays, mapped to the 0-4 indices. Arrays 0, 1, and 2 define 10 identified objects, with each object being corresponding to one element in each array.

Google ML Kit is used for Face Detection and MobileFaceNet model is used for Face Recognition which is trained on MS-Celeb-1M (Microsoft Celeb Dataset). This framework uses deep learning algorithm that retains the data as faceprint and maps the facial features of a person. In application, When the faces are detected, bounding box is retrieved for each detected face. So, we can get a proper resolution image to feed the recognition phase. Face cropping is done by translating the portrait bitmap to the face’s origin and scaling to match the DNN input size.

# 8.2 implementation technique

TensorFlow API is mainly used for object detection that is applicable using different models based on convolution neural network. CNN algorithm is an artificial neural network, a kind of deep neural network most popularly used for analyzing images. In CNN, an input image goes through a bunch of layers. Convolutional layer takes a patch from input and apply a set of filters to the given input then this activated data goes into another layer which called “POOLIM” layer. By using certain functions to sum up sub-regions, the pooling layer reduces the size of feature maps. Afterwards, the feature map is converted as vector like a Neural Network in Fully Connected layer. Multiple results are produced to which fully connected layer pick the top 3 or 5 best cases for consideration. Then highest probability is selected and considered as output. This is how an image visualized through CNN. Camera API with JSON is used for the implementation of color recognition feature. The JSON file containing 1600 colors with hexadecimal code. When user tap on screen, the color focused in the pointer will compare with saved colors from JSON file. For Text Reading, the app uses Google Mobile Vision APIs for Optical character recognition (OCR). The TextRecognize object processes images and determines what text appears within them.

# 8.3 Website Design

We have developed our website using HTML, CSS and Bootstrap framework for responsiveness and jQuery for animation. The UI of our website is included in chapter 6.

# 8.4 Mobile Application Design

We have developed android application using Android Studio. For UI we used XML. The application is implemented in JAVA. TensorFlow was used for training the deep network and to detect multiple objects in an uploaded image. The User Interface of our application is included in chapter 6.

# 8.5 Deployment

The application is published on Google Play Store. Users can download the application easily and can use it with all features for free. There are 10+ downloads and 2 versions have been updated. Application will be updated from time to time. For efficient use of application, users have to check for the updates and install them.

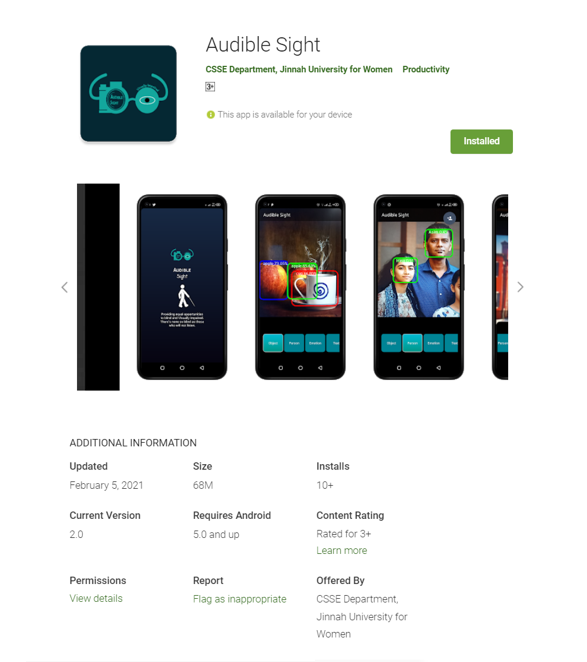


Figure 9.1: App on Play Store

# Chapter 9

# Testing

# 9.1 Test Plan

Test Plan purpose is to describe that all the software testing activities, scope, requirements and implementation are working as per the actual results for the validation of the high quality of the product. Test plans helps:

* To understand the details of software testing to people who are outside of the testing team.
* To guide tester what need to be followed.
* To management team to re-used and review important aspects that are include in test plan documentation like test estimation, test strategy and scope.

# 9.2 Unit Testing

This test is applied on each module to find whether or not each module is properly working. independently unit tested and validated the supporting APIs which are used to provide different functionality to ensure that each feature should work as it is defined.

# 9.3 Integration Testing

Integration testing performs after the unit testing for different modules and components that need to be integrated to perform specific task. In Audible Sight, a type of Incremental testing approach is employed and integration testing is done using white box testing technique in which we tested the flow of specific input through code, expected output, each statement and function on individual basis.

# 9.4 Functional Testing

Functional testing is the testing that is done to check whether the system is performing its functions properly. Moreover, it is also used to check whether the system is performing functions as described in the requirements document.

* Audible Sight functional testing started with object detection feature to check whether the objects are being detected accurately.
* On Facial Expression Recognition to check whether the app is recognizing 7 facial expressions accurately.
* On Person Identification to check whether the app is detecting the person accurately whose image is saved in database.
* On color identification to check that colors are detected accurately.
* On text reading to check whether the app is reading computer generated text correctly.

# 9.5 System Testing

This testing is done to ensure whether the application is working properly on the intended systems. Application was developed on android platform so for testing purposes it was installed on several android mobile phones. The testing results showed that application will work smoothly on android version 8 or above. Moreover, some specifications of Samsung phone are not supporting few modules of this application.

# 9.6 Stress Testing

Stress testing performs for extreme condition that can occur at peak time or at stressful condition on system. It’s necessary to perform for each system because it uncovers or discover some condition that can occur like memory leak, resource loss bugs, data corruption and etc. Stress testing is performed on Audible Sight application to check if it’s able to work correctly under pressure or not. So, to check memory leakage and performance of application, at least 30th names saved or input for person identification.

# 9.7 Performance Testing

Performance testing is performed to check if the application is running smoothly under certain load or action. Its check response time of the system. Audible Sight features are tested for all features as it responds quickly on every click. All the features are working on real time that start working automatically after the camera open by user.

# 9.8 Acceptance Testing

In acceptance testing, application is test with actual user to know if it’s fulfilling its requirements or not. Audible Sight is tested with actual user. Blind resource foundation targeted to test from actual users. Application features is tested by blind users and modify these features according to their requirements.

# 9.8 Test Cases

### **9.8.1 GUI Testing**

This testing has been done to test the Graphical User Interface which comprises three test cases and test case IDs are abbreviated as GUI.

Table 9.1: GUI Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| GUI-01 | Check whether the blind user is able to access the app. | Blind user can easily open/access the app. | Blind user can easily open/access the app. | PASS |
| GUI-02 | Check whether the talkback feature is working properly. | Talkback feature is working smoothly. | Talkback feature is working smoothly. | PASS |
| GUI-03 | Check whether features switching is easily done through talkback. | Feature names are properly called to navigate within the app. | Feature names are properly called to navigate within the app. | PASS |

### **9.8.2 Object Detection Module Testing**

This testing has been done to test the object detection module which comprises four test cases and test case IDs are abbreviated as OD which stand for Object Detection.

Table 9.2: Object Detection Module Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| OD-01 | Check whether this module is accurately detecting people. | Detect persons along with number of persons present there. | Detect persons along with number of persons present there. | PASS |
| OD-02 | Check whether this module is accurately detecting fruits. | Detect fruits along with their names. | Detect fruits along with their names. | PASS |
| OD-03 | Check whether this module is accurately detecting animals. | Detect animals along with their names. | Detect animals along with their names. | PASS |
| OD-04 | Check whether this module is accurately detecting indoor objects. | Detect objects along with their names. | Detect object along with their names. | PASS |

### **9.8.3 Emotion Recognition Module Testing**

This testing has been done to test the Emotion Recognition module which comprises three test cases and test case IDs are abbreviated as ER which stand for Emotion Recognition.

Table 9.3: Emotion Recognition Module Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| ER-01 | Check whether the module is able to recognize the expression. | Feature can easily recognize 7 expressions. | Feature can easily recognize 7 expressions. | PASS |
| ER-02 | Check whether the feature will recognize expression of person wearing any mask or veil. | No expression is recognized as the face is covered. | No expression is recognized as the face is covered. | PASS |
| ER-03 | Check whether features switching is easily done through talkback. | Feature names are properly called to navigate within the app. | Feature names are properly called to navigate within the app. | PASS |

### **9.8.4 Color Recognition Module Testing**

This testing has been done to test the Color Recognition module which comprises three test cases and test case IDs are abbreviated as CR which stand for Color Recognition.

Table 9.4: Color Recognition Module Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| CR-01 | Check whether the module is able to detect different colors. | Feature is recognizing different shades of colors. | Feature is recognizing different shades of colors. | PASS |
| CR-02 | Check whether the feature is recognizing color where screen is focused properly. | Color us recognized where screen is focused by touch. | Color us recognized where screen is focused by touch. | PASS |
| CR-03 | Check whether features is clearly and properly telling colors name. | Feature is clearly telling colors names. | Feature is clearly telling colors names. | PASS |

### **9.8.5 Text Reading Module Testing**

This testing has been done to test the Text Reading module which comprises four test cases and test case IDs are abbreviated as TR which stand for Text Reading.

Table 9.5: Text Reading Module Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| TR-01 | Check whether the module is able to read English text. | Feature will read the English text clearly. | Feature will read the English text clearly. | PASS |
| TR-02 | Check whether the feature is able to read computer generated text. | Feature is reading computer generated text smoothly. | Feature is reading computer generated text smoothly. | PASS |
| TR-03 | Check whether features is easily reading numbers along with words. | Feature is able to properly read numbers along with words in the text. | Feature is able to properly read numbers along with words in the text. | PASS |
| TR-04 | Check whether feature is able to read hand written or Urdu text. | Feature is unable to read hand written or Urdu text. | Feature is unable to read hand written or Urdu text. | PASS |

### **9.8.6 Person Identification Module Testing**

This testing has been done to test the Person Identification module which comprises four test cases and test case IDs are abbreviated as PI which stand for Person Identification.

Table 9.6: Person Identification Module Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| PI-01 | Check whether the blind user is able to save the person's image. | Blind user can easily save the person's image. | Blind user can easily save the person's image. | PASS |
| PI-02 | Check whether the image saving option of feature is easily accessible. | Image saving option is easily accessible. | Image saving option is easily accessible. | PASS |
| PI-03 | Check whether text and voice input accessibility for saving person's name is working properly. | Voice and text input accessibilities are working properly. | Voice and text input accessibilities are working properly. | PASS |
| PI-04 | Check whether using text input numbers are saved in person's name. | Numbers are not allowed to enter to save person's name. | Numbers are not allowed to save person's name. | NOT PASS |

### **9.8.7 Stress Testing**

The stress testing is performed to check application performance under pressure, at least 30 names are saved for person identification and test case ID is abbreviated as ST which stand for Stress Testing.

Table 9.7: Stress Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| ST-01 | Input 30 names for person identification. | It should recognize all saved person without crashing. | It recognizes all the person in front of camera without being crash. | PASS |

### **9.8.8 Voice Testing**

This testing is performed to check if voice input feature is working properly which comprises three test cases. Test case IDs are abbreviated as VT which stand for Voice Testing.

Table 9.8: Voice Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Objective** | **Expected Result** | **Actual Result** | **Test Status** |
| VT-01 | If its easily taking names input through voice. | The name should be entered on first time when user input name by voice. | It enters name on first time. | PASS |
| VT-02 | Voice should be clear enough to understand by the user. | User should be able to understands the objects through voice. | User is able to recognize object through voice. | PASS |
| VT-03 | Check if voice is responding on every object that comes in front of camera. | Voice should pronounce all objects. | Voice is not responding on every objects. | NOT PASS |

# Conclusion

Audible Sight is an android application developed on machine learning techniques. It comprises five features Object Detection, Face Recognition, Emotion Recognition, Color Recognition and Text Reading. We aimed to support blind in a way that they can get along with their surrounding in easiest and convenient way. As we humans want to move independently but most blind people do their basic routine task through any guidance. This hurdle restricts them to depend on others for their living. So, we have provided the feature of object detection in our app by which blind can easily orientate themselves in their indoor environments. Blind people are also unable to recognize about the known person when they are interacting with people. To solve this problem, we have implemented the person identification feature to help them in recognizing their relatives and friends. Socializing is one of the biggest issues faced by blind when meeting with different people because they don’t know how the person is reacting while talking to them. For this purpose, emotion recognition feature is provided in our app to read the facial expressions of people.

We have combined object, emotion and person identification functionalities using TensorFlow API and text reading form Google Vision API and color recognition from CameraAPI with JSON into single Application to achieve our objective. For object detection, we have used mobile\_net\_ssd model trained on COCO dataset that containing 330,000 images. For Face Recognition, we have used MobileFaceNet model trained on MS\_Celeb\_1M face dataset containing 3.8 M images. The results are converted into sound using Text-to-speech library and then sound is transmitted to the user. In future, several advancements can be made in this application with more features that play a vital role in helping blind people such as age, gender and Urdu Text Reading. Also, to help them navigate in the outdoor environment by informing them any important places passing by like hospitals, parks and restaurants.

# References

|  |  |
| --- | --- |
| [1] | Y. Z. Erin Brady, "Visual challenges in the everyday lives of blind people," in *SIGCHI Conference on Human Factors in Computing Systems*, 2013. |
| [2] | J. John, "Facial Expression Recognition System for Visually Impaired," in *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, 2019. |
| [3] | H. A. Hanen Jabnoun, "Object detection and identification for blind people in video scene," in *Conference: 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, 2015. |
| [4] | M. A. S. Yasmin Mussarat, "Face Recognition: A Survey," *JOURNAL OF ENGINEERING SCIENCE AND TECHNOLOGY REVIEW,* vol. 10, no. 2, pp. 166-177, 2017. |
| [5] | G. S. Pawan Mishra, "A Study on Video Surveillance System for Object Detection and Tracking," in *3rd International Conference on “Computing for Sustainable Global Development*, New Delhi (INDIA), 2016. |
| [6] | A. T. Dumitru Erhan, "Deep Neural Networks for Object Detection," in *Advances in Neural Information Processing Systems*, 2013. |
| [7] | P. A. S. E. F. Prof. Hassan Soliman, "Face Recognition in Mobile Devices," *International Journal of Computer Applications,* vol. Volume 73, no. 2, URL: https://www.researchgate.net/publication/271069265\_Face\_Recognition\_in\_Mobile\_Devices, p. (0975 – 8887), 2013. |
| [8] | K. S. V. S. C. S. Aesha Shah, "Built-in Face Recognition for Smart Phone Devices," *International Research Journal of Engineering and Technology (IRJET),* vol. 04, no. 01, Jan -2017. |
| [9] | Y. T. A. A. Chucai Yi, "Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons," *IEEE/ASME Transactions on Mechatronics ,* vol. 19, no. 3, URL: https://ieeexplore.ieee.org/document/6517218, pp. 808 - 817, 2014. |
| [10] | B. M. T. Z. Ruxandra Tapu, "DEEP-SEE: Joint Object Detection, Tracking and Recognition with Application to Visually Impaired Navigational Assistance," *Advanced Research and TEchniques for Multidimensional Imaging Systems Department,* pp. URL: https://www.mdpi.com/1424-8220/17/11/2473/htm, 2017. |