ABSTRACT

With the rise of development around software engineering, humanity hasn't implemented enhancements that we can use for visually impaired people. With the development of artificial intelligence and machine learning techniques, engineers can use to infuse "intelligence" even in dumb computers and with the ease of accessibility, we can extend this "intelligence" to our smartphones to help the visually impaired people understand their surroundings and receive a helping hand during their day-to-day activities.

Our mobile application bridges the gap between visually impaired people and the visual world by leveraging the power of Deep Learning which can be made accessible even on low-end devices with a lucid User-Interface that allows them to better understand the world around them.

Our primary purpose is to leverage and study how Deep Learning architecture and easy prototyping tools can help us develop applications that we can quickly render, even on low-end devices. With this application, we aim to have a one-stop solution to allow blind or partially blind People to understand their surroundings better and cope with the dynamic world ahead of them.

Our mobile application allows the Users to leverage Image Captioning Architecture to generate real-time insight into their surroundings while using Natural Language Processing to speak out lucidly. The cornerstone of our mobile application is its User Interface which would infuse a lucid experience for the User with its ease of handling and use.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION
ML	Machine Learning
IDE	Integrated Development Environment
UI	User Interface
API	Application Programming Interface
NLP	Natural Language Processing
CNN	Convolutional Neural Network
UX	User Experience

CHAPTER 1

INTRODUCTION

Vision is one of the most important human senses and helps us understand the perception of the environment around us. This information is not available through conventional methods for visually impaired people, and they have to depend on external interventions to utilize it. However, it leads to a lot of problems in their safety and mobility, with the visually impaired people having to rely on external factors for everyday activities. The rise of artificial intelligence and mobile application development has fused the user experience to provide the best possible user experience for the general mass. Applications like Tik Tok, Instagram, Amazon, and Spotify use real-time machine learning algorithms to accelerate their data-driven insights and bring the best possible user experience. However, not much research and implementation have been done by researchers and engineers to help visually impaired people understand their visual scenes. Existing work has been mainly focused on helping understand the components and characteristics of the environment, which is not fully scalable. This project aims to propose a solution for visually impaired people to better understand their environment, with specific features implemented to allow better them to navigate and interact with the environment.

1.1 MOTIVATION

At present, the development of assistive technology is still in its infancy. Assistive technology (AT) is helpful for people who are striving to maximize their efficiency and work output but are limited by their disabilities. Assistive technologies can help people with specific disabilities carry out their tasks much better, reduce the social and emotional burden and help them contribute to the professional outlook of their career.

Assistive technology for visually impaired people saw initial traction in the 2010s, developing watches, yardsticks, special glasses, and other devices. However, the real boost came in the mid-2010s when mobile application development boomed. The mobile application development boom was fueled by the rise of the Android and iOS platforms. The capability to embed high-functioning applications in the user's devices was deemed proper to bring this technology to light.

The development of machine learning frameworks focused on mobile applications, like Tensorflow Lite, also helped herald the integration of assistive technology into the mobile application development process. However, the hype died down very slowly even though mobile application development has continued to grow. It has been because of a lack of a firm business plan, lack of a technical roadmap, and lack of specific talent to build a team.

It has introduced a lot of half-baked applications in the market since the development of assistive technology has been a slow and inefficient process. While picking up this problem, we wanted to focus on the visually impaired population specifically. Today, India is home to around 12 million visually impaired people, one-third of the world's total visually impaired population. To prevent the visually impaired population from being left behind, we decided to focus on building a mobile application with modern machine learning and deep learning techniques and frameworks.

1.2 IDE

Visual Studio Code (VSC) is a desktop-based code editor developed by Microsoft. It is used to create and edit code and debug programs. We used VS Code for developing the mobile application alongside an Android 11-based smartphone as an emulator for development, testing, and debugging. Jupyter Notebooks were used for machine-learning modeling and creating the deployment models. The

environment was managed using Anaconda, an open-source distribution of Python and R, to perform data science and machine learning.

1.3 OBJECTIVES

The main aim of this project was to develop a mobile application that would capitalize on machine learning models developed by us to help visually impaired people to navigate and understand their surroundings much better. The most fundamental challenge was integrating our machine learning models while ensuring a smooth user experience and keeping the application size small and flexible. The project provides multiple features like:

- All machine-learning-powered features, except image captioning, work wholly offline and do not require any internet connection.
- All offline features work in real-time and don't need any processing time for models to make predictions.
- Every feature, including image labeling to currency detection, uses text-to-speech to give the application user a more natural voice.
- The application features an intuitive user experience with a minimum number of buttons, each having a large size.

1.4 OVERVIEW OF THE PROJECT

Chapter 1, deals with the introduction to the project where the existing system is being discussed.

Chapter 2, a detailed description of the literature survey of the papers which are referred during the course of the project was summarized.

Chapter 3, gives a brief explanation on the aim and scope of the project. Here the proposed system has been compared with the existing system. The software and

hardware requirements are provided along the system design, architecture overall project.

Chapter 4, the Results and Discussion along with the screenshots of each module has been depicted.

Chapter 5, deals with the summary and conclusion of the project. It also includes the future scope of the project.

CHAPTER 2

LITERATURE SURVEY

In the paper "Intelligent eye: A mobile application for assisting blind people", published by **Milios Awad**, **Tarek Mahmoud**, et al., an Android-based mobile application is presented and showcased. It presents a user-friendly interface for blind people to navigate the environment using a single application. The application features colour detection, banknote recognition, object recognition and light detection. The paper also presents how the application can be more practical for the end-users and provide the desired features.

In the paper "A new Android application for blind and visually impaired people" by **Piotr Kardys**, **Damian Huderek**, et al., a novel approach to using voice commands is proposed to provide visually impaired people with some assistance while using the mobile application. It features some new capabilities, including sending, receiving text messages and using a phone book alongside monitoring their battery or positioning.

In the paper "BLIND GUARD: Designing Android Apps for Visually Impaired Persons", by **Mozibul Hoque**, **Monir Hossain**, et al., new features around the protection & security of visually impaired users are proposed. It showcases how easier it would be for visually impaired people to walk and speak with others while using the application. It also adds a new intelligent stick that can allow the visually impaired to use the application without any assistance and thus have new assistive features, like a location tracker and more. Alongside the new features, it can also help the visually impaired navigate by providing the best land route while travelling to a destination.

In the paper "BlindAid - Android-based Mobile Application Guide for Visually Challenged People" by **Gagani Senarathne**, **Hansi De Silva**, et al., the authors discuss three primary components: face recognition and obstacle identification and critical navigation. These components are proposed to drive a seamless user experience for the users and help them understand the environment better.

The paper "Image Captioning Methods and Metrics" by **Omkar Sargar** and **Shakti Kinger** was used to understand image captioning better and how Computer Vision (CV) and Natural Language Processing (NLP). Image Captioning is a method for extracting meaningful information from images, and it is made possible by identifying objects, actions, and their relationships and generating a description for the image. Captioning has been used across various fields: self-driving cars, assistive technology, medical imaging, etc. The paper determines the approach for image caption generation using the Convolutional Neural Network (CNN) and Generative Adversarial Network (GAN) model in a deep learning framework. We have adopted this methodology to develop an encoder-decoder architecture. Herein CNN is utilized for image vector generation, and LSTM is used to generate a logical sentence using the NLP modelling done previously. The paper also provides a detailed description of the architecture and the results obtained.

The paper "Real-Time Object Detection Android Application using Machine Learning Concepts for Visually Impaired People" by **Anil Thapa, Nilesh Wani,** et al. discussed Machine Learning and Google's TensorFlow Lite object detection API that provides a flexible and fast solution for accessing TensorFlow Lite interpreter and performing inference. We have utilized the proposed machine learning workflow to develop our workflow for building offline-first machine learning models.

CHAPTER 3

SCOPE AND REQUIREMENT ANALYSIS OF PRESENT INVESTIGATION

In this chapter, the proposed system has been compared with the existing system. The software and hardware requirements are provided.

3.1 EXISTING SYSTEM

The applications featured a clunky user interface that wasn't helpful to visually impaired people in the existing system. After extensive user research, we discovered that speech-controlled inferences are usually not so good for people with visual impairments and might lead to misunderstandings, thus putting the user in jeopardy. Additionally, the machine learning models proposed by the researchers were not so efficient for real-time inference, mainly while serving the machine learning models through a RESTful API. The image captioning system proposed by the researchers was not so efficient either and suffered network latency. It leads to delayed responses and thus brings regression problems.

3.2 PROPOSED SYSTEM

The project proposes the development of a new full-fledged mobile application to help visually impaired people with their daily activities. The application features a seamless, easy-to-use & intuitive user interface with offline-first capabilities. It allows the visually impaired people to navigate through the application quickly and perform tasks such as travelling, inferencing their environment, getting in touch for help, identifying text & currency notes and more. The mobile application has been put into real-life production and has been made available to a group of beta testers who have validated the efficacy of the application.