FACULTY OF ENGINEERING AND TECHNOLOGY
DEPARTMENT OF COMPUTER ENGINEERING AND INFORMATICS

FINAL YEAR PROJECT REPORT

A Smart Glasses System for the Visually Impaired People using an Object Detection Model

By

SSEMUJJU RAYMOND
BU/UG/2020/1976
+256 751 709 546/+256762051794
Email: ssemujjuraymond0@gmail.com

SUPERVISOR
Dr. ALUNYU ANDREW EGWAR

A Final Year Project Report Submitted to the Department of Computer Engineering and Informatics in Partial Fulfillment for the Award of the Degree of Bachelors of Science in Computer Engineering of Busitema University

June, 2024

DECLARATION

I SSEMUJJU RAYMOND hereby declare that the information provided in the report is my own original gathered authentic work and has never been submitted before to the Department of Computer Engineering and Informatics of Busitema University main campus and any another institution of high education.

Signature...…………………………… Date………………………………

APPROVAL

This project titled a smart glasses system for the visually impaired using an object detection model has been submitted with the approval of the supervisor.

……………………………………….. ………………………………………

(signature) (date)

Dr. Alunyu Andrew Egwar (Academic Supervisor)

DEDICATION

I dedicate this report to my lovely family, friends and more especially my mentor Mrs. Ocora Mary N Dawn who have always been there for me whenever they can. Thanks a lot for their love, prayers, emotionally, financially and academic support during the due course of my academics. May the Almighty God bless them abundantly.

ACKNOWLEDGEMENT

Great thanks to the Almighty God for always being with me and I also want to thank Him for the gift of life, a loving family and loving friends who have stood by me at all times.

I would like to say thank you to Department of Computer Engineering and Informatics. Also, I appreciate my parents for supporting me and my dreams no matter the cost.

Finally, I thank Dr. Alunyu Andrew Egwar, a lecturer at the department of Computer Engineering and Informatics for supervising my final year project.

I am really grateful for your individual contribution to my success; may the good Lord bless you all.

ABSTRACT

To date there are challenges experienced by visually impaired individuals in perceiving their surroundings resulting in navigation and accessibility problems. Therefore, to address this limitation, this project designs a system using a microcontroller, object detection model, and camera to assist blind people in recognizing their surroundings, including faces, objects, and obstacles. The visually impaired individuals receive output in form of audio from the system. The design method is used to develop, integrate and test the system components. A prototype and results of such tests were compiled, reported and will be submitted for examination to the department of Computer Engineering and Informatics, Busitema University.

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
IEEE	Institute of Electrical and Electronic Engineers
OCR	Optical Character Recognition
TTS	Text-to-Speech
YOLO	You Only Look Once
CNN	Convolutional Neural Network
GPU	Graphics Processing Unit
GSM	Global System for Mobile computing
GPS	Global Positioning System
LED	Light Emitting Diode
USB-TTL	Universal Serial Bus-time to live
PCB	Printed Circuit Board
DAC	Digital to Analog Convertor
TPU	Tensor Processing Unit
FPs	Frames Per second

LIST OF TABLES

LIST OF FIGURES

Table of contents:

CHAPTER ONE: INTRODUCTION

1.1 BACKGROUND

Blindness is a global health problem affecting millions of people worldwide. According to the World Health Organization, there were 39 million blind and 246 million visually impaired people in 2010 [1]. Recent health measures have reduced the number of people blind due to infectious diseases. However, injuries related to aging are increasing. Cataracts are still the leading cause of blindness worldwide, except in industrialized countries [1].

According to the International Agency for the Prevention of Blindness, about 0.19% of Uganda's population is blind (83,000), and 1.10% has moderate to severe vision impairment (475,965) [2]. The leading causes of vision impairment and blindness in Uganda are river blindness from black flies, cataracts, diabetic retinopathy, glaucoma and age-related macular degeneration [3] [4].

Blindness and low vision can have significant impacts on the quality of life, education, employment, and social inclusion of the affected individuals. As individuals lose their ability to work, participate in education, and contribute to the economy, the overall impact reverberates across communities and the nation.

Many technologies such as white canes, guide dogs, Braille alphabet and audio books have been developed to assist the blind and visually impaired. However, these methods have some limitations, such as requiring special training or being expensive and difficult to obtain. Therefore, more advanced and cheaper technologies are needed to help blind and visually impaired people see and interact with their environment.

One of the new technologies that can meet this need is smart glasses. Smart glasses can be used for many purposes such as education, guidance and entertainment [5]. For blind and low vision people, smart glasses can provide important information about their environment, such as the name, location and distance of objects, as well as exposure effects, words or faces. This will help them navigate their environment safely and freely and access information they would not normally have access to [5].

Several research projects and commercial products have explored the use of smart glasses for blind and low vision people. For example, the Microsoft Seeing AI is an app that runs on smart phones and can describe scenes, objects, text, and people using the phone's camera and voice feedback. The eSight is a headset that uses high-definition cameras and screens to enhance the residual vision of low vision people [1].

1.2 PROBLEM STATEMENT

The problem that this project addresses is the inefficient accessibility and navigation for the visually impaired people that results from their inability to perceive their surrounding environment.

This is a problem because blind people in Uganda face many challenges and risks in their daily lives due to the lack of adequate infrastructure, facilities, and services for them. For instance, they may encounter uneven roads, potholes, open drains, traffic, animals, or crowds, which can cause accidents, injuries, or loss of orientation. Moreover, they may have difficulty accessing information that is presented visually, such as signs, labels, or menus, which can limit their opportunities and choices.

Therefore, there is a need to design and implement a low-cost, user-friendly, and robust technology that can help blind and visually impaired people in Uganda to navigate their surroundings, identify people and text, and enhance their independence and well-being.

1.3 OBJECTIVES OF THE PROJECT

The main objective of this project is to design and implement a prototype of smart glasses for the visually impaired people in Uganda that can detect and identify obstacles in their environment using an object detection model.

The specific objectives of this project are:

i. To conduct a literature review on the existing literature and technologies on smart glasses and object detection models for the visually impaired people, and identify the gaps and opportunities for improvement.
ii. To develop a user interface for the smart glasses that is suitable and accessible for the visually impaired through the design and assembly of the hardware components.
iii. To design and develop the smart glasses system that can capture images from the camera, process them with the object detection algorithm and generate audio output.
iv. To test and validate the smart glasses system and the object detection algorithm.

1.4 SIGNIFICANCE

Once implemented, this project will contribute to the improvement of the quality of life, education, employment, and social inclusion of blind people in Uganda, as well as to the advancement of the knowledge and innovation in the field of assistive technologies for blind and low vision people. It will provide blind people in Uganda with a technology that can help them to perceive and interact with their environment more safely and independently, by detecting and avoiding obstacles and providing them with useful information. It will enhance the accessibility and affordability of smart glasses for blind and low vision people, by designing and implementing a device that is low-cost, lightweight, and offline.

1.5 SCOPE

The project focuses on designing and implementing smart glasses for blind people in Uganda, who have no or very little residual vision.

The project does not cover other types of visual impairments, such as color blindness, or night blindness, or other types of assistive technologies, such as white canes, guide dogs, or braille.

The smart glasses were tested and evaluated in a controlled environment, such as the computer laboratory and a closed room, and in an uncontrolled environment, such as the compound and an open street.

The smart glasses are based on low-cost and lightweight hardware, such as ESP32-CAM, Arduino Nano, and an MP3 Voice Playback Module as the main components.

CHAPTER TWO: LITERATURE REVIEW

This chapter presents a literature review on the existing smart glasses and object detection models for blind and low vision people. The chapter is divided into four parts. In the first part, key terms and concepts that are used in the project, such as smart glasses, object detection, and blindness are explained. The second part examines the related works and projects that have been developed for blind and low vision people. In the third part, existing systems and products that are available in the market, are examined and compared in terms of functionality, usability, and affordability. The fourth part reviews the proposed system and describes the main features, components, and functions of the system, and how they work together to achieve the objectives.

2.1 Key Terms and Concepts

The following were the key terms and concepts related to the smart glasses for the blind and visually impaired people project based on the research carried out.

2.1.1 Smart Glasses

Smart Glasses are wearable devices that can augment the user's vision with computer-generated information, such as images, text, or sound. Smart glasses typically consist of a camera, a display, a speaker, a microphone, and a processor, and can communicate with other devices, such as smartphones or computers, via wireless connections [6].

2.1.2 Object Detection

Object detection is a computer vision task that aims to locate and identify objects in an image or a video. Object detection models use machine learning algorithms to learn the features and patterns of different objects, such as cars, bikes, pedestrians, etc., and to output the class and the bounding box of each object in the input image or video [7].

Computer vision and artificial intelligence methods play a key role in the development of smart glass systems. It is not possible to build a smart glass system without computer vision methods such as object detection and recognition methods because the input data is an image or a video. Object detection and recognition has garnered the attention of researchers, and numerous new approaches are being developed every year [7].

2.1.3 Blindness

Blindness is a condition of complete or nearly complete vision loss. Blindness can be caused by various factors, such as diseases, injuries, or genetic disorders. Blindness can be classified into two types: congenital blindness, which is present at birth or develops shortly after, and acquired blindness, which occurs later in life. Blindness can also be categorized by the degree of vision loss, such as total blindness, which means no light perception, or functional blindness, which means some light perception but no usable vision [8].

Vision impairment or blindness may be caused by several reasons, such as, cataract (94 million), unaddressed refractive error (88.4 million), glaucoma (7.7 million), corneal opacities (4.2 million), diabetic retinopathy (3.9 million), trachoma (2 million), and others [2].

In Uganda, the prevalence of blindness is estimated to be 1.2%, with the main causes being cataract, glaucoma, and trachoma. According to the International Agency for the Prevention of Blindness, about 0.19% of Uganda's population is blind (83,000), and 1.10% has moderate to severe vision impairment (475,965) [9].

2.1.4 Computer Vision

Computer vision is a branch of artificial intelligence that leverages data from films, photographs, and other visual inputs to assist systems in making judgments based on that data. In simple words, computer vision helps systems to understand what they see.

Human vision is similar to computer vision; however, humans have an advantage. Human vision benefits from lifetimes of context to teach it how to distinguish objects apart, how far away they are and if they are moving [10].

In order to execute computer vision, a system must be taught. It must employ cameras, data, and algorithms, unlike the human sight and brain. Hundreds of processes can be handled in minutes in some instances due to the quality of the model learned and its system. In terms of speed, this can almost match human capabilities, but nothing near its efficiency [10].

2.2 Related Works and Projects

The following were the different projects and works related to the smart glasses for the blind and visually impaired people project based on the research carried out.

2.2.1 Smart Walking Cane

This is also called the Autonomous Walking Stick for the Blind Using Echolocation and Image Processing. The aim of this project is to develop and build a smart walking stick that will aid visually impaired persons in identifying obstacles and getting to their destination [11].

The Assistor was the name given to this construction. The Assistor relies on echolocation, image processing, and a navigation mechanism to function. The Assistor detects things using ultrasonic sensors that echo sound waves. The model has a Smartphone app which is used to lead the user to their destination using GPS (Global Positioning System) and maps, and an image sensor is used to identify the items in front of the user and for navigation by capturing runtime photographs [11].

This project makes use of an ultrasonic sensor. It's complemented with a GPS navigation system. The navigation system has been pre-programmed to assist the user in getting to their preferred place. For each circumstance, the user receives auditory feedback. Image matching is also included in the system. Using algorithms, determine the best route to the destination, to assist the user in getting to their goal [11].

A disadvantage to this is that picture processing is not real-time. The photos taken must be processed once they have been captured. This necessitates a significant amount of computational power and time. Before the picture has finished processing, the risk has happened. The model compensates for this flaw by including an ultrasonic sensor. It's possible that the solution won't work in every instance [11].

2.2.2 Low-Cost Ultrasonic Smart Glasses for Blind

This project constructs a smart glass that uses the ultrasonic sensor for detection. This project focuses on navigating, which is one of the most crucial attributes for the blind. This method is very inexpensive, saves time in construction, and is light and portable. The project's components are easily available, which speeds up building [12].

The initiative, as excellent as it appears to be and as simple as it appears, is not without flaws. The ultrasonic sensor utilized in this approach has a number of flaws. This sensor is unable to effectively portray and anticipate the user's environment. It takes only one step ahead [12].

The sensor has a restricted range to go along with this. It is impossible to properly analyze the surroundings in order to provide an appropriate user route. Furthermore, only one ultrasonic sensor was employed in the center of the glasses in this project. As a result, only forward movement may be detected [12].

2.2.3 Tap-tap see application

Tap-tap See is a mobile camera application designed specifically to help the blind and visually impaired iOS users identify objects they encounter in their everyday lives. It is powered by Cloudsight.ai image recognition API. It utilizes the device's camera and Voice Over functions to photograph objects and identify them out loud for the user [13]. The user simply taps anywhere on the screen to take a picture.

TapTapSee can photography any two or three-dimensional object at any angle and speak the identity back to the user. Its features include picture recognition, ability to repeat the last identification, barcode and QR code reader and auto-focus notification [13].

2.2.4 Smart Shoes

The shoes sync with a user's phone over Bluetooth to an app that pulls one's path from Google Maps which allows the shoes to keep track of where they're going. Once destination and route are set, the shoe buzzes to tell you which direction to turn. A buzzing right shoes means to hang right and a buzzing left shoe means to hung the other way [14].

2.3 Existing Systems Comparison Table

The table below shows a summary of the existing systems related to the smart glasses project for the visually impaired using an object detection model. The table shows a comparison of their strengths and their weaknesses.

PROJECT	STRENGTHS	WEAKNESSES
Smart Walking Cane	➢ Includes face recognition. ➢ Includes panic button. ➢ Easy to use.	➢ Sees through limited angles. ➢ Processing isn't in real time. ➢ Doesn't recognize objects.
Low-Cost Ultrasonic Smart Glasses for Blind	➢ Cheap. ➢ Light and portable. ➢ Components easily available	➢ Limited range. ➢ Only forward movement is detected
Tap-tap see application	➢ Easy to use. ➢ Can detect any number of objects in an image.	➢ High processing time. ➢ Works on captured images only not real time. ➢ Requires the capture of many images to detect all objects.
Smart Shoes	➢ Light and portable. ➢ Can be customized as a fitness buff.	➢ Processing must be done on mobile phone. ➢ Expensive. ➢ It doesn't inform the user about obstacles.

Table 1Table showing the comparison of the existing systems.

2.4 Developed System

The developed system is a prototype of smart glasses for blind people in Uganda that can detect and help the visually impaired people avoid obstacles in their environment using an object detection model. The system consists of three main components: the glasses, the microcontroller, and the headphones. The system works as follows:

- The glasses are equipped with an ESP32-CAM board. The ESP32-CAM captures the video scene in front of the user and runs the object detection model. The board runs a squeeze and attention custom object detection algorithm that can recognize and locate common objects and obstacles in the environment, such as people, tables, computers, desktops, laptops, chairs, cups, stairs, walls, doors, among others. It then sends signals to the microcontroller.

- The microcontroller (Arduino Nano) is used for controlling and managing the system functions. It receives signals from the ESP32-CAM board and runs the corresponding function to trigger the audio files in the MP3 Player Voice Playback Module.

- The MP3 Player Voice Playback Module stores the different audio files for the object classes and it is used to provide the user with audio feedback from the system through the headphones. The audio feedback consists of the object information and the navigation guide. The object information tells the user the name, and the location of the objects in the scene, using natural language and spatial audio. The navigation guide alerts the user of the direction options to take in order to avoid the obstacles.

CHAPTER THREE: METHODOLOGY

This chapter describes the methodology that was used to design, implement, and evaluate the smart glasses and the object detection model for blind people in Uganda. The chapter is divided into four parts. The first part describes the data collection process, which involves capturing video data from different outdoor environments in Uganda. The second part describes the data analysis process, which involves using various techniques to enhance the low-light images, augment and annotate the data. The third part describes the model development and training process. Part four describes the requirements analysis, which involves hardware and software requirements.

3.1 Data Collection

The data collection method involved capturing image and video data from different outdoor environments in Uganda. The data was used to train and test the object detection model, as well as to simulate the scenarios that the blind users may encounter in their daily lives.

The data collection method consisted of the following steps:

- Selecting the locations and the scenarios for the data collection. The locations and the scenarios that were representative of the common and challenging situations that the blind users face in Uganda, such as streets, compounds, markets, closed rooms, among others. The locations and the scenarios also covered a variety of objects and obstacles that the object detection model could be able to recognize, such as chairs, cups, people, computers, poles, walls, doors, among others.
- Capturing the video data. A camera (specifically a phone camera) was used by a sighted person who could act as a proxy for the blind user. The person moved around the selected locations and scenarios, and recorded the video of the scene in front of them. The video was captured at a resolution of 640x480 pixels and a frame rate of 30 frames per second.
- Extracting Key frames. The video data was converted into key frames, which were the images that represent the most important moments of the video. The key frames were selected at a regular interval, such as every five seconds. These settings were achieved

by the help of VLC media player through the scene filter preference settings. The scene video filter is then checked in the all-filter settings. A total of 1008 key frames were obtained from the video scenes.

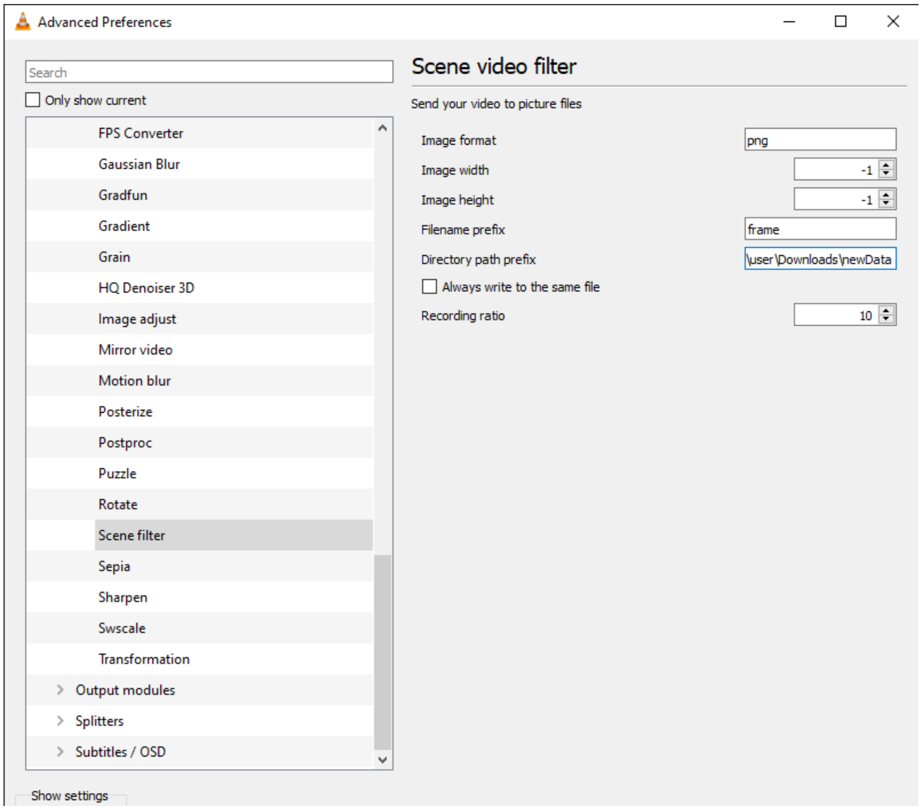


Figure 1 Scene Video Filter settings in VLC media player

3.2 Data Analysis

The data analysis method involved enhancing the low-light images, augmenting the data using various techniques and annotating the data. The data analysis method improves the quality and the quantity of the data, and make it more suitable and diverse for training and testing the object detection model.

The data analysis method consists of the following steps:

- Enhancing the low-light images using a deep learning model. The low-light images are the images that are captured in dark or dim environments, such as at night or indoors. The low-light images can have poor visibility, contrast, and color, which can affect the performance of the object detection model. To enhance the low-light images, a deep learning model, such as the Retinex-Ne was used. This model learns to adjust the brightness, contrast, and color of the low-light images, and produce more clear and realistic images [15].

Figure 2 Enhancing low-light images with Retinex

- Augmenting the data using various techniques. The data augmentation is the process of applying transformations to the original data, such as rotation, flipping, cropping, scaling, noise, blur, among others, to create new and different data. The data augmentation increases the size and the diversity of the data, and makes it more robust and generalizable for the object detection model. The data augmentation was done using tools, such as ImageAugmenter, [16] which applies various transformations to the images and the labels. This entire process yielded 1,932 images in total that were to be used in training the custom model.

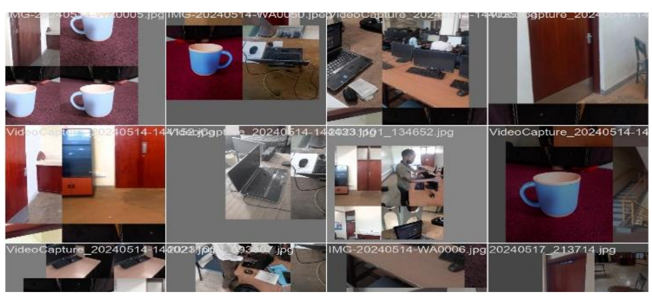


Figure 3 Image Augmentation using ImageAugmenter

- Data partitioning and formatting: The obtained dataset was divided into three parts; training, validation and testing. The data was also shuffled to avoid any natural bias. The model has a fixed 244 x 244 input layer, so there was need to format any input image before feeding it to the model (to train or to predict).

- Data annotation. The key frames were then labeled using a tool, such as LabelImg, which helps to draw bounding boxes and assign class labels to the objects and obstacles in the images. The labels follow the MS COCO dataset format, which is a standard format for object detection datasets. The labels were stored in an XML file that could be converted to TXT files, which contains the information about the image name, the image size, the bounding box coordinates, and the class name of each object and obstacle [17].

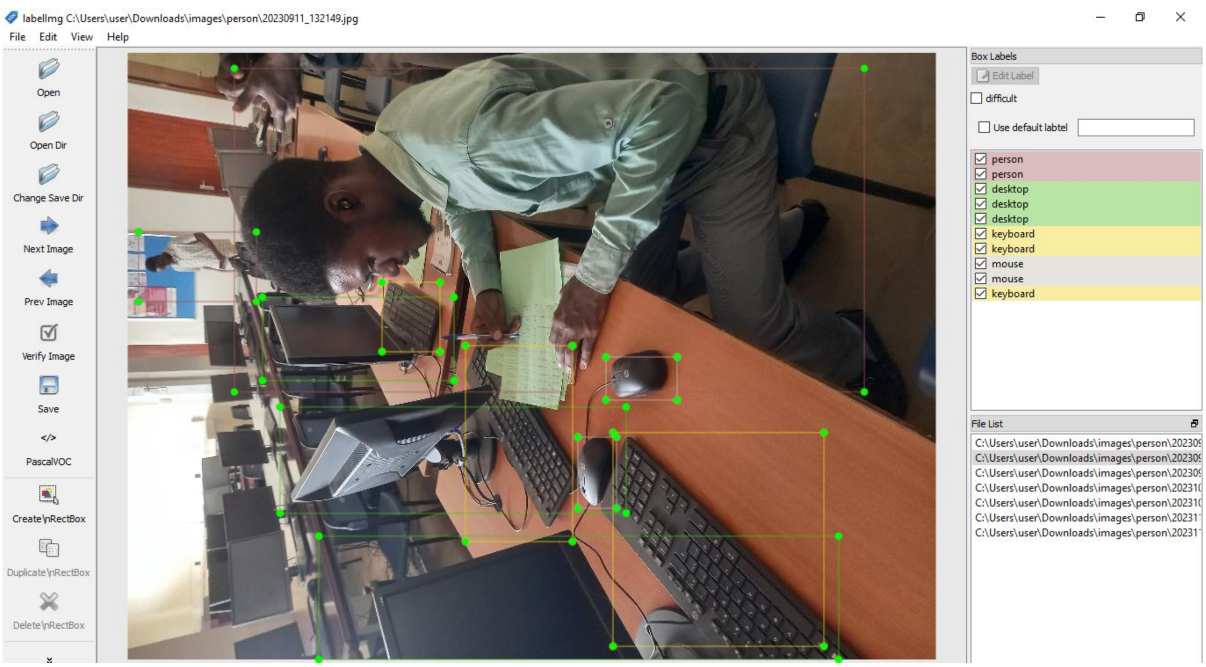


Figure 4 Data Annotation using LabelImg

3.3 Model development and Training

For development, a Deep Learning Object Detector was designed from scratch using TensorFlow. Instead of using a predefined model, each layer in the convolutional neural network which was then trained to detect both the object bound box and its class.

The object detection model developed has a classifier and a regressor for the classification of the object label and the regression of the bound-box coordinates. Since the model must implement two tasks (classification and regression), two different Loss Functions were needed: One for the classification task such as Categorical Cross entropy and one for the bound box regression such as Mean Squared Error. Hyperparameters like training and validation sets, epochs, learning rate, optimizers and so on were then set before training could be done.

Due to the requirement for a GPU or TPU when training, training of the model was done on Google Colab, a cloud-based development environment that provides access to powerful computing resources online and free of charge.

After training, the training performance, and the model evaluation were obtained. The obtained model was then exported as an Arduino library that could be integrated with the sketch code.

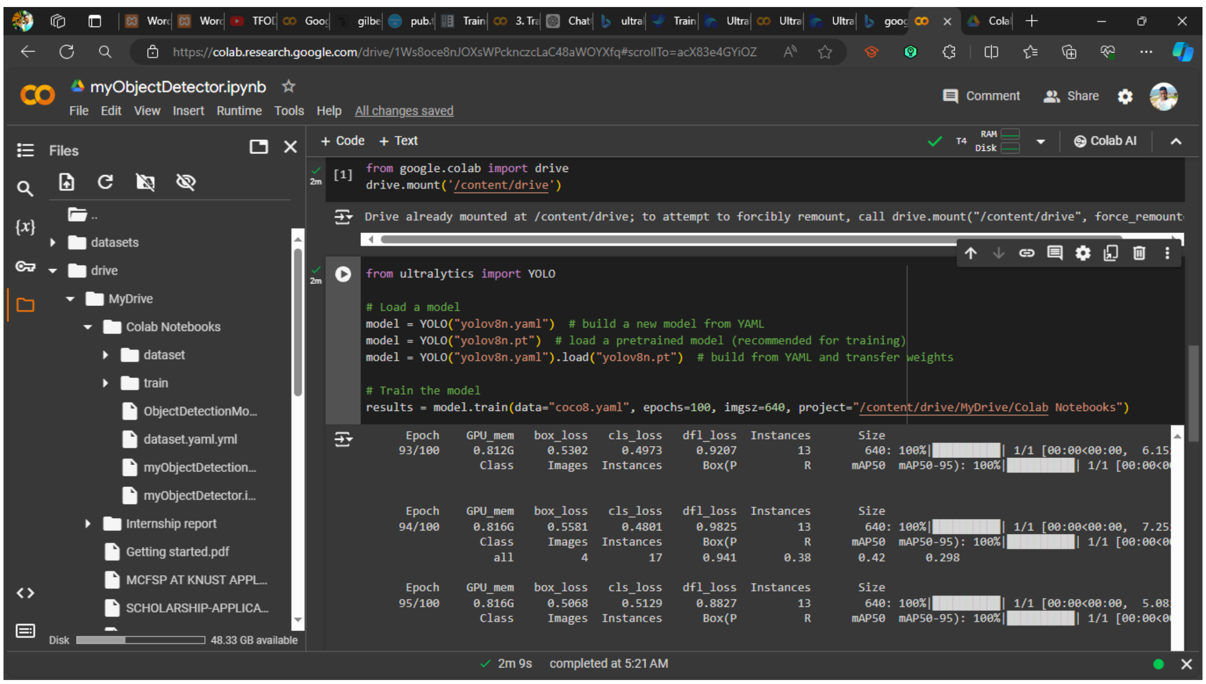


Figure 5 Training a custom model on Colab

3.4 Requirements Analysis

The requirements for designing the system were divided into two which include the following:

3.4.1 Hardware Requirements

The components and their design specifications include:

- **Microcontroller: (ESP32-CAM)**

The ESP32-CAM is a camera module with the ESP3-S chip, and a microSD card slot to store images taken with the camera. It is widely used in various IoT applications, suitable for home smart devices, wireless control, positioning, monitoring among others. Some of its design specifications are included in the table below.

Figure 6 ESP32-CAM board

Table 2 Design Specifications for ESP32-CAM board

PARAMETER	SPECIFICATION
Power Consumption	180mA@5V or 310MmA@5V
Power Supply Range	4.75V – 5.25V
Operating Temperature	$-20°C \sim 70°C$
Image Output Format	JPEG, BMP, GRAYSCALE

- **Microcontroller: (Arduino Nano)**

The Arduino Nano is a small complete board based on the ATmega328.it works with a Mini-B USB cable. Some of its design specifications are included in the table below.

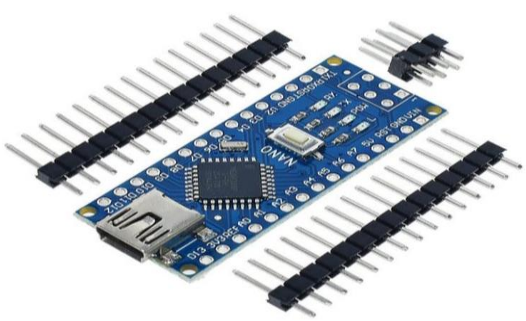


Figure 7 Arduino Nano board

Table 3 Design Specifications for Arduino Nano board

PARAMETER	SPECIFICATION
Operating Voltage	+5V
Input Voltage	+7V ~ +12V
Output Voltage	+5V, +3.3V
Rated Current	40mA/pin

- **MP3 Player Voice Playback Module**

The DY-SV5W is an independently developed sound module that integrates IO segment trigger, UART serial port control, standard MP3 and 7 other operating modes. Some of its design specifications are included in the table below.

Figure 8 MP3 player module

Table 4 Design Specifications for MP3 player module

PARAMETER	SPECIFICATION
Supply Voltage	DC 5V
Built-in	5W class D amplifier, 24-bit DAC
Playable formats	MP3, WAV

Other hardware components include; switch, PCB, safety glasses, LEDs and resistors, USB-TTL Serial Conversion Adapter module, lithium-ion batteries and a battery holder.

3.4.2 Software Requirements

These include software such as:

- **Arduino IDE:** This was used for writing, compiling and uploading the system code to the microcontroller boards i.e.; Arduino nano and the ESP32-CAM.
- **Google Collab:** This was used to train the object detection model with the labelled dataset as it provides the required resources such as GPU and TPU.
- **LabelImg:** This was used for annotating the different images in the custom dataset.
- **Fritzing:** This was used to create the breadboard view assembly of components and the schematic circuit of the system.
- **Draw IO:** This was used to design the logical design of the system through a flowchart.
- **PyCharm Community Edition IDE:** This was used to write the python code for the custom object detection algorithm in order for it to be trained.
- **Proteus:** This was used to simulate the working instances of the system.

CHAPTER FOUR: SYSTEM ANALYSIS AND DESIGN

This chapter gives a description and analysis of the developed system, project design and its implementation in a detailed form. The chapter provides an overview of the financial analysis, requirement analysis, and system design of the project, which are aimed at achieving the project's main objective.

4.1 Functional Analysis

The functional analysis of the system is determined by identifying and defining the functions that the project must perform in order to achieve its objectives. Functional analysis is divided into functional and non-functional requirements.

4.1.1 Functional Requirements

These refer to the purposes, the system is to perform. Functional requirements of the project include the following;

- Scene capture. The smart glasses system captures image scenes ahead the user.
- Object detection algorithm. The system runs a squeeze and attention custom object detection algorithm that can recognize and locate common objects and obstacles in the environment.
- Object recognition and interpretation. The system recognizes and interprets a number of surrounding objects in real time, both indoors and outdoors such as people, computers, cups, doors, trees, stair cases, among others.
- Audio Output. The system produces audio output for the different corresponding data for the detected objects and the navigation options thereafter.

4.1.2 Non-functional Requirements

These refer to the criteria that the system must meet. Nonfunctional requirements define how the system must perform and they include the following;

- Usability: the system is user friendly to the visually impaired individuals.

- Performance: the object detection and processing are fast enough to give real time assistance.
- Reliability: the system functions properly without frequent failures and downtimes which ensures continuous object detection.
- Portability: the system is wearable and can easily be moved from one location to another.
- Accessibility: the system can easily be accessed by the visually impaired users.
- Scalability; the system is able to accommodate more components for complex scenarios without significant degradation of performance.

4.2 Requirements Analysis

This involves specifying the relevant requirements for the design and implementation of the system.

- Use case development: use cases were developed to outline typical scenarios of system operation that included outdoor environments such as market places and streets and indoor environments such as computer laboratory and lecture rooms.
- Requirement prioritization: Requirements were prioritized based on there impact on the system functionality. High priority requirements included the esp32-cam, Arduino nano and mp3 audio player module.
- Feasibility study: A feasibility study was conducted to assess the technical and economic viability of the system. It included an analysis of available technologies for object detection and accessibility.

4.3 System Design

This involves specifying the architecture, components and interfaces of the system that were developed to meet the project's requirements. The system design is categorized into the system block diagram, system logical design and system physical architecture.

4.3.1 System Block Diagram

The figure below shows the sequence of fabrication and connecting of the different modules of the smart glasses. It depicts the working principle of the different components.

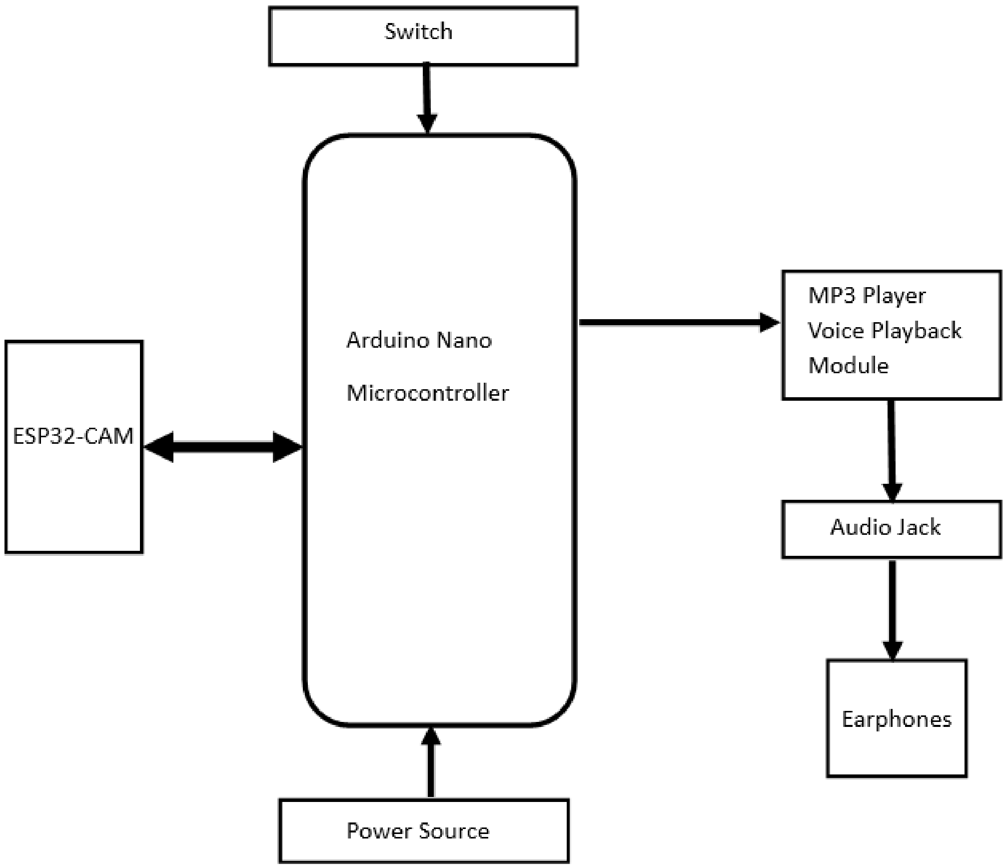


Figure 9 Block diagram showing the smart glasses system

4.3.2 System Flow Chart

The figure below shows the logical design of the smart glasses system. It depicts the step-by-step process through a procedure or system especially using connecting lines and a set of conventional symbols.

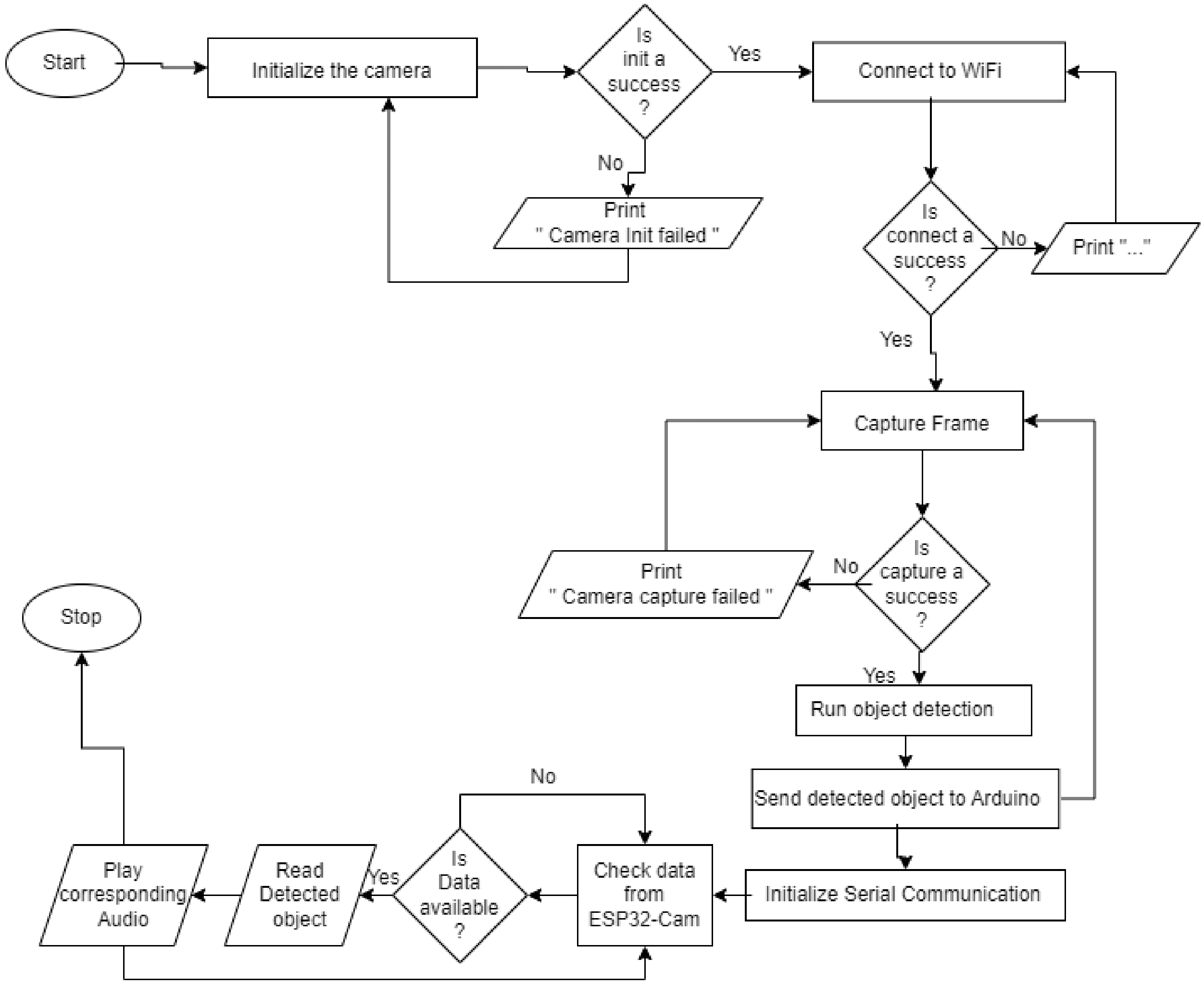


Figure 10 flowchart of the smart glasses system

4.3.3 System Circuit Diagram

The figure below shows the logical representation that defines the elements of the smart glasses system. It depicts the assembly of the different components of the system.

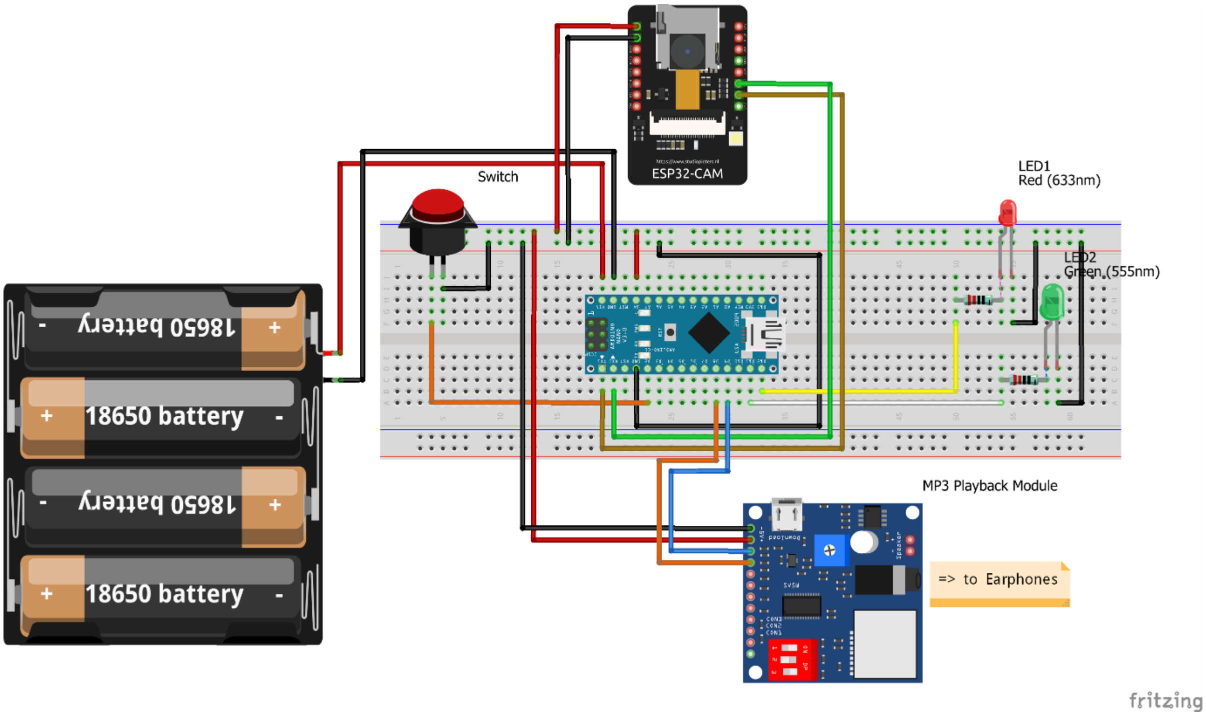


Figure 11 Fritzing Breadboard View showing the assembly of components

4.3.4 Physical Design

The figure below shows the output of fabrication and connecting of the physical components of the smart glasses. It depicts a working prototype of the different physical components.

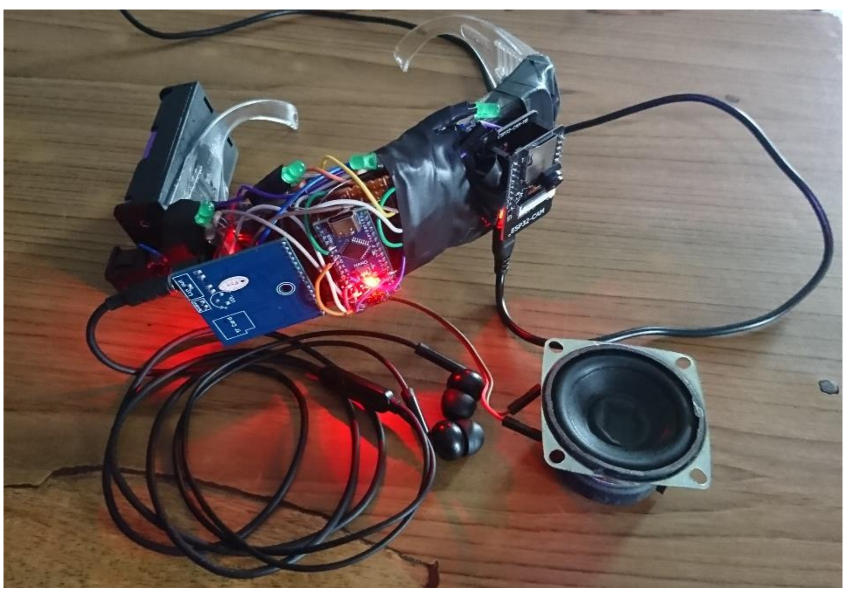


Figure 12 Physical design of the smart glasses system

CHAPTER FIVE: IMPLEMENTATION AND TESTING

This chapter describes the development of the design and results of the system. The development process was a series of steps that included the following; Implementation is partitioned into development environments, system components and implementation steps. Then system testing, verification and validation.

5.1 System Implementation

This describes all the processes followed when designing and developing the system. They include following:

5.1.1 Development Environments

These comprised of programming languages, frameworks and development tools.

- The basic programming languages were: C-language (for Arduino sketch codes), python (for algorithm development, computer vision and backend processing), and UNIX (for batch commands during training on GPU).
- The frameworks were: Proteus (for simulation), Google Colab (for model training), Drawio (for flow chart design), and Fritzing (for circuit diagram design).
- The Development Tools were: VS Code (a text editor), Python IDE (for running the algorithm), Arduino IDE (for code generation), and PyCharm Community Edition Development IDE (for developing the object detection algorithm).

5.1.2 System Components

These comprised of the data processing module, system interface and Audio Output Module.

- The Data processing module was implemented using the machine learning models and convolutional neural networks to detect, recognize and interpret objects in video scenes captured by the cameras.
- The System interface was developed to give real-time assistance to the visually impaired individuals.
- The Audio module was developed to provide the visually impaired users with real-time audio output in accordance to the objects detected by the algorithm.

5.1.3 Implementation Steps

These included the following steps.

- Setup and configuration. This involved the initialization setup and the necessary configurations for both the hardware and the software environments.
- Algorithm development. This involved the structuring, configuring, training and exporting the custom object detection model on a custom dataset.
- Integration. This involved the interfacing of the algorithm with the hardware devices and running them in unison. The trained model was exported as an Arduino library and then added to the IDE where the hardware devices would access it.
- Deployment. This involved the final set up of the object detection model and the smart glasses system as a whole.

5.2 Testing and Validation

The testing and validation method involved evaluating the performance and the usability of the smart glasses and the object detection model using quantitative and qualitative methods, such as accuracy, speed, recall, precision, user satisfaction, and user feedback. The testing and validation method ensured that the smart glasses and the object detection model can meet the expectations and the requirements of the blind users in Uganda, and that they could provide them with effective and satisfactory assistance. The testing and validation method consisted of the following:

5.2.1 Unit testing

Individual units and modules of the system were tested to see that they perform what is required accurately and at satisfactory level.

The object detection model was uploaded to Edge Impulse in order to obtain its evaluation and it returned the following test results.

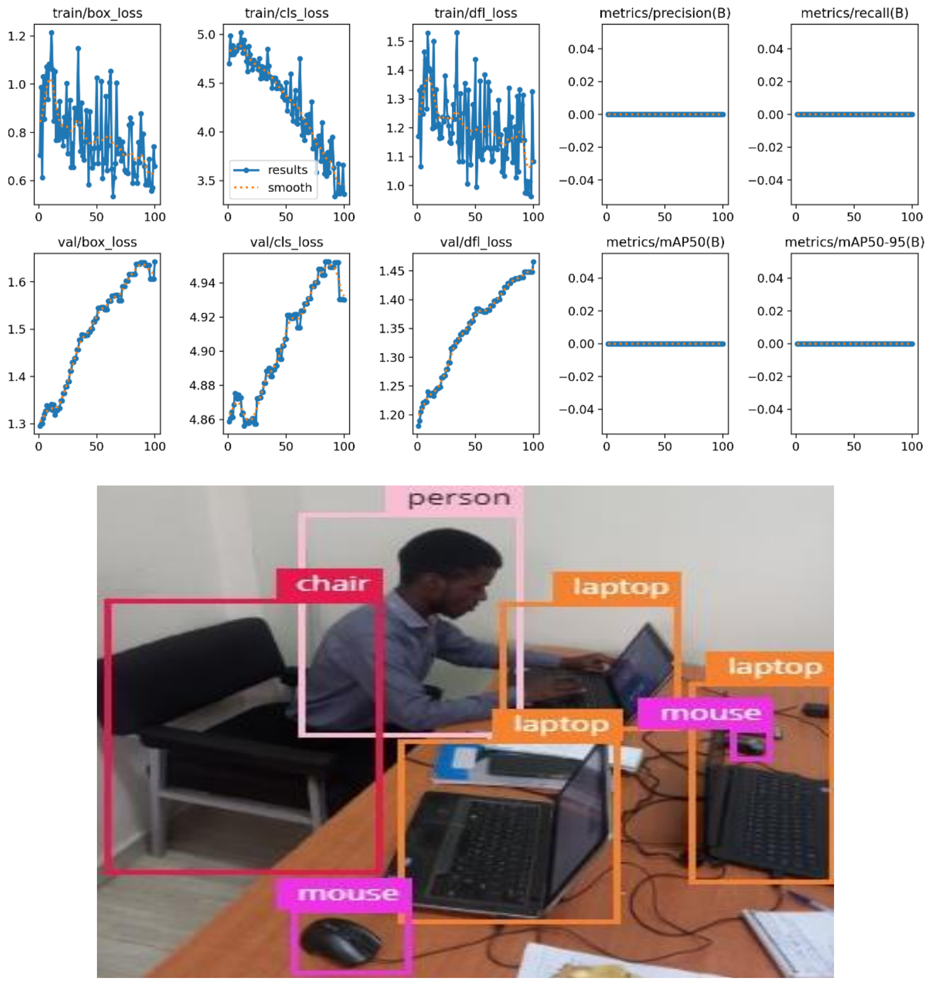


Figure 13 training results for the custom object detection model

The model returned an accuracy of 88.38% and a last training performance on the validation set F1 Score of 55.5% on the validation and the test datasets after model testing.

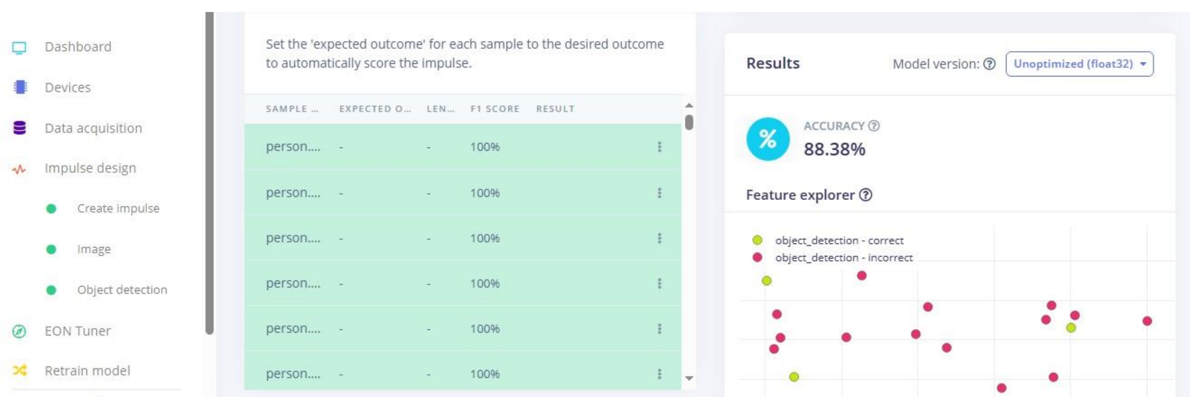


Figure 14 model results showing accuracy of 88.38%

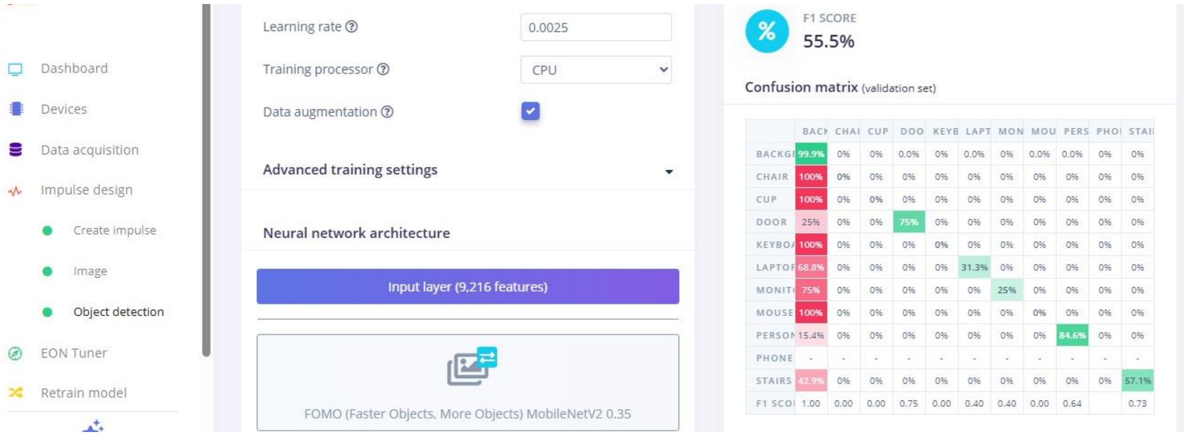


Figure 15 model results showing the F1 Score of 55.5%

The ESP32-CAM was also tested individually with a code example of the camera webserver which returns an IP address (http://192.168.1.67) in the initialization information. This url address opens the webpage for the camera with the complete number of controls.

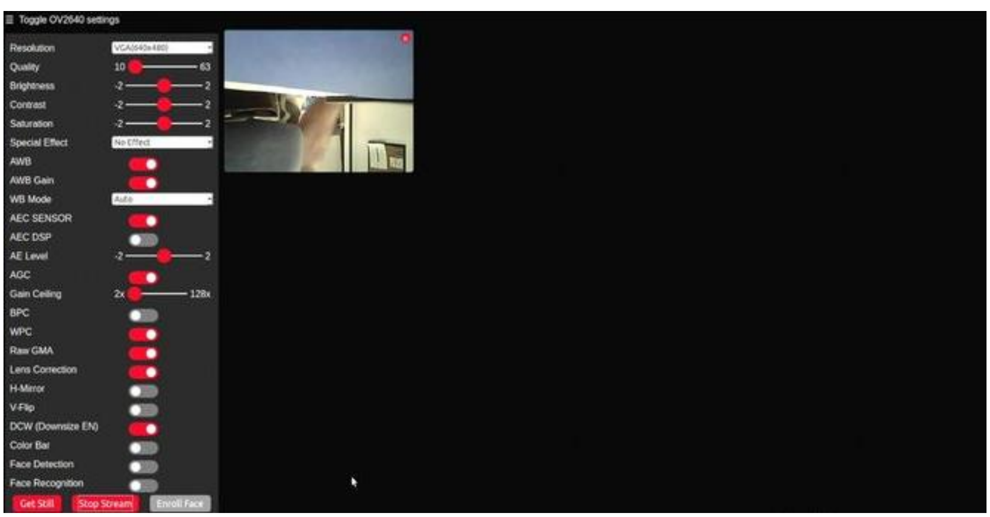


Figure 16 unit testing for an ESP32-CAM board

5.2.2 Integration testing

Different modules of the system were brought together to see that they still perform the required tasks without any significant drop in the level at which they perform these tasks.

The object detection module was compiled and uploaded to the ESP32-CAM board where the board was able to successfully run the model.

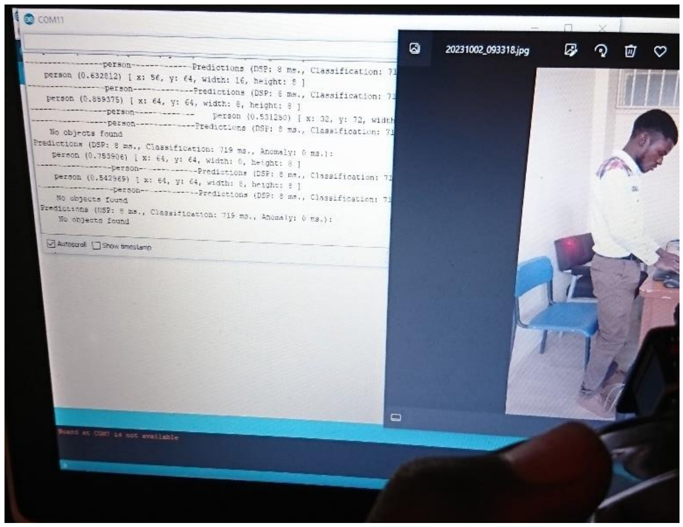


Figure 17 integration testing of Object detection algorithm and esp32-cam

The different modules of the system were able to successfully perform there required tasks even after integration. The Arduino nano was able to trigger the corresponding audio files based on the detected objects.

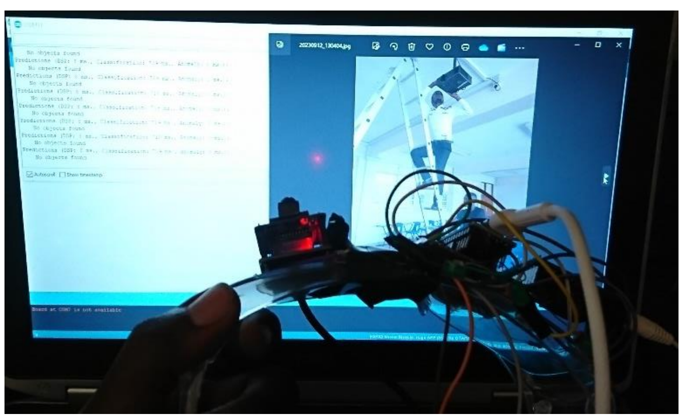


Figure 18 integration testing of the modules in the system

5.2.3 System testing

The system was tested and it fulfilled the required functional requirements. The overall system yielded the following test results: Test results for the object detection model. The tables below show the different predictions of objects, made by the algorithm with their accuracy.

PREDICTION	ACCURACY	X	Y	WIDTH	HEIGHT
Laptop	0.566406	56	56	16	8
Laptop	0.628906	56	64	8	8
Laptop	0.820312	64	64	16	8
Laptop	0.570312	64	64	8	8

Table 5 Predictions for object class 'Laptop'

PREDICTION	ACCURACY	X	Y	WIDTH	HEIGHT
Door	0.500000	16	40	8	8
Door	0.945312	32	40	8	8
Door	0.816406	16	32	8	8
Door	0.546875	24	56	8	8

Table 6 Predictions for object class 'Door'

PREDICTION	ACCURACY	X	Y	WIDTH	HEIGHT
Person	0.933594	64	64	16	16
Person	0.988281	48	40	24	32
Person	0.996094	56	56	24	24
Person	0.996094	56	56	24	16

Table 7 Predictions for object class 'Person'

PREDICTION	ACCURACY	X	Y	WIDTH	HEIGHT
Stairs	0.996094	16	16	24	16
Mouse	0.863281	8	16	16	32
Monitor	0.542969	72	48	8	16
Keyboard	0.714844	40	72	8	8

Table 8 Predictions for other object classes

The figures below show the results of the custom object detection algorithm as observed from the serial monitor of Arduino IDE and the Ultralytics hub where the algorithm is hosted to run with the necessary resources.

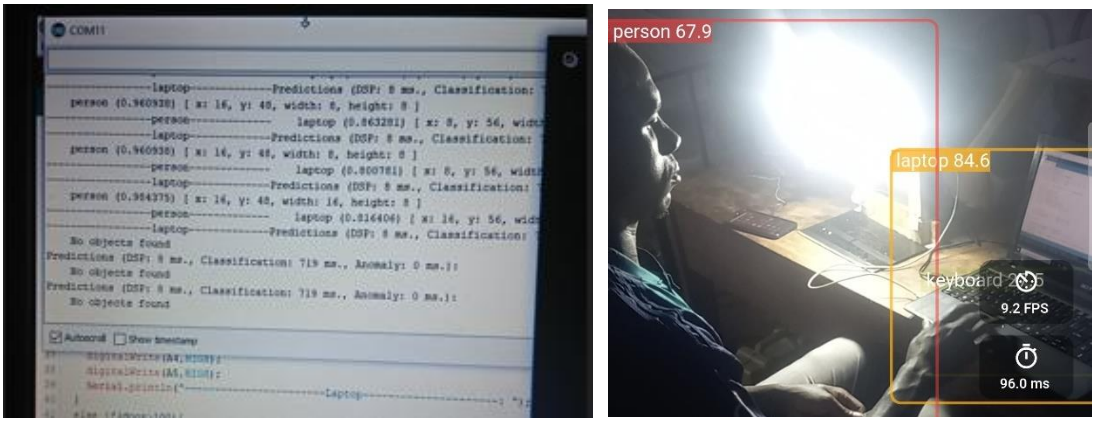


Figure 19 Predictions made by the object detection model

Test results for the audio output. The figure below shows the results of the audio output from the MP3 audio module as viewed from the serial monitor of the Arduino IDE.

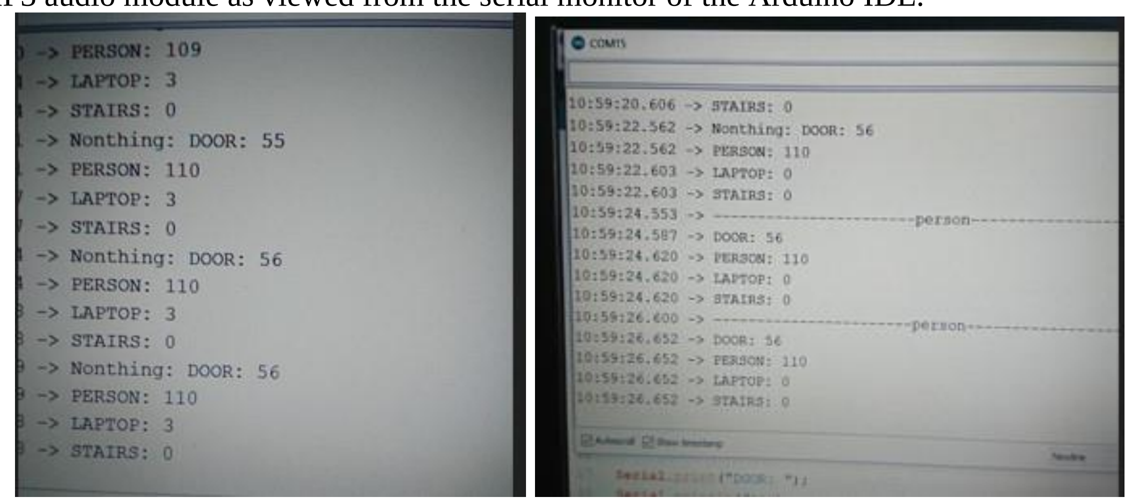


Figure 20 Serial monitor showing trigger of the audio files

The results indicate that the glasses have been fully implemented. The esp32 camera results are shown in Figure 19. This guarantees that the camera detects objects. When the camera identifies an item, as shown in Fig 19, it then plays an audio file for the blind person to hear and respond to as shown in figure 20.

CHAPTER SIX: DISCUSSIONS AND RECOMMENDATIONS

This chapter contains the work summary, critical analysis, challenges faced, recommendations for future works and the conclusion.

6.1 Work summary of the work done

The main objective of this project was to design and implement a smart glasses system for the visually impaired people using an object detection model which would give an audio output to the user about the different objects in there surrounding and, in the meantime, provide a solution to inefficient accessibility and navigation that results from their inability to perceive their environment which was successfully achieved.

6.2 Critical Analysis/appraisal of work

A smart glasses system for the visually impaired people using an object detection model was successfully designed as it was stated in the main objective. The specific objectives were also successfully achieved and it is guaranteed that the system will help a number of the visually impaired individuals across the country as they navigate through there surrounding.

The smart glasses system can be employed within different environments such as indoor environments like classrooms, houses among others and outdoor environments such as streets, market places among others, as the test results depicted the accuracy in these different operation environments.

The test results revealed that the ESP32-CAM is not the most compatible hardware for completing this project. The camera's frame rate is too low to be considered dependable, with a range of 0.5 to 4.5. This is just too sluggish, especially when assisting a blind person in determining what to expect. It also over heats up when in operation for a longer period of time which could disrupt the proper running of the object detection model. Its frame-rates and the resolutions is also too low which could interfere with the quality of the image scenes captured for analyzing by the algorithm.

6.3 Challenges faced

The smart glasses system was successfully deployed however, there were some challenges that were faced during and after the development process. These challenges included the following;

- Insufficient power distribution to other components by the ESP32-CAM board. This was solved through the introduction of the Arduino nano that was used to distribute power to all the other components.
- The computer that was used for development, lacked the ability and the resources to train a custom object model such as the Graphics Processing Unit (GPU). This was solved by enlisting the help of Google Colab an online platform that provides these resources.
- Challenge uploading code to the ESP32-CAM board using an FTDI module. This was solved by employing the use of the USB-UTL serial conversion adapter module which enabled easy and smooth upload of the code.
- The developed model suffered from overfitting which is a condition that occurs when a model is too complex and it performs well on training data but poorly on new, unseen data. This was solved by using transfer learning of fine-tuning a pretrained modal to a custom dataset, reducing the epochs hyperparameters from 1000 to 100, and also reducing the number of hidden layers.
- The ESP32-CAM board had very few GPIO pins which made it impossible to trigger a number of audio files for all the objects detected which require separate pins per audio file.
- Poor internet connection while training the custom object detection model on the online platform of Google Colab which caused several downtimes of the runtime during sessions that could disconnect the Graphics Processing Unit (GPU) leading to loss of files for the trained model during the training process.

6.4 Recommendations

The system has laid a good foundation for more research on how machine learning can be used to develop systems that help people with disabilities such as blindness. I recommend that the following can be done on the system.

- For quicker processing speed and FPs rate, I recommend the use of a high processing board such as a Raspberry Pi or Google Coral Dev board TPU that can easily and quickly run and process heavy algorithms without any delays.
- I recommend future works to include a GSM (Global System for Mobile Communication) that can send an alert message to the caretakers in case of an emergency situation.
- I recommend future works to include a GPS (Global Positioning System) that can be used to give more advanced and detailed navigation guidelines to the blind users for a more suitable device and also track the movement of the visually impaired by their caretakers.
- I recommend future works to implement the distance estimation of the different obstacles and objects in front of the blind user using techniques such as stereo vision.
- I recommend the use of a pre-trained model such as YOLO v9 for the object detection instead of a custom object detection algorithm for a higher accuracy, precision and scalability.

6.5 Conclusion:

The smart glasses system will be an effective solution to address the concern of inefficient accessibility and navigation for the visually impaired people that results from their inability to perceive their surrounding environment in Uganda and the world at large. By implementing this system, it will successfully provide a means to enhance and improve the quality of life of the visually impaired.

References

[1] I. A. C. (http://ieeeauthorcenter.ieee.org/), "Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)," IEEE Xplore, 14 - 15 September 2023. [Online]. Available: https://doi.org/10.1109/ICSEIET58677.2023. [Accessed 25 February 2024].

[2] O. (https://optimadesign.co.uk/), "Country Map & Estimates of Vision Loss - Uganda," The International Agency for the Prevention of Blindness (IAPB), 23 February 2021. [Online]. Available: https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/countries/uganda/. [Accessed 25 February 2024].

[3] R. Athman, "River Blindness increases in Uganda," Africa News, 27 November 2021. [Online]. Available: https://www.africanews.com/2021/11/27/river-blindness-increases-in-uganda/ . [Accessed 14 March 2024].

[4] T. Abet, "Why Uganda has high blindness cases," MONITOR, 04 April 2021. [Online]. Available: https://www.monitor.co.ug/uganda/special-reports/why-uganda-has-high-blindness-cases-3348236. [Accessed 14 March 2024].

[5] I. Xplore, "Smart glasses for blind people using obstacles detection," IEEE Conference Publication, 14 September 2023. [Online]. Available: Smart glasses for blind people using obstacles dethttps://ieeexplore.ieee.org/document/10303495/. [Accessed 14 March 2024].

[6] A. Georgiou, "'Smart Glasses' Developed That Can Help Blind People Recognize Objects," NEWSWEEK DIGITAL LLC, 25 October 2023. [Online]. Available: https://www.newsweek.com/smart-glasses-help-blind-people-recognize-objects-1837042. [Accessed 25 February 2024].

[7] T. S. &. G. Balakrishnan, "YOLO glass: video-based smart object detection using squeeze and attention YOLO network," SpringerLink, 31 December 2023. [Online]. Available: https://doi.org/10.1007/s11760-023-02855-x. [Accessed 25 February 2024].

[8] L. Steffen, "These AI-Powered Smart Glasses are Changing the Lives of Blind People," Intelligent Living Media Ltd, 21 March 2022. [Online]. Available: https://www.intelligentliving.co/ai-powered-smart-glasses-changing-lives-blind-people/. [Accessed 25 February 2024].

[9] seva.org, "Uganda Fact Sheet," September 2023. [Online]. Available: http://www.seva.org/pdf/Seva_Country_Fact_Sheets_Uganda.pdf. [Accessed 25 February 2024].

[10] I. R. blog, "What is computer vision?," IBM, [Online]. Available: https://www.ibm.com/topics/computer-vision. [Accessed 25 February 2023].

[11] G. D. N. N. a. K. A. A. Krishnan, "Autonomous walking stick for the blind using echolocation and," *2nd International Conference on Contemporary Computing and Informatics (IC3I),* pp. Pages. 13 - 16, 2016.

[12] R. A. e. al, "Low cost ultrasonic smart glasses for blind," *8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON),* pp. Pages. 210 - 213., 2017.

[13] F. v. d. H. &. P. Regtien, "Wearable navigation assistance - a tool for the blind," *University of Twente, The Netherlands, Measurement Science Review,* vol. Volume 5, no. Section 2, 2005.

[14] A. Kooser, "Smart shoes vibrate you in the right direction using Google Maps," CNET, 29 July 2014. [Online]. Available: https://www.cnet.com/tech/mobile/bluetooth-smart-shoes-vibrate-you-in-the-right-direction/. [Accessed 05 June 2024].

[15] I. C. Publication, "Smart Glasses Application system for visually impaired people based on deep learning," IEEE Xplore, 1 February 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9181366. [Accessed 25 February 2024].

[16] R. Tatkare, "Object detection and translation for blind people using deep learning," 2021. [Online]. Available: https://www.academia.edu/63399837/Object_Detection_and_Translation_for_Blind_People_Using_Deep_Learning. [Accessed 25 February 2024].

[17] I. C. Publication, "Smart Glasses: a visual assistant for the blind," IEEE Xplore, 10 March 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9751975/. [Accessed 25 February 2024].

[18] I. A. Center, "Computing, Analytics and Networks (Indo-Taiwan ICAN)," IEEE Xplorer, 7 - 15 February 2020. [Online]. Available: https://doi.org/10.1109/indo-taiwanican48429.2020. [Accessed 25 February 2024].

[19] M. &. C. J. Mukhiddinov, "Smart glass system using deep learning for the blind and visually impaired," Electronics, 10(22), 2756, 2021. [Online]. Available: https://doi.org/10.3390/electronics10222756. [Accessed 25 February 2024].

[20] G. Jocher, "YOLOv8," Kaggle, May 2023. [Online]. Available: https://www.kaggle.com/code/ultralytics/yolov8. [Accessed 13 February 2024].

APPENDICES / INDEX:

1. System Code:

- Esp32 cam:

The full project code has been uploaded on my GitHub account and can be accessed via this link: [hansel112/Smart-Glasses-esp32-cam (github.com)]

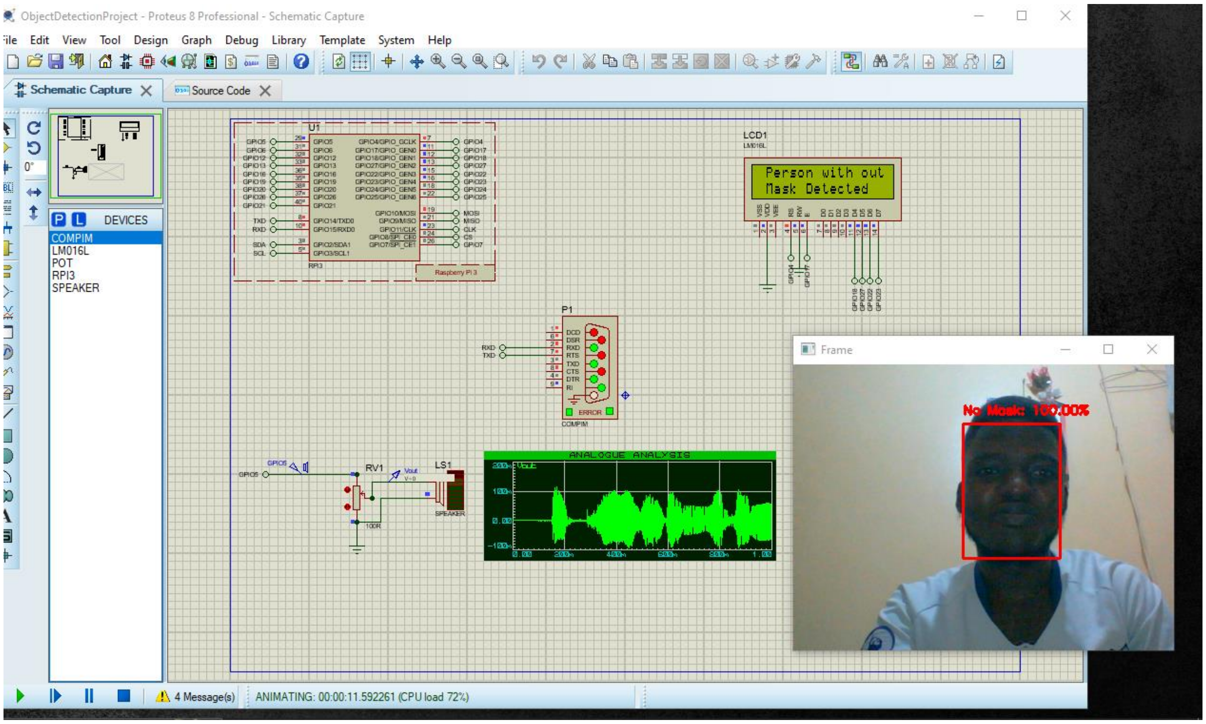
- Arduino nano:

```
void setup() {
// put your setup code here, to run once:
  // initialize the serial monitor
  Serial.begin(9600);

  // define the audio pins for the object classes
  pinMode(A3,OUTPUT); //DOOR
  pinMode(A4,OUTPUT); //STAIRS
  pinMode(A5,OUTPUT); //PERSON
  pinMode(A1,OUTPUT); //LAPTOP
}

void loop() {
// put your main code here, to run repeatedly:
  // define the ldr pins for the object classes
  int laptop=analogRead(A0);
  int door=analogRead(A6);
  int person=analogRead(A7);
  int stairs=analogRead(A2);

  // print the ldr values
  Serial.print("DOOR: ");
  Serial.println(door);
  Serial.print("PERSON: ");
  Serial.println(person);
  Serial.print("LAPTOP: ");
  Serial.println(laptop);
  Serial.print("STAIRS: ");
  Serial.println(stairs);
```

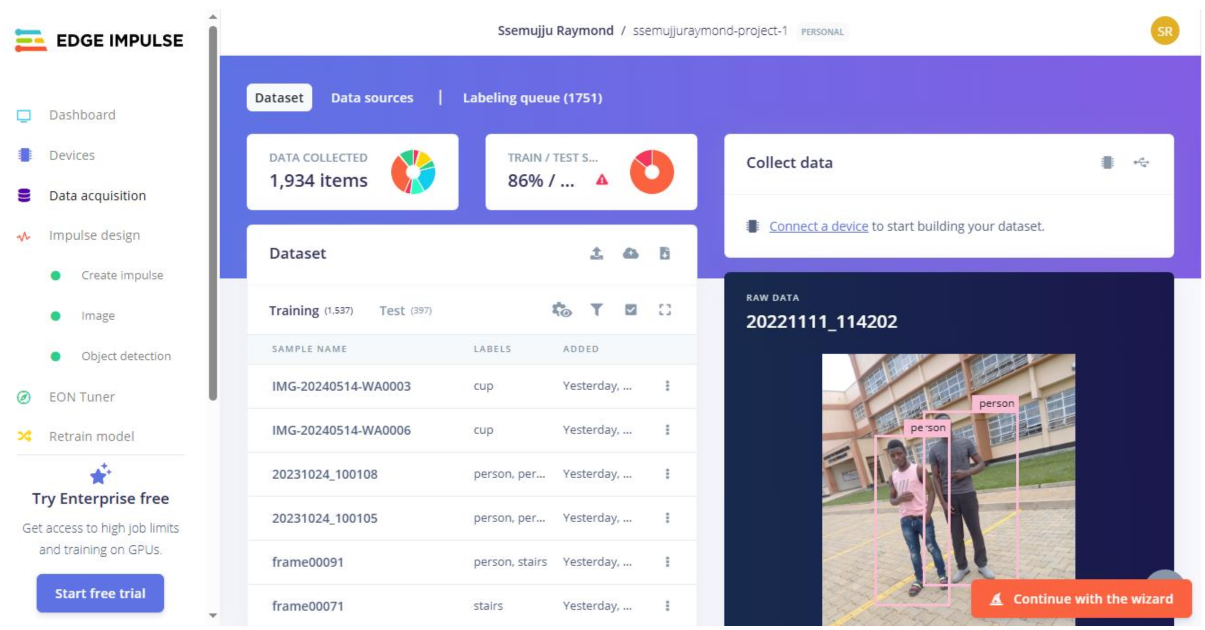
```
  //coditions to trigger the audio files
  if(person>109){
    digitalWrite(A5,LOW);
    digitalWrite(A3,HIGH);
    digitalWrite(A4,HIGH);
    digitalWrite(A1,HIGH);
    Serial.println("-----------------------person---------------------: ");
  }
  else if(laptop>3){
    digitalWrite(A1,LOW);
    digitalWrite(A3,HIGH);
    digitalWrite(A4,HIGH);
    digitalWrite(A5,HIGH);
    Serial.println("-----------------------Laptop---------------------: ");
  }
  else if(door>57){
    digitalWrite(A1,HIGH);
    digitalWrite(A3,LOW);
    digitalWrite(A4,HIGH);
    digitalWrite(A5,HIGH);
    Serial.println("-----------------------door---------------------: ");
  }
  else if(stairs>56){
    digitalWrite(A1,HIGH);
    digitalWrite(A3,HIGH);
    digitalWrite(A4,LOW);
    digitalWrite(A5,HIGH);
    Serial.println("-----------------------stairs---------------------: ");
  }

  else{
    Serial.print("Nonthing: ");
    digitalWrite(A1,HIGH);
    digitalWrite(A3,HIGH);
    digitalWrite(A4,HIGH);
    digitalWrite(A5,HIGH);
    }

}
```

2. Development platforms:

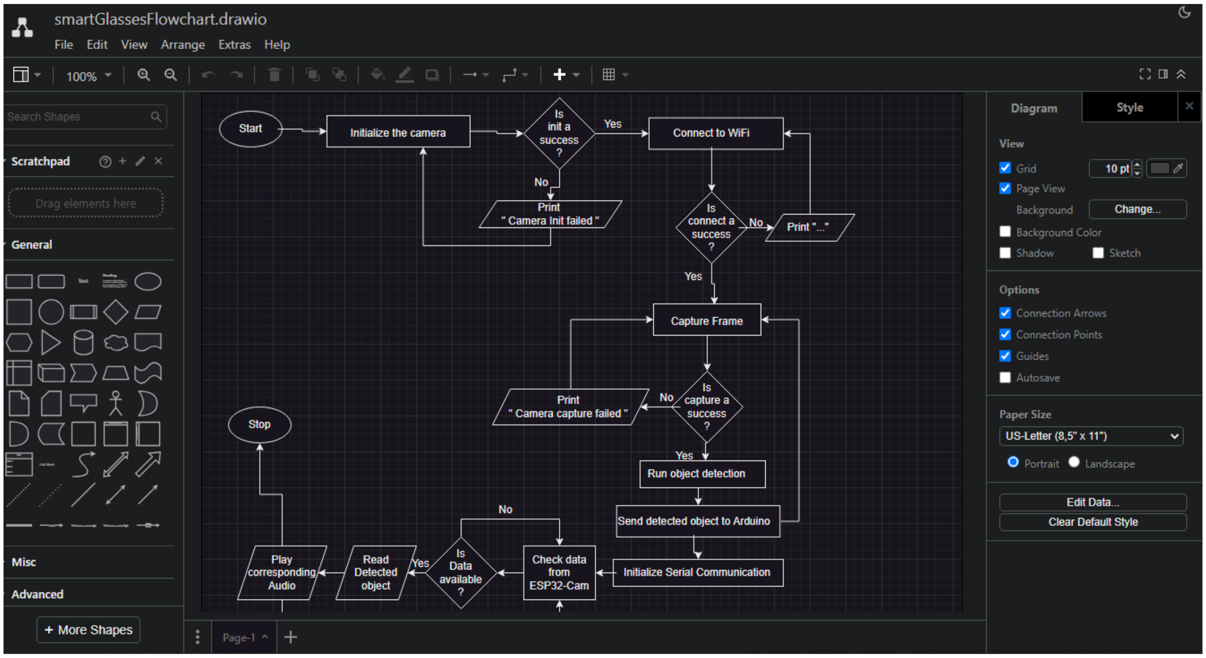
- Proteus simulation:



- Edge impulse dataset compilation:



- Drawio workspace flow chart:



- Fritzing schematic circuit diagram: