# Real-time Smart Navigation System for Visually Impaired People

Project ID: 2022-256

Project Proposal Report

Weragoda W. R. J. M.  IT19008424

Supervisor: Mrs. Sanjeewi Chandrasiri

B.Sc(Hons) Degree in Information Technology Specializing in Software Engineering

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2022

# Real-time Smart Navigation System for Visually Impaired People

Project ID: 2022-256

Project Proposal Report

B.Sc(Hons) Degree in Information Technology Specializing in Software Engineering

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2022

# DECLARATION

"I declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books)."

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

| Name | Student ID | Signature |
|---|---|---|
| Weragoda W. R. J. M. | IT19008424 | Jonani. |

Signature of the supervisor:

## ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude for the immense support, guidance and motivation provided by my supervisor Mrs. Sanjeewi Chandrasiri which always helped me for the successful completion of my undergraduate research. Her enthusiasm towards research motivated us to engage in competitions and to meet new people from the industry which otherwise would not have been possible.

Moreover, our special thanks go to our seniors for the greatest guidance and support given us in order to end this research successfully. Furthermore, we would like to thank our parents for their patience, time and providing resources to acquire the needful and for other expenses. Finally, I express my sense of gratitude to my teammates, family, friends, to one and all, who directly or indirectly have extended their support throughout this project.

## ABSTRACT

The main focus of this project is to give a better and more efficient solution for navigation issues of the visually impaired people. Visual impairment has always made it difficult for someone to go about their daily tasks normally. Their problems are countless, especially in public areas. It has always been difficult for people to read traffic signs to protect their safety when traveling on roads and crossings. The Vision Atlas of the International Agency for the Prevention of Blindness estimates that there are 295 million people with moderate-to-severe visual impairment and 43 million people who are completely blind[1]. For a long lime, white cane and guide dogs have become the solution for this matter[2]. But the only help they can provide is for locating a remote spot. The objective is to develop a portable, self-contained device that will let visually impaired people move through both familiar and unfamiliar places on their own without the use of guides. There are already a variety of technological gadgets available for providing assistance to a remote place, but these are rather pricey, or can be used for only one purpose. This introducing system is capable of identifying obstacles and potholes, walk lanes, faces of known individuals and road signs. Also, it contains an alternative solution for the accessible pedestrian signals. It allows user to make the decision whether it is safe to cross the road, or they should wait for the next turn. Furthermore, the system informs the blind person via voice through the headphone. Through this system most of the problems and challenges of visually impaired people faced during their navigation have solved. System allows them to travel with increased independence, safety, and confidence.



**Keywords**

Visually impaired, Blind Navigation, Image processing, Traffic Sign

# TABLE OF CONTENT

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATION

| | |
|---|---|
| CNN | Convolutional Natural Network |
| IOT | Internet of Things |
| ICT | Information and communication Technology |
| AI | Artificial Intelligent |
| ANN | Artificial Neural Network |
| CPU | Central Processing Unit |
| | |
| | |

# 1. INTRODUCTION

## 1.1 BACKGROUND LITERETURE

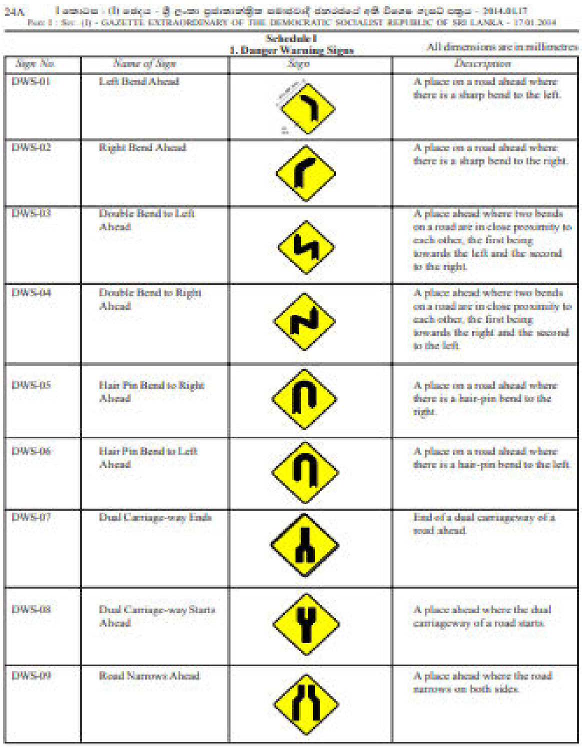
To provide road users and drivers with the information they need for safe driving and navigation, traffic signs are put on the side of the road or above the roads. There is occasionally a potential that drivers will overlook a traffic sign due to poor weather, heavy traffic, or inattentiveness, which could result in an accident. Therefore, it is crucial to automatically detect and recognize traffic signs in order to warn the driver of the danger. Road sign boards are particularly useful for giving drivers directions while they are on the road. These come with some unique instructions on the specific roads. The following types of signs are included in regulatory signage[3]:

(A) A Prohibitory Sign is required to inform motorists of the current restriction order on a route or a specific part of a road. A sign must be taken down when the prohibition it refers to is lifted or the relevant prohibition period is over.

(B) A Restrictive Sign must inform motorists of any restrictions that apply to the use of a road or a specific area of a road, as well as any removal or cancellation of those restrictions.

(C) A Mandatory Sign must make a mandatory instruction clear to drivers of moving vehicles.

(D) A priority sign is used to inform drivers of the priority that will be given to other drivers in specific conditions, or to inform later drivers of the priority that will be given to them by the former drivers in that situation, or to indicate that the priority has ended.



24A — Part I : Sec. (I) - GAZETTE EXTRAORDINARY OF THE DEMOCRATIC SOCIALIST REPUBLIC OF SRI LANKA - 17.01.2014

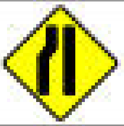
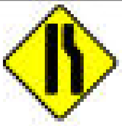
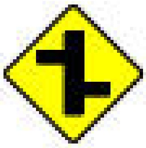
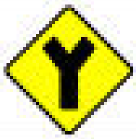
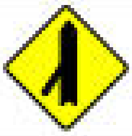
Schedule I

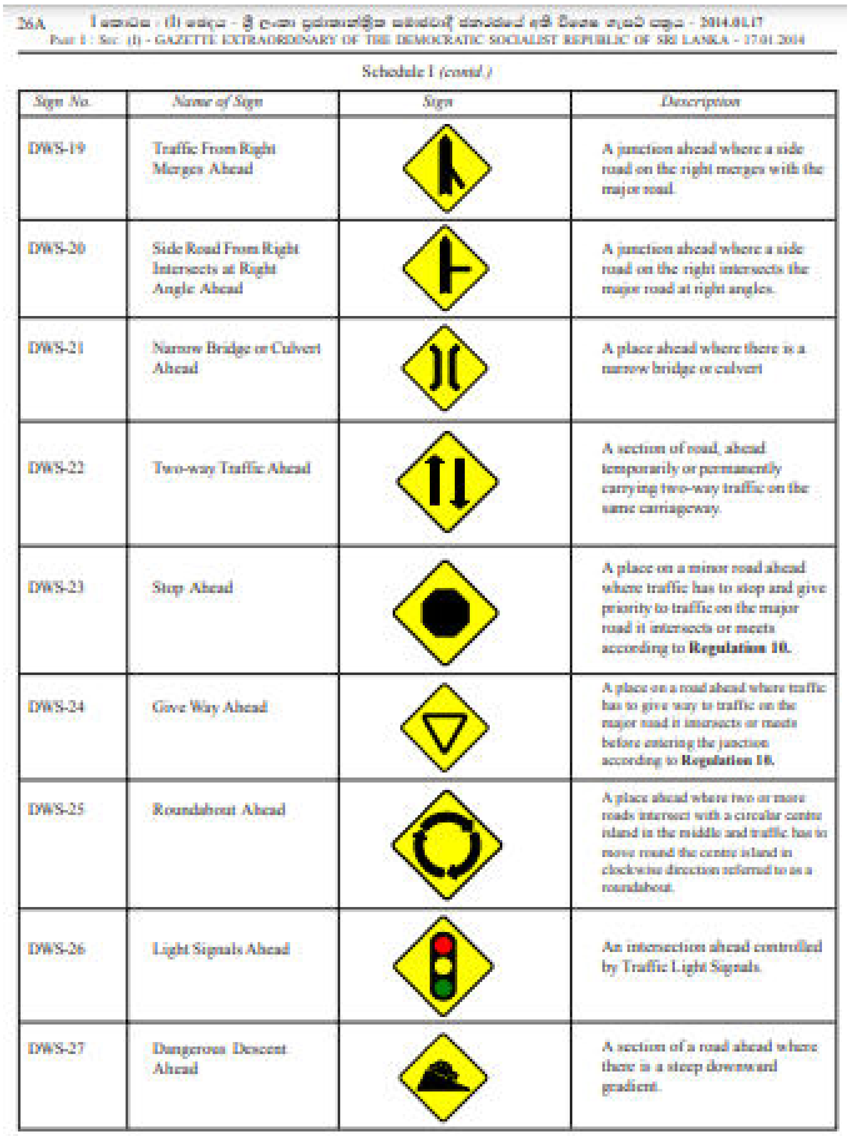
1. Danger Warning Signs

All dimensions are in millimetres

| Sign No. | Name of Sign | Sign | Description |
|---|---|---|---|
| DWS-01 | Left Bend Ahead | | A place on a road ahead where there is a sharp bend to the left. |
| DWS-02 | Right Bend Ahead | | A place on a road ahead where there is a sharp bend to the right. |
| DWS-03 | Double Bend to Left Ahead | | A place ahead where two bends on a road are in close proximity to each other, the first being towards the left and the second to the right. |
| DWS-04 | Double Bend to Right Ahead | | A place ahead where two bends on a road are in close proximity to each other, the first being towards the right and the second to the left. |
| DWS-05 | Hair Pin Bend to Right Ahead | | A place on a road ahead where there is a hair-pin bend to the right. |
| DWS-06 | Hair Pin Bend to Left Ahead | | A place on a road ahead where there is a hair-pin bend to the left. |
| DWS-07 | Dual Carriage-way Ends | | End of a dual carriageway of a road ahead. |
| DWS-08 | Dual Carriage-way Starts Ahead | | A place ahead where the dual carriageway of a road starts. |
| DWS-09 | Road Narrows Ahead | | A place ahead where the road narrows on both sides. |

Schedule I *(contd.)*

| *Sign No.* | *Name of Sign* | *Sign* | *Description* |
|---|---|---|---|
| DWS-10 | Road Narrows on the Left Side Ahed |  | A place ahead where the road narrows on the left hand side. |
| DWS-11 | Road Narrows on the Right Side Ahed |  | A place ahead where the road narrows on the right hand side |
| DWS-12 | Cross Roads Ahead |  | A junction ahead of two intersecting roads. |
| DWS-13 | Staggered Junction Ahead with First Side Road to Left |  | A place on a major road with a staggered junction ahead where two side roads intersect the major road close to each other with the first side road towards the left and the second side road to the right. |
| DWS-14 | Staggered Junction Ahead with First Side Road to Right |  | A place on a major road with a staggered junctions ahead where two side roads intersect the major road close to each other with the first side road towards the right and the second side road to the left. |
| DWS-15 | 'T' Junction Ahead |  | A place ahead where one road meets another to form a "T" junction. |
| DWS-16 | 'Y' Junction Ahead |  | A "Y" junction of three roads ahead meeting at a common place. |
| DWS-17 | Traffic From Left Merges Ahead |  | A junction ahead where a side road on the left merges with the major road. |
| DWS-18 | Side Road From Left Intersects at Right Angle Ahead |  | A junction ahead where a side road on the left intersects the major road at right angles. |



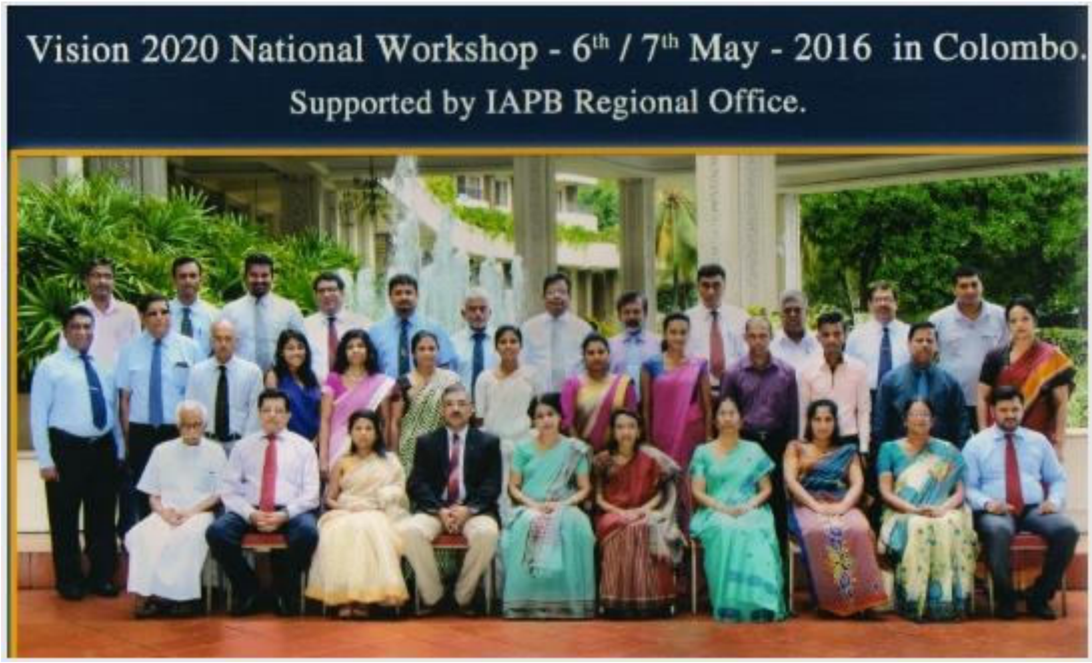
26A Part I : Sec. (I) - GAZETTE EXTRAORDINARY OF THE DEMOCRATIC SOCIALIST REPUBLIC OF SRI LANKA - 17.01.2014

Schedule I *(contd.)*

| *Sign No.* | *Name of Sign* | *Sign* | *Description* |
|---|---|---|---|
| DWS-19 | Traffic From Right Merges Ahead | | A junction ahead where a side road on the right merges with the major road. |
| DWS-20 | Side Road From Right Intersects at Right Angle Ahead | | A junction ahead where a side road on the right intersects the major road at right angles. |
| DWS-21 | Narrow Bridge or Culvert Ahead | | A place ahead where there is a narrow bridge or culvert |
| DWS-22 | Two-way Traffic Ahead | | A section of road, ahead temporarily or permanently carrying two-way traffic on the same carriageway. |
| DWS-23 | Stop Ahead | | A place on a minor road ahead where traffic has to stop and give priority to traffic on the major road it intersects or meets according to **Regulation 10.** |
| DWS-24 | Give Way Ahead | | A place on a road ahead where traffic has to give way to traffic on the major road it intersects or meets before entering the junction according to **Regulation 10.** |
| DWS-25 | Roundabout Ahead | | A place ahead where two or more roads intersect with a circular centre island in the middle and traffic has to move round the centre island in clockwise direction referred to as a roundabout. |
| DWS-26 | Light Signals Ahead | | An intersection ahead controlled by Traffic Light Signals. |
| DWS-27 | Dangerous Descent Ahead | | A section of a road ahead where there is a steep downward gradient. |

It might be difficult for those who are visually challenged to detect and recognize traffic signs. They have been using various methods to make their life easier. Most of the times visually impaired people use white cane to navigate around. They also employ guide dogs, who have been taught to assist the blind in navigating. The issue is that the dog can only travel to places where they have already received training. So, these people cannot be fully independent due to this weakness. They had to rely on their caregiver or other others to assist them in getting around, going places, etc. Due to this issue, these people are unable to live alone.

There are 295 million people with moderate-to-severe visual impairment and 43 million people who are completely blind, according to a 2021 study by the International Agency for the Prevention of Blindness's Vision Atlas. According to the World Health Organization Global Action Plan for Universal Eye Health, population-based data on visual impairment are necessary to estimate the need for services, asses s service delivery, and highlight priority that need to be addressed[4]. A nationwide poll was started since Sri Lanka lacked any national statistics. It covers five main conditions. On May 6 and 7, 2016, VISION 2020 Sri Lanka and the Ministry of Health hosted a nationwide workshop on an action plan for eye health in cooperation with IAPB Southeast Asia[5]. The government offers free medical care, which includes vision care. The major goal of the workshop was to develop modification ideas for Sri Lanka's National Eye Health Action Plan 2013–2018 based on the results of their most current blindness survey, which was conducted in




2014–2015, and to include monitoring indicators for the WHO Global Action Plan in it. According to the report, Blindness affects 1.7% of Sri Lankans over 40. Similar to this, the survey revealed that 1.6% and 15.4% of the sample population had significant vision impairment and visual impairment, respectively. In Uva province (Southeast), the prevalence of blindness is as high as 2.9%, whereas it is as low as 0.29% in the South. Nearly 67% of all blindness is caused by cataracts, which is the major cause of blindness. With a visual acuity cut off of 3/60, it was discovered that 85.4% of people have undergone cataract surgery. 59.7% and 75.1%, respectively, of the eyes that underwent cataract surgery had good presenting and best corrected visual acuity results. The ability of those who are blind to engage with their surroundings will be considerably increased by the creation of a successful visual information system.

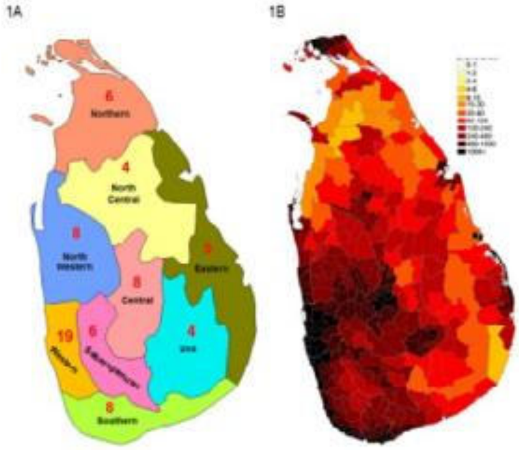


*Figure 1:Map of Sri Lanka showing location of survey clusters in each Province*

**Sampling**

The survey selected a group of individuals aged 18 years to estimate the prevalence of disability and a nationally representative sample of adults aged 40 years to estimate the prevalence of blindness. There were 25 districts and all nine provinces covered. The districts served as the main sample units.

**Sample size**

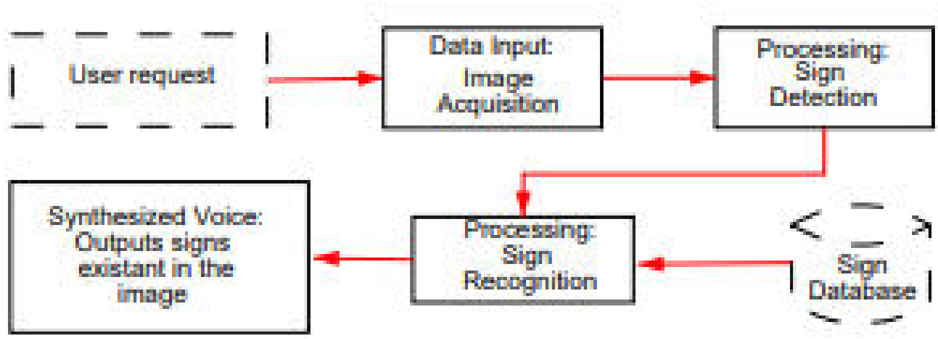
Based on earlier data from the South Asia region, the sample size was calculated using the following parameters: prevalence of blindness (presenting vision) among people aged 40 years: 2.5%; confidence interval: 95%; acceptable error: 0.02; precision: 80%; design effect: 1.5; and response rate: 85%.

There were around 6,600 samples. For accurate district-level estimates to be generated, this was not a large enough sample size, however it was designed to produce precise estimates of the extent of blindness and vision impairment on a national scale, include the main contributors to ocular illness, impairment, and blindness among those over 40.

| *Province/District* | *Number of clusters* | | |
|---|---|---|---|
| | *Total* | *Rural* | *Urban* |
| Western Province | 19 | 12 | 7 |
| Colombo | 8 | 2 | 6 |
| Gampaha | 7 | 6 | 1 |
| Kalutara | 4 | 4 | 0 |
| Central Province | 8 | 7 | 1 |
| Kandy | 4 | 3 | 1 |
| Matale | 2 | 2 | 0 |
| NuwaraEliya | 2 | 2 | 0 |
| Southern Province | 8 | 8 | 0 |
| Galle | 3 | 3 | 0 |
| Matara | 3 | 3 | 0 |
| Hambantota | 2 | 2 | 0 |
| Northern Province | 6 | 6 | 0 |
| Jaffna | 2 | 2 | 0 |
| Kilinochchi | 1 | 1 | 0 |
| Mannar | 1 | 1 | 0 |
| Vavuniya | 1 | 1 | 0 |
| Mullaitivu | 1 | 1 | 0 |
| Eastern Province | 5 | 5 | 0 |
| Batticaloa | 2 | 2 | 0 |
| Ampara | 2 | 2 | 0 |
| Trincomalee | 1 | 1 | 0 |
| North Western Province | 8 | 8 | 0 |
| Kurunegala | 5 | 5 | 0 |
| Puttalama | 3 | 3 | 0 |
| North Central Province | 4 | 4 | 0 |
| Anuradhapura | 3 | 3 | 0 |
| Polonnaruwa | 1 | 1 | 0 |
| Uva Province | 4 | 4 | 0 |
| Badulla | 3 | 3 | 0 |
| Monaragala | 1 | 1 | 0 |
| Sabaragamuwa Province | 6 | 6 | 0 |
| Ratnapura | 3 | 3 | 0 |
| Kegalle | 3 | 3 | 0 |
| Total | 68 | 60 | 8 |

The ability of those who are blind to engage with their surroundings will be considerably increased by the creation of a successful visual information system. It has been stated that a person who is blind seeks the same kinds of cognitive information as a person who is seeing does[6]. A sighted individual typically uses maps and signs to find their way around a new airport or city. The information offered by signs would also be advantageous to those who are blind.

The wearable system will have four modules in total(Figure 1). A head-mounted camera in the first module is utilized to take pictures at the user's request.




The second component, a sign detector, examines the image that was acquired by the camera and searches for regions that correspond to a sign. A sign recognizer in the third module assigns each image region to one of the signs in its database. The speech synthesizer in the fourth module, which is the last one, outputs details about the signs in the image.

## 1.2 RESEARCH GAP

The following systems were discovered during the search for comparable methodologies used to address the problem prior to moving forward with the project. There was no system in Sri Lanka that dealt with the aforementioned problems directly, although there were several systems that dealt with the major issue, "Navigation Systems for People with Visual Impairments."

Presenting a unique method for traffic detection and recognition based on an ADS, Po-Cheng Shih, Chi-Yi Tsai, and Chun-Fei Hsu (ADAS)[7]. The basic goal of ADAS is to gather important data for the driver so that they can exert less effort while driving safely. Other approaches for the detection and identification of traffic signs are explained by Sumi K.M., Arun Kumar M.N., Ph.D., Po-Cheng Shih, Chi-Yi Tsai, and Chun-Fei Hsu. One such method uses the maximum stable external region as blob detection in a picture[7]. This technique accurately detects traffic signs, and the histogram of oriented gradient algorithm is utilized to extract the feature vector for each traffic sign.

Another crucial technique for the detection and recognition of traffic signs is explained by Feng Lin, Yan Lai, Lan Lin, and Yuxin Yuan[8], Ayoub Ellahyani, and Mohamed El Ansari[9]. This method is nothing more than support vector machines. SVM is made to address the issue of binary classification. SVMs are supervised learning algorithms that generate repeatable pattern recognition performance. Finding the best hyperplane to divide the data sets so that the margin between them is maximized is the aim of SVM. The key benefits of SVM are that it can be theoretically studied using ideas from computational learning theory and that, when applied to actual problems, it produces good results. However, SVM's large dimensionality and longer time requirements are its drawbacks.

At least a few attempts have been made to help those who are vision impaired with their daily routines. Most initiatives improve the tools that blind people already use with new cutting-edge technology. This keeps the amount of stuff blind people need to carry the same. Human aid services are also available remotely, in which blind people can receive telecommunications-based aid for a variety of tasks. The latter is typically used by big organizations.

A text-to-speech application was developed in 2010 by the Hewlett-Packard and Mobile Speak Pocket businesses to make mobile phone technologies accessible to people with visual impairments [10]. This software converts the visual data from the phone into speech. Users may browse the internet, chat with friends, and move around the many screens of their mobile phones when equipped with the phone's functioning buttons, all without the need for perception. The Intel Reader [11] devices, another effort from Intel in a related direction, have text recognition software that reads material aloud. When combined with object recognition systems, this is quite beneficial. Because text alone can occasionally be used to display navigational information. The Intel Reader takes text-containing images and converts them into audio by using text recognition and image processing techniques.

By supporting people who are blind with grocery shopping, Foo et al. presented yet another intriguing piece[12]. It can be highly expensive for grocery businesses to help persons who are blind buy food. They may be able to buy food on their own with the use of this application.

Given that eating is a daily necessity, this might make life easier for persons who are blind or visually impaired. The object identification, audio localization, and text-to-speech notifications employed in this program are fundamental computer vision techniques. One of the hallmarks of the Grozi project was its ability to help people find aisles using sound localization and identify items and aisle segments using object recognition. A multi-dimensional microphone system was utilized to pinpoint the location of the actual 3-dimensional scene in each isle after a specific audio for that isle was repeatedly played. Although this is a workable method, it can only be used in quiet settings. which is highly unlikely given how noisy grocery stores frequently are during business hours. It may be more efficient to use an image-based recognition technology that can read QR codes put at the start of aisles.

**Table 1: Research gap table**

| | Research A [1] | Research B [2] | Research C [3] | Research D [4] | Researc h E [5] | Research F [6] | **Smart Navigation Tool** |
|---|---|---|---|---|---|---|---|
| Indoor Sign Detection | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Outdoor Sign Detection | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Notify user via Voice mode | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ |
| Identify text in signs | ✓ | ✗ | ✓ | ✗ | ✗ | ✗ | ✓ |
| Overcome lighting conditions. | ✗ | ✓ | ✗ | ✗ | ✓ | ✓ | ✓ |
| Overcome different viewing angels | ✗ | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Occlusion and Cutter | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |
| Cover all the colours and shapes that signs can posses | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✓ |

The majority of the studies and apps cited previously are incapable of making decisions. The majority of the study is focused on developing new machine learning models and identifying new technologies that can more effectively detect basic road signs and indoor signage. Research A[1] is the only one which is able to detect indoor and outdoor sign, texts in signs and which is able to notify user via voice mode. Research B[2] is the only system which has abled to overcome lighting conditions and different viewing angels' problems. Notifying user via voice mode has achieved by research A, B,C and D. So, there is not any other system which is able to fulfil decision making. In this introduced system, user will be able to make decisions using the tool. When there are two pedestrian crossings near by the user, they will notify about the closest pedestrian crossing. So, user can make decision about, which pedestrian crossing they are going to use.

## 2. RESEARCH PROBLEM

There are almost 250 million visually impaired persons in the globe, however the majority of interior situations remain inaccessible to them[13]. The need to support blind people's navigational needs has been overlooked by technological advancements over the last few decades. Those with vision impairments frequently need more help with daily tasks than people with other disabilities. A visual impairment restricts one's capacity for learning, social interaction, and object recognition. Even though there are important attempts that assist in giving blind individuals environmental feedback (through noises), they have not yet been standardized or sufficiently spread to a sufficient range of interactive devices. Select of the initiatives include the voice feedback systems on some elevators and grocery store kiosks, however these only cover a small portion of the environments where blind individuals need this type of environmental feedback.

Various autonomous systems have used computer vision and image processing techniques during the past 20 years. In this area of research, enormous progress has been made. Current research in computer vision demonstrates that recognition systems with a high enough recognition rate can recognize and differentiate a variety of items, including license plates, road signs, handwriting, and more[14,15,16].

Facial recognition, malignant cell detection, and other medical illness analyzers are important technologies that emerged from this field[14], [17], [18], [19], [20]. Only a small number of articles, meanwhile, have focused on computer-visual aids for the blind. Additionally, the majority of publications are not yet available as finished goods[21], [22]. Most of the existing options are hypothetical and under development.

According to the figures above, a sizeable portion of the population is affected by blindness or other types of vision problems. Additionally, it is evident that the bulk of them are over 50, which results in a disability brought on by their visual restrictions. One of the main causes of impairment, particularly in older persons, is vision loss [23]. This condition will lower quality of life, which will raise melancholy, anxiety, and other illnesses owing to restricted motion. According to Pawel Strumillo, vision loss is the most severe sensory impairment, accounting for about 5% of all cases. 90% of an individual's capacity for multisensory perception has been lost [24].

However, due to the fact that the majority of visually impaired people live in developing nations and struggle to make ends meet, these traditional methods of mobility aids for the blind can only be as discussed above. As a result, it makes it more difficult for them to carry out their daily tasks and restricts their ability to integrate into society. A visually impaired individual faces difficulties with even the most fundamental tasks, including autonomously moving from one location to another, which further limits their ability to participate in society and the economy [27]. Mobility assisting devices, such as braille signs and labels, magnifiers, special lights to aid low vision, and canes stretched so far, though still far from perfect, are used to treat this issue and attempt to ease the sensory handicap. Modern technology has advanced with numerous goods to close this gap. Two significant study fields were formed within this framework. To enable independent, secure movement from one place to another, the first one deals with the recognition of surfaces and potholes, while the second one deals with the sensing things around the user[26][25]. Since the currently available technical solutions for the above-mentioned objectives are expensive, aim of this innovation is to develop a cheaper alternative for this[26].

In earliest stages of this research, an informal survey of a visually impaired focus group was conducted. The focus group identified various challenges associated with blind navigation. The obstacles faced by the blind and visually impaired are numerous and complex, even with the aid of a conventional cane. Many of the blind people we spoke with still have some vision. One gentleman spoke about how retinitis pigmentosa caused him to lose his vision. He described his eyesight as having a severe loss of peripheral vision and having a lot of trouble with Night vision, contrast, and color. Any of the people is unable to read plain text or efficiently utilize a cell phone and computer without assistive technology. Thankfully, conference attendees acknowledged use a variety of assistive technology to access printed content. Reading a computer or cell phone screen is achievable by inverting the contrast and enlarging the text.

To access printed material, people without light perception employ optical character recognition (OCR) and other voice-over technology. The majority of those attending the meeting used a long white cane for navigating. The long white cane has a number of advantages, according to attendees. Navigation greatly benefits from its primary purpose of detecting the position, size, and kind of impediments. By marking curbs and steps, the cane creates a clear path and helps prevent falls. It is used to locate door openings, which can be challenging if there are surrounding large windows or if the door is constructed of glass. Attendees did mention, however, that they frequently found the cane to be insufficient for all of their navigational needs.

Due to external environmental difficulties brought on by a crowded background, obscured objects, the change in illumination, and other restrictions, Still unresolved are numerous computer vision issues. One of the difficulties is the computational difficulty of pattern recognition and visual understanding. More research efforts are necessary because there is still much that has to be accomplished in this area of study.

Developing a strong and efficient indoor sign identification system is one of the objectives of the proposed research, which also aims to further computer vision research. With the use of this device, visually handicapped persons will be assisted in navigating new situations in real-time. We will only currently accept a limited number of interior navigational indicators that are positioned on plain backgrounds with minimal color distraction. This system will alert users to the presence and orientation of a sign in front of them in order to effectively help them. Python is used to implement the suggested system in order to maximize performance.

*Figure 2: RGB image*

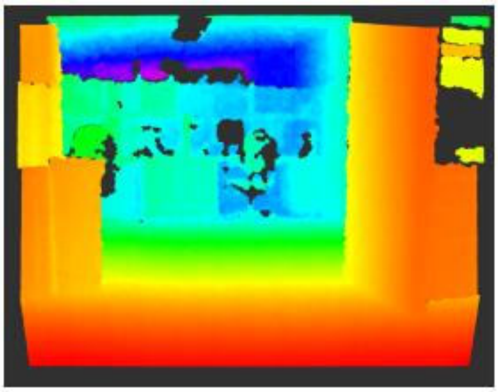


*Figure 3: Point cloud representation in 1*

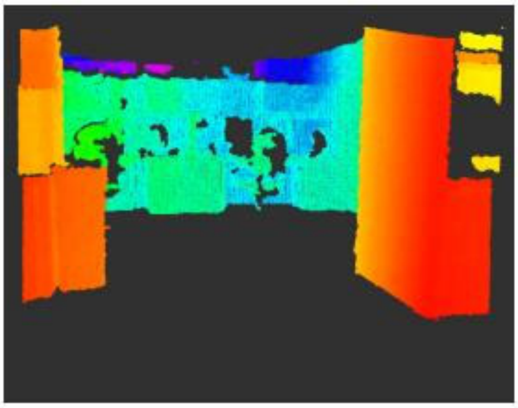


*Figure 4: Point cloud in 2 with the ground and abovehead points removed*

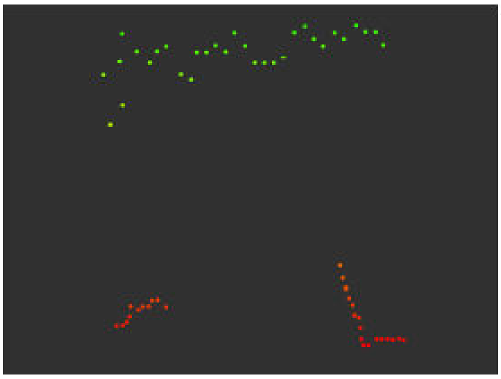


*Figure 5: Aerial view of the resulting nearest point laser scan.*

Using a cane does not eliminate the chance of falling, which could occur, for instance, if a person trips when using stairs or if they don't get enough information about where and how a curb is shaped. The requirement to entirely dedicate an arm and hand's function to cane use is another frequent source of annoyance. It can happen when bringing a plate of meals in a cafe or when using a cane and a basket simultaneously in a grocery shop. In densely populated locations with many pedestrians, using a cane presents another problem. Many participants had experienced agitation and reactivity when they unintentionally bumped into or made touch with a sighted person's heel. The frequency of touch with sighted people significantly increases when navigating through dense crowds. Last but not least, many people describe the widespread occurrence of banging their heads on low-hanging objects, such as flowerpots or tree branches.

When asked what kind of technology they prefer there were so many different answers. There were many thoughts about comprehending their environment better. Sometimes, GPS navigation smartphone apps can guide you to a destination, but they cannot convey the existence of nearby details. The existence of vehicles, trees, benches, and other objects piqued people's curiosity. The use of lightweight, compact devices is preferable to those that are large and heavy. Because some people have hearing loss, vibration feedback is especially useful. Others agreed, stating that they would prefer to hear descriptions of the objects' locations and their distances.

*Figure 6: white cane, Xtion Pro Live, EliteBook 850, and Nintendo Wii remotes.*

The development of GPS, mobile technology, and other technologies has enhanced blind people's ability to navigate the world. Local navigation, on the other hand, has not made as much progress as technology in terms of object identification, path recognition, and safety. Although there are many electronic mobility aids to buy, the usage of those things by the blind is not very common [26]. The goal is to create an additional obstacle identifying system for the long white cane that overcomes its shortcomings and improves on its advantages. This is achieved by employing sensing techniques and algorithms that support local navigation for the blind using 3D data. The techniques are going to be used in a portable device that helps the blind by giving them more sense about their surroundings. Given the anticipated growth in the number of blind people, the necessity for a device that can assist the blind in efficiently, safely and successfully navigating their environment without falling, tripping, or bumping into barriers is even more critical. According to one organization, population growth and aging might cause the number of blind individuals to quadruple over the following three decades.

# 3. RESEARCH OBJECTIVES

## 3.1 MAIN OBJECTIVES

Wayfinding and navigation are aided by signage, which helps blind persons navigate unfamiliar areas. Proposed system will offer an innovative camera-based strategy for automatically detecting and recognizing restroom and road signage in the context of surrounding settings in this research. The proposed system will present a real-time system that can recognize a variety of indoor and outdoor navigational indicators placed over clear backgrounds to help them with this difficulty. The sign selection will be made up of common samples from a variety of indoor and outdoor sign kinds. The strategy is to determine the region of interest (ROI) that contains the sign in a taken image and then extract this region for classification using image processing techniques.

## 3.2 SPECIFIC OBJECTIVES

Given that they provide directional information, road signs help individuals navigate in a big way.

1). Capture images using a camera.

a). The suggested remedy may include a handheld or wearable gadget.

2). Detect the presence of a road sign on a given image.

a). Detect the shape of the road sign using Shape detection algorithm.

3). Classify the sign by color and shape.

a). Unify the colors of the image. (Red, Green, Blue, Yellow) (Using machine learning algorithm called K-means Clustering)

i). Identify prohibiting signs – red
ii). Identify common guiding-signs – green
iii). Identify caution signs – yellow
iv). Identify room-labelling signs - blue

b). Sign Recognition: SURF

4). The sign's backdrop should be hidden.

5). If the sign has been recognized, speech sounds to alert the user to the specific sign.

# 4. METHODOLOGY

## 4.1 REQUIREMENT GATHERING AND ANALYSIS

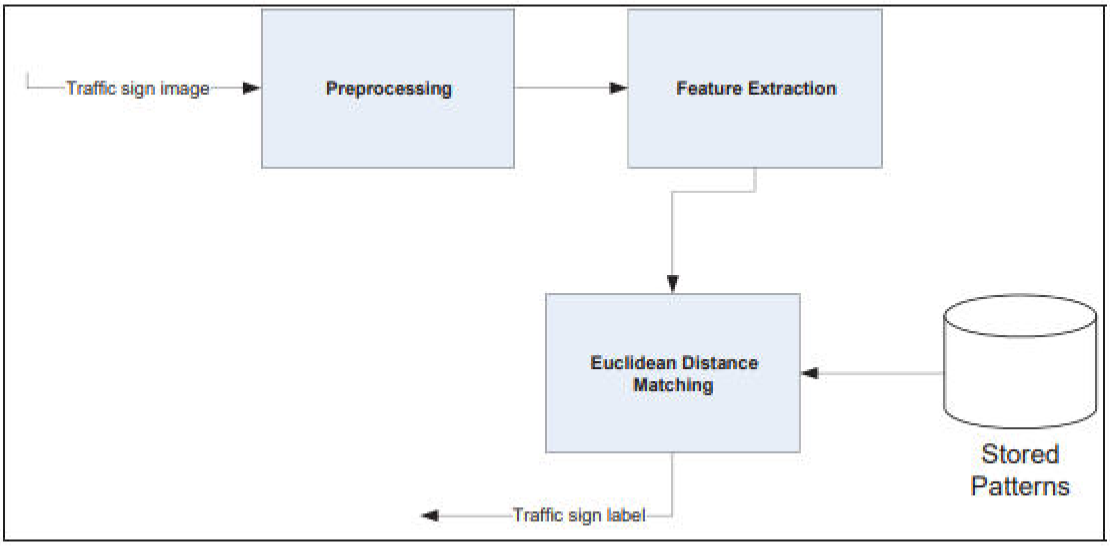



*Figure 7: Architecture Diagram*

### 4.1.1. Pre-processing

The image is enhanced by this module, which also applies thresholding and thinning to the image to make it appropriate for feature extraction. Image enhancement techniques are often applied in a variety of image processing applications where the subjective quality of the images is essential for human interpretation. Any subjective assessment of the quality of an image must take contrast into consideration. Contrast is created when two nearby surfaces reflect light with different luminance's. To put it another way, contrast is the visual difference that makes an object stand out against its surroundings and the background. Contrast in visual perception is determined by how one object differs from other things in terms of color and brightness. Because Absolute brightness is less responsive to our visual system than contrast, humans are able to perceive the environment consistently despite the significant variations in lighting conditions. To address problems with image processing, a variety of contrast enhancement techniques have been developed and applied.

In this research, adaptive histogram equalization [28] is used to improve the images. The contrast enhancement and computer image processing method known as adaptive histogram equalization is used to enhance images' contrast and has a track record of success. Unlike standard histogram equalization, it is different. Adaptive histogram equalization typically has two issues. picture of a stop sign Extraction of Feature Preprocessing Matching via Euclidean Distance Archived Patterns Optional settings for adaptive histogram equalization are available, especially in homogenous zones to restrict the contrast, to reduce the amount of noise that a traffic sign's label's slow movement and excessive amplification of the noise it makes in the image. While histogram equalization applies to the entire image, adaptive histogram equalization only applies to discrete, or tile-based, areas of the image. We can improve the contrast of each tile by employing adaptive histogram equalization, causing the output region's histogram to roughly resemble a given histogram. The combining of adjacent tiles is called adaptive histogram equalization(regions) after conducting equalization to remove artificially induced boundaries, making it useful for strengthening local contrast and edge definitions in each region of an image.

Ketcham, Hummel, and Pizer each separately developed the fundamental components of the adaptive histogram equalization (AHE) method. This version of the technique involves equalizing each pixel's histogram based on surrounding or adjacent pixels or pixels in the region where the pixel is located. In other words, each pixel is assigned an intensity proportionate to its rank in the pixels in the area(s) around it. An image with a variety of intensity levels and a contextual geographic size demands time for the adaptive histogram. Similar to how histogram equalization is traditionally done, the transformation functions are obtained from the histograms:

As with conventional histogram equalization, the transformation functions are obtained from the histograms:

$$g[x, y] = \frac{CDF[f[x,y]] - CDF_{min}}{(N \times M) - CDF_{min}}(L - 1).$$

Eq. (1)

CDF is the cumulative distribution function calculated as

$$CDF[j] = \sum_{i=1}^{j} h[i] .$$

Eq. (2)

In the above Eq. (2), the h[i] is the histogram bucket. It is calculated as

$$h[i] = \sum_{x=1}^{N} \sum_{y=1}^{M} \begin{cases} 1, if\ f[x, y] = i \\ 0, \quad otherwise \end{cases}.$$

Eq. (3)

Each pixel's f(x, y) value is assigned by the histogram to one of the L evenly spaced buckets. The CDF (min) is the smallest non-zero value of the cumulative distribution function, where M x N is the size of the image and L=256. The image is partitioned into small blocks for adaptive histogram equalization, and each of these blocks is histogram equalized. Because their neighborhood would not entirely be contained inside the image, pixels near or outside the image boundary must be given particular treatment. This holds true, for instance, for the pixels in the illustration to the left of or above the blue pixel. The image can be made larger by mirroring the pixel lines and columns with regard to the image boundaries. Copying the pixel lines on the border is not recommended since it would result in a neighborhood histogram with a sharp peak.

Similar to erosion or opening, the thinning technique is used to eliminate certain foreground pixels from binary images[29]. Only binary images can be thinned, and the result is another binary picture. A morphological process called thinning is influenced by a structural factor. The expanded type of binary structuring components used for thinning is detailed under the hit-and-miss transform. A morphological procedure known as the "thinning algorithm" is used to eliminate certain foreground pixels from binary images. While discarding the majority of the original foreground pixels, the topology (extent and connectedness) of the original region is preserved. Below is a result of a thinning process on a straightforward binary image.

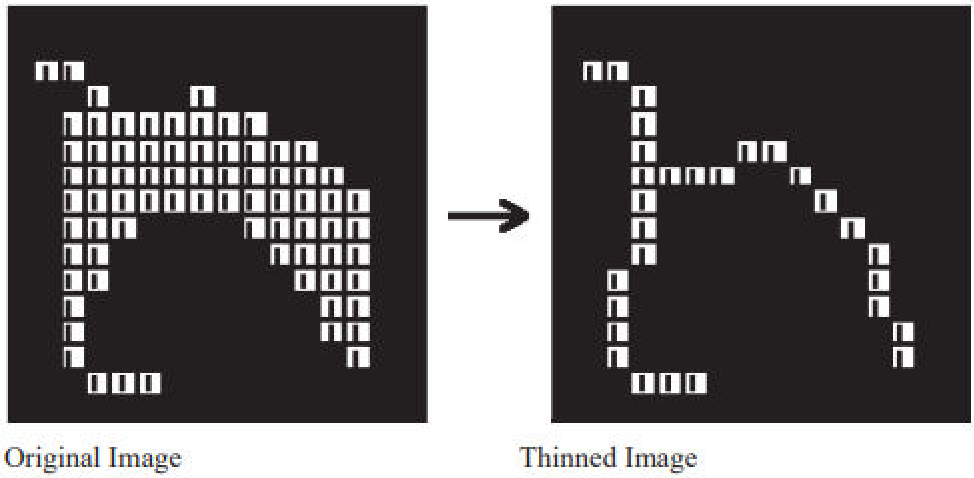



*Figure 8: results of a thinning operation on a binary image*

Four 3 x 3 templates are used in the thinning operation to scan the image. These four templates are displayed in Figure 3.

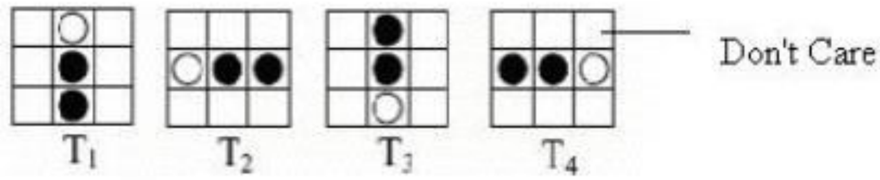



*Figure 9: template for thinning*

### 4.1.2. Feature Extraction

This module creates three horizontal and three vertical lines, then determines where traffic signs cross these lines at those intersections, using those intersections as features.

### 4.1.3. Matching

This module compares the characteristics of an image of a traffic sign with a stored pattern and assigns the matching label of the stored pattern to the person to whom the characteristics of the traffic sign are most similar. Maximal matching criteria are used in the matching process. It is the shortest distance between two feature vectors in this work. Consider that the feature vectors of the two traffic sign images are x1, x2, x3, x4,..., xn, and y1, y2, y3, y4,..., yn, respectively. The calculation for the matching distance between A and B is shown below.

$m = \sum_{i=1}^{N} F(x_i , y_i)$.

Eq. (5)

The function $F(x_i, y_i)$ is calculated as

$$F(x_i, y_i) = \begin{cases} 1 , \ if \ x_i == y_i \\ 0, \ if \ x_i \neq \ y_i \end{cases} .$$

Eq. (6)

### 4.1.4. Use Case Diagram

The project's object-oriented design is explained below. Figure 4 depicts a use case diagram, a particular kind of behavioral diagram produced from a use case study. Its objective is to provide a graphical picture of a system's functioning in terms of the players, their objectives, and any interdependencies among those use cases.

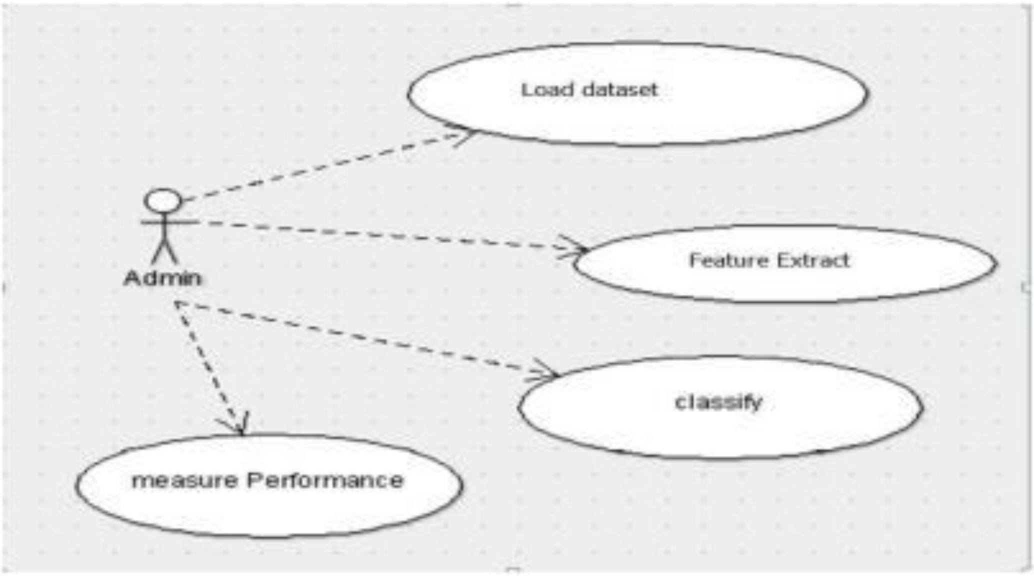



*Figure 10: use case diagram*

The user known as "Admin" has the following capabilities.

1. Data set loading

2. Features of extract

3. Sort out any input symbol you see.

4. Measure the performance.

In the Unified Modelling Language (UML), a class diagram is a form of static structural diagram that illustrates the classes, attributes, and relationships between the classes in a system to describe the system's structure.

### 4.1.5. Class Diagram

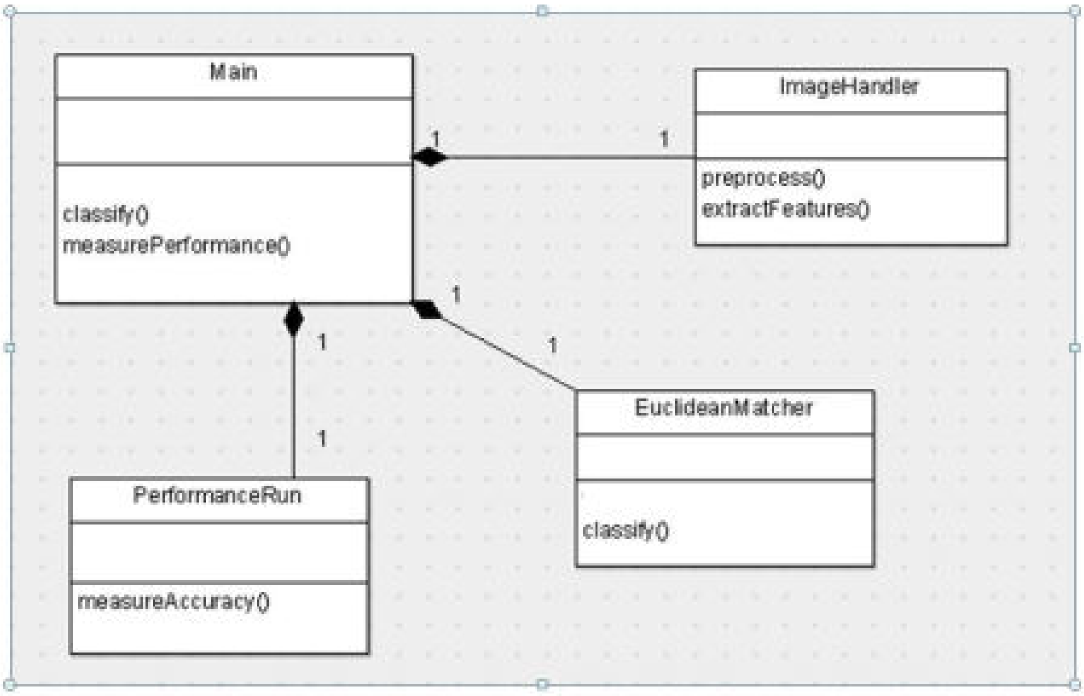



*Figure 11: Class Diagram*

The classes are listed after.

1. Main

2. Picture Taker

3. Geometric Matcher

4. Execution Run

The user interface class is the main. The pre-processing and feature extraction functions of the image are implemented by the Image Handler. The functionality of assigning the symbol to one of the traffic symbol labels is implemented by Euclidean Matches. The performance measurement functionality is implemented by the Performance Run class. In the Unified Modelling Language, a sequence diagram is an interaction diagram that illustrates the order and relationship between several activities (UML). It is a chart construct. Below are the sequence diagrams.

### 4.1.6. Loading Sequence

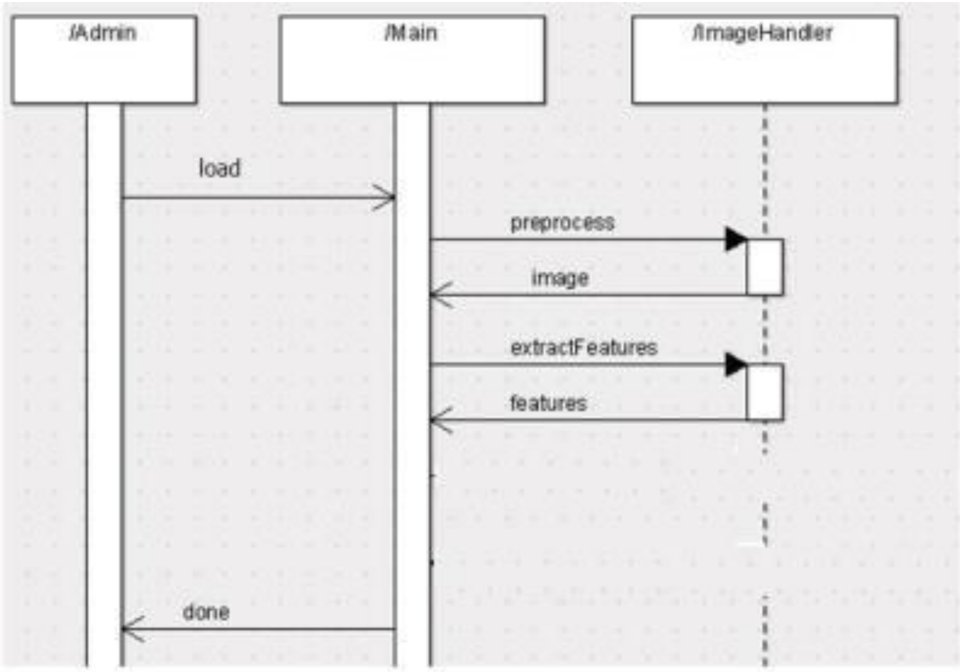
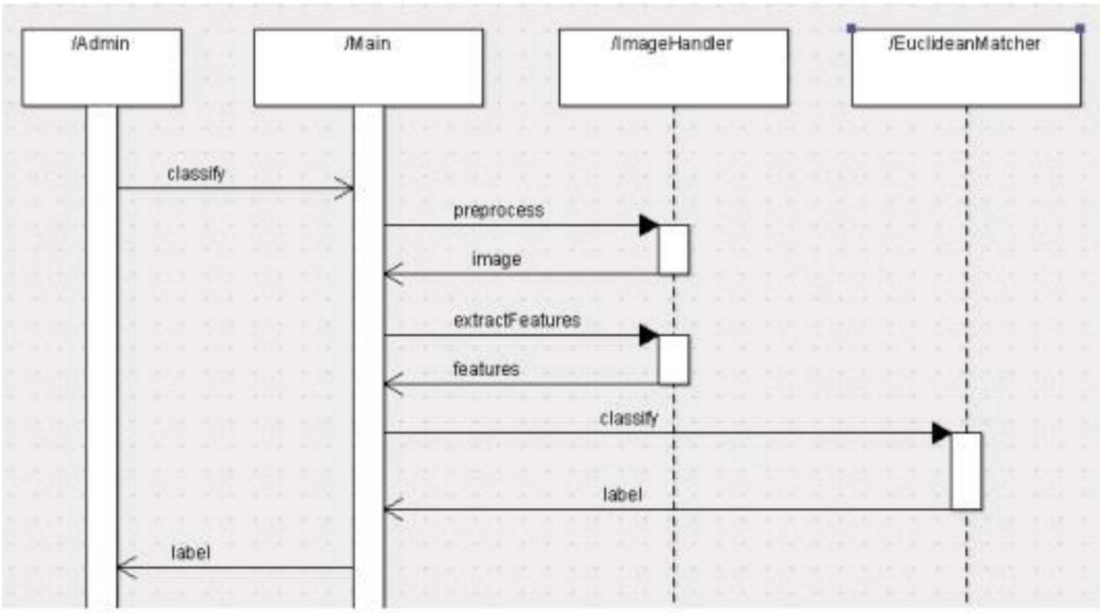



*Figure 12: loading sequence*

Admin calls load the folder containing the images of the traffic signs. Each of the pre-processed traffic sign photos has a feature that is extracted.

### 4.1.7. Classification Sequence Diagram




When the categorize function is called, an unknown traffic symbol image is provided. To discover the closest match, the image is preprocessed, its features are extracted, and they are compared to those of a stored image. The result is the label of the stored repository's closest-matching image.

### 4.1.8. Data Flow Diagram

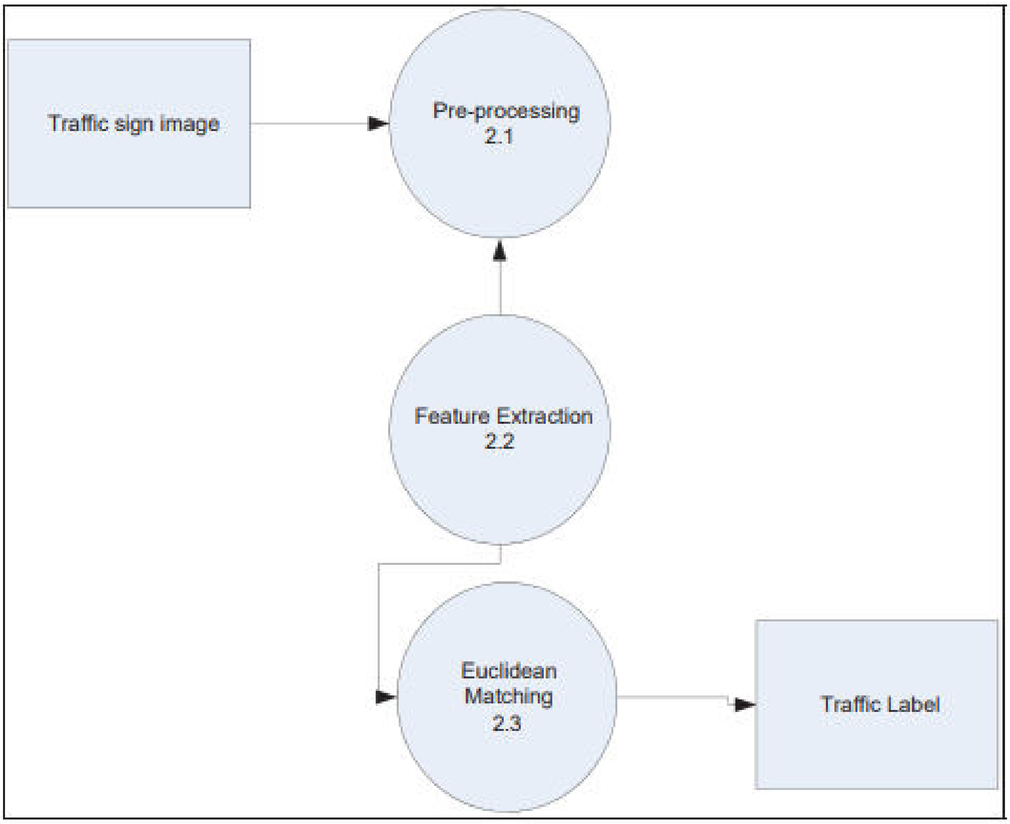
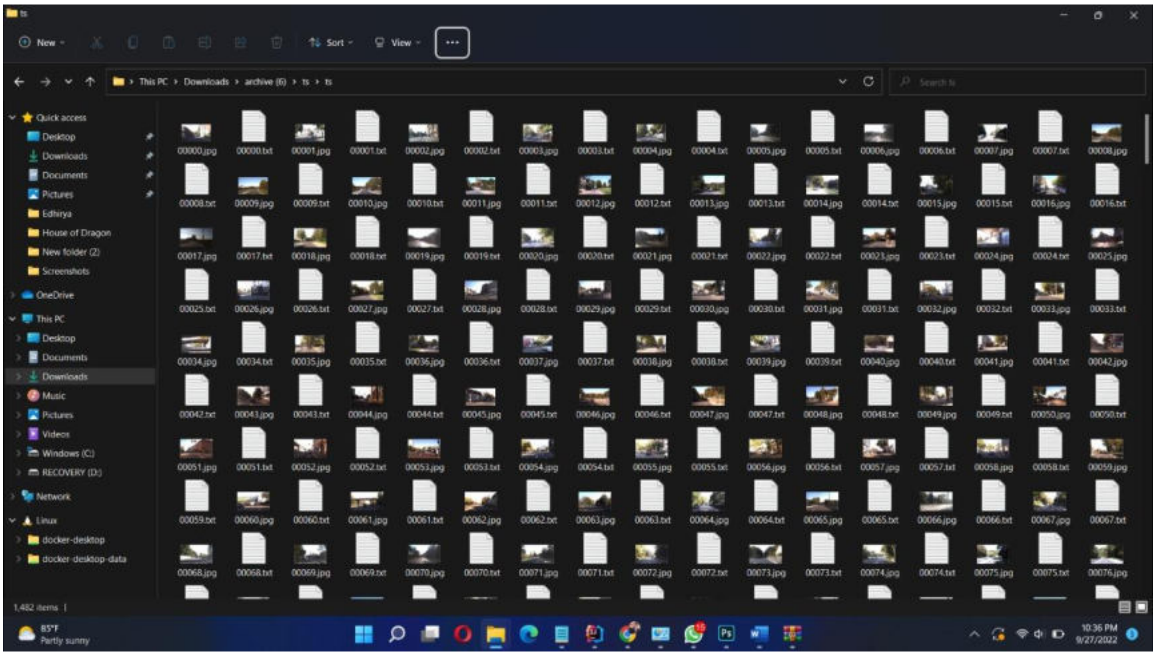



*Figure 13: classification flow*

The input image is preprocessed, its features are retrieved, and a match is made between those characteristics and those of the image stored in the repository. The label of the image with the best match is output.

### 4.1.9. System Architecture Diagram



**Feasibility Study**

The feasibility study is useful to determine the proposed project or plan is practical or not

**1. Technical Feasibility**

Project members should have some expert knowledge in web app development and knowledge in software architectures as well as frameworks.

**2. Financial Feasibility**

The proposed sub-component should work perfectly without any errors or failures. The component should be more reliable, high-performance, and less expensive. Limited cost for resources and needs of the component.

**3. Market Feasibility**

The proposed system should useful product for the industry and provide the best service for the users as well as should have competition with other systems available in the market.

**4. Organizational Feasibility**

A Member should be responsible for each phase of the software life cycle and especially like requirement analysis phase. The final output of the product should satisfy the identified requirements of the users.

## 4.3 System Designs

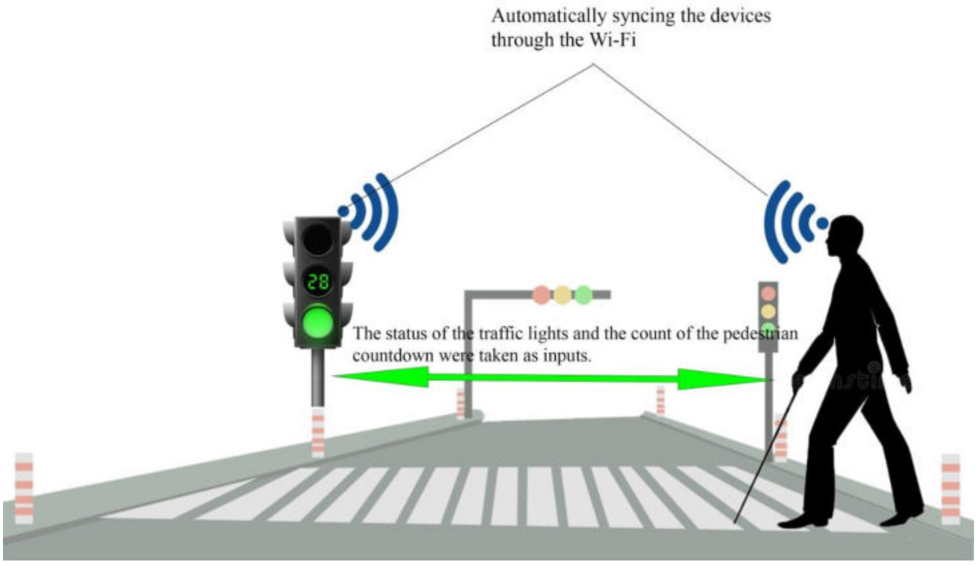
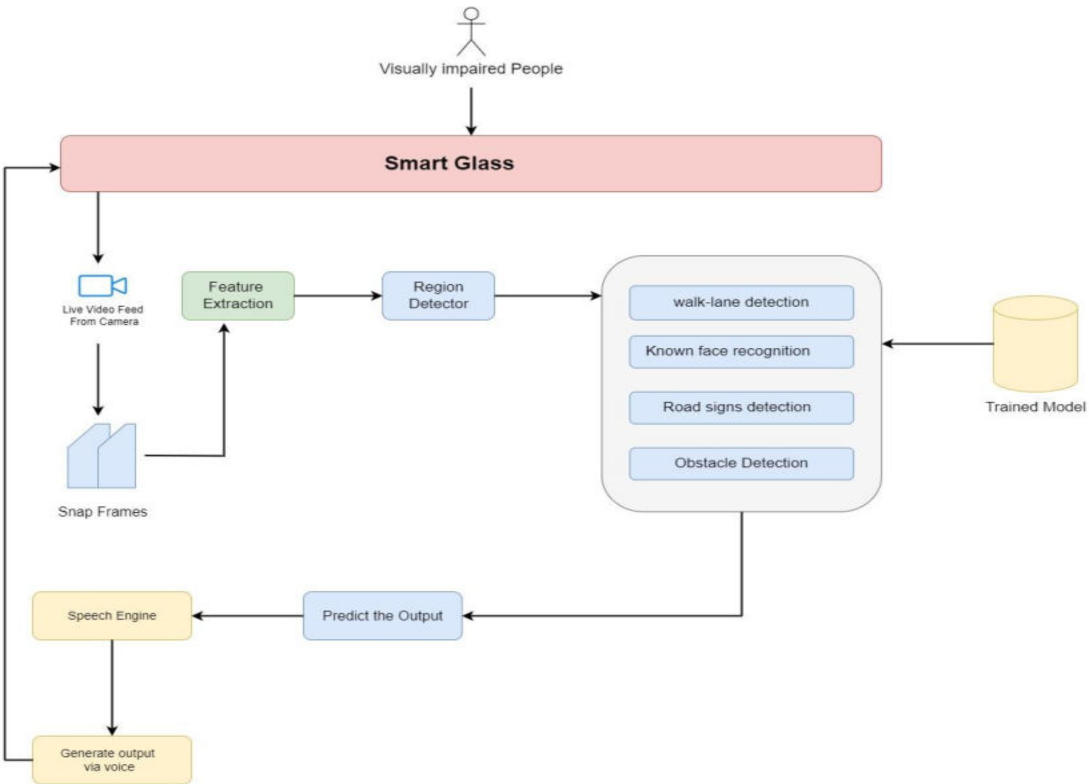
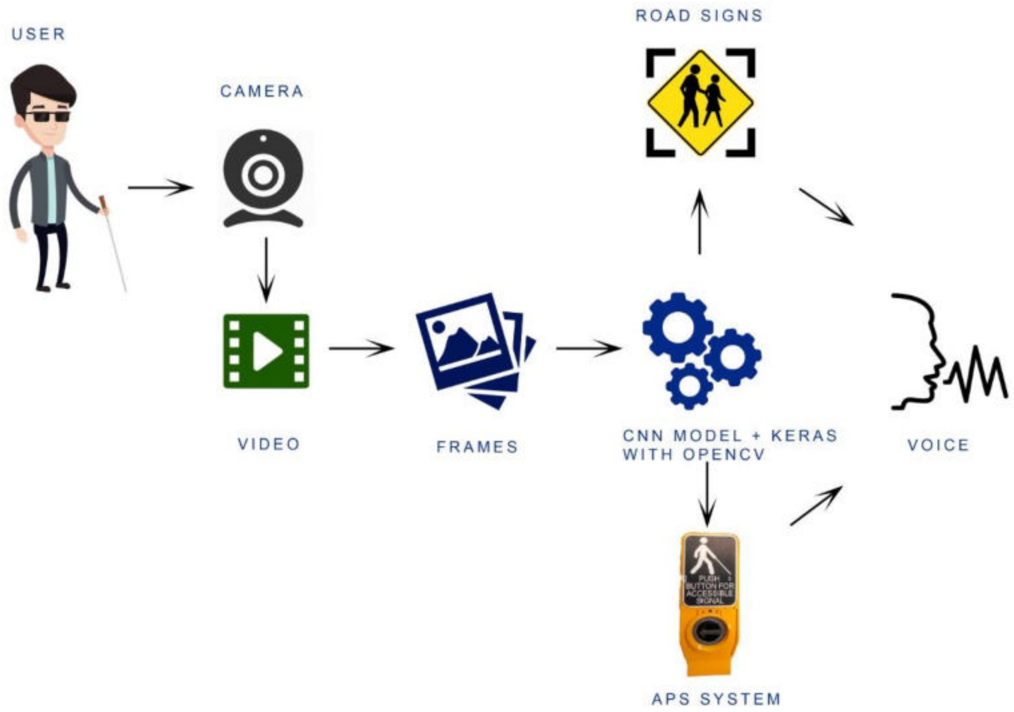



*Figure 14: Alternative system for APS system*

### 4.3.1. Overall System Diagram



### 4.3.2. Design Diagrams for the component




## Unified Detection Model – YOLO

Joseph Redmon first introduced the YOLO model in his work "You only look once, Unified, Real-time object detection." One neural network is utilized in the algorithm's design, to predict each bounding box's bounding boxes and class labels based on the input of a picture. It had speeds of up to 45 frames per second and up to 155 frames per person on speed optimized versions of the model, but it had lower prediction accuracy, primarily because of more localization errors.

To begin using the model, the input image is first divided into a grid of cells, with each cell being in charge of predicting a bounding box if a bounding box's center falls within it. Each grid cell has a bounding box that includes an evaluation criterion for quality called a confidence score, the x, y, width, and height. Additionally, a class prediction is based on every cell.

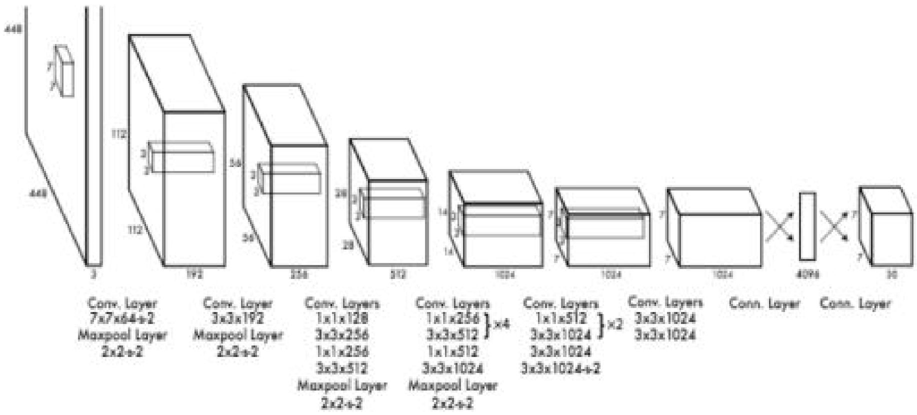



*Figure 15:YOLO Structure*

An example will be given to add greater emphasis. A 7 x 7 grid, for instance, might be used to divide a picture into cells that each forecast two bounding boxes, yielding a total of 94 predictions. By fusing the bounding boxes with confidences, the map of class probabilities, and the bounding boxes themselves, the final set of bounding boxes and class labels is produced.

The YOLO was not without flaws; the algorithm had some restrictions on the amount of grids it could operate on, in addition to some other problems that will be discussed later. First, the model employs a 7 7 grid, and because each grid can only identify one object, it caps the number of objects that can be detected at 49. The model also has a problem with near detection, which prevents it from recognizing several objects in a grid cell because each grid can only detect one object at a time. Thirdly, since an object's location could extend outside a grid, there is a chance that the model will mistakenly identify the object more than once. The aforementioned difficulties faced when running YOLO made it quite clear that the system's localization error and other issues needed to be fixed. As a result, YOLOv2 was developed as an enhancement to address the problems and queries raised by its forerunner. As a result, both real and localization problems were heavily rectified in the new 30 version. In their 2016 work "YOLO9000: Better, Faster, Stronger," Joseph Redmon and Ali Farhadi upgraded the model to significantly improve model performance.

*Figure 16:YOLO Network Architecture*

**Agile**


*Figure 17: agile methodology*

A range of iterative software development approaches that focus on customer needs are referred to as "agile software development.", and Self-organizing cross-functional teams collaborate to come up with solutions. Agile methodologies or Agile processes encourage a disciplined project management approach that encourages regular inspection and adaptation, a leadership philosophy that promotes teamwork, self-organization, and accountability, a set of engineering best practices that enable the rapid delivery of high-quality software, and a business plan that connects development with customer needs and corporate goals. Agile development is any method of software development that follows the guidelines of the Agile Manifesto. A group of fourteen prominent figures in the software industry collaborated to write the Manifesto, which is a reflection of their understanding of the ways that are successful and unsuccessful in the field.

*Figure 18: agile methodology*

Local and global image properties can be loosely divided into two types. A single feature vector is produced for each image as a result of the computation of global characteristics like texture descriptors throughout the entire image. The picture patches that surround these points are characterized by local features, which are computed at numerous locations throughout the image. For each image, the outcome is a set of feature vectors. The dimensionality of all feature vectors is the same, but the number of features produced by each image varies depending on the interest point detector utilized and the image content. Global features are easier to utilize with a conventional classification algorithm because they provide a more compact representation of an image (e.g. support vector machines). Local features, on the other hand, have a number of characteristics that make them more suitable for our application. Local features are computed at many places of interest in an image, making them more resistant to clutter and occlusion and obviating the need for segmentation. Given the sign detector's current level of imperfection, we must account for inaccuracies in the sign's outline. Local features have also proven to be quite effective in a variety of object recognition applications.

### Color Segmentation.

The images will be changed from their traditional red, green, and blue (RGB) format to a hue-saturation-value (HSV) format in order to separate the pure color information from the superfluous intensity information. A straightforward median filter will be used to equalize the intensity values after extracting segments of saturation and hue from HSV images. We're looking for the following colors for the indoor signs: red, green, blue, and yellow. To distinguish the sign from the backdrop, a binary mask will be produced for each color. This will be done by performing a bitwise operation on the color of interest to cancel out all other complementary hues.

### Shape detection.

A Canny filtered image's convex areas' contours are what the shape identification method uses to identify shapes. We'll extract all contours in the image and group vectors that create convex regions to find forms. These regions' shapes will then be approximated. Andrew Kirillov's work was used to create our shape extraction methods. To get all of the contours in the image, the technique employs OpenCV's f indContours() function. This divides contours that are close to each other into separate elements. We go through each element, selecting from the collection AllConvexShapes any elements with inner angles smaller than 180.

### 1.1.1. Hardware solution

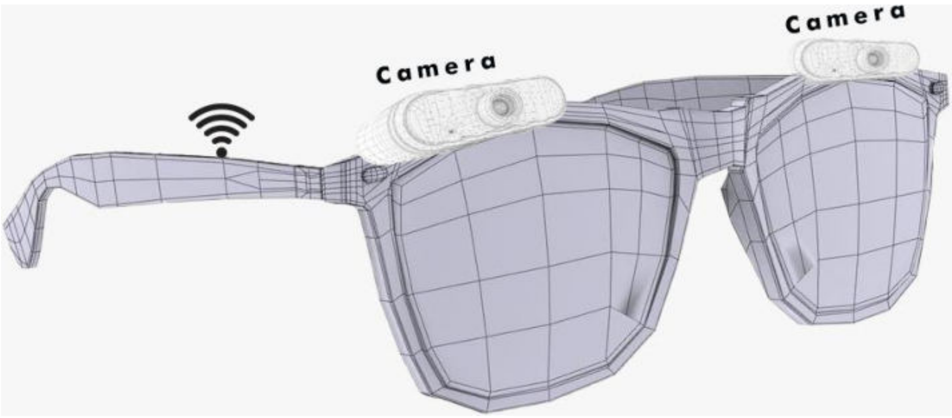



*Figure 19: hardware solution*

Nearby impediments and moving things, including translucent objects like glass, can be detected by the two cameras incorporated in the glasses. The device can also recognize faces, read signs, and identify items using artificial intelligence. Finally, the high-quality speakers included in the package can deliver audible commands.

# 5. IMPLEMENTATION AND TESTING

## 5.1. IMPLEMENTATION

In order for our model to produce better outcomes, we must first gather a vast amount of data regarding Sri Lankan traffic signs. Second, a variety of outside factors affect how accurate TSR is. A defogging algorithm is discovered. For the hazy photos, this model has a greater accuracy rate. Since there is extremely little chance that there will be foggy days, we won't have much opportunity to take our own foggy photos, therefore we'll find a ton of them to use as data. We will contrast two various deep learning techniques. At the same time, we will research how fog affects recognition outcomes. Ultimately, we will discover a system that has a quick recognition time and a high level of TSR accuracy.

With the advancement of deep learning and artificial intelligence, TSR is also developing quickly. To aid drivers on the road safely, many high-end vehicles now feature TSR driver assistance systems. Only photos shot on sunny days, with unobstructed traffic signs, and especially in poor weather with hard lighting conditions, have a high accuracy with the current TSR. If traffic signs are obstructed, output may include false alerts or no recognition. The thesis then goes on to explain how to address the issues, how the causes impact TSR, and what the distinctions are.

**Impact of Raining on TSR**

TSR on photographs shot on rainy days is a particularly challenging task since, unlike rain or snow, which have dynamic motions of visual objects, fog is immobile and has no evident motions. For TSR, which already needs real-time object detection, all of these present significant difficulties. Raindrops vary in size, speed, and shape, and their locations are essentially random. Raindrops with a shallow bulk will fall through the air slowly when there is little light available. Droplets in the air impair sight and block the camera, which could lead to traffic signs being misread before a camera deforms.

These minute water droplets make a raindrop path if they land on car glass. The traffic signs on the captured photographs will become blurry if it is raining heavily since the raindrops will be moving quickly enough to cover the windows like a curtain. Following a rainstorm, water on the ground acts as a mirror, reflecting traffic signs and perhaps altering or obscuring the TSR. A TSR system with a visibility improvement module, a convolutional encoder, mixed with a convolutional decoder, and a dark channel was proposed by Chen et al. as a solution to these issues. Utilizing photos from wet days increased accuracy in the studies by 50%.

**Impact of Illumination on Traffic Sign Recognition.**

For TSR, illumination is necessary. A traffic sign or camera that receives direct sunlight during the day will be overexposed. Driving at night is frequently too dark, and the lights of the automobile will reflect off of the road signs. This will result in inaccurate traffic sign recognition. To address these issues, Greenhalgh presented a traffic sensing technique that modifies the incoming light's intensity and enhances TSR. When paired with the geometric characteristics of traffic signs, it efficiently lessens the impact of shadows and illumination on traffic recognition, improving the accuracy of recognition under a variety of lighting circumstances. While this was happening, Kwangyong proposed a byte-MCT and traffic sign-based AdaBoost classifier for TSR, first using SVM to assess the potential regions and then CNN to extract visual features. This method produced excellent results with significant robustness under varied lighting circumstances.

**Impact of Fog**

When using a digital camera to take a picture outside in hazy conditions, the ambient light will be significantly scattered, resulting in blurred image details and the lack of feature extraction and other associated activities. We discovered that the haze photos have particular features through the analysis of numerous clear and hazy images of the same location. The hazy picture and classifier for pattern classification are automatically identified by the study of these image features. The traits can be summed up as follows. The traits can be summed up as follows.

First off, the contrast of haze photos is significantly lower than that of clear images, and when scene structure information increases, the degree of contrast attenuation exhibits features of exponential attenuation. Second, the image's key details, such as its edges and contours, are obscured. The color saturation is lower, and there is color distortion and offset in the image. The dynamic range narrows, the distribution of picture pixel values is condensed, and the pixel histogram shows a uniform distribution. The image has a lot of low-frequency information and very little high-frequency information, according to the perspective of the image frequency domain. The photos' grayscale level changes uniformly, the amount of detailed information is decreased, and the edge of the image contour is weakened in terms of visual effect and look.

The factors that contribute to haze-related blurring and image quality reduction are listed. When the weather is hazy, aerosol particles are suspended in the atmosphere. Aerosol particles will interfere with the medium used to transmit light, result in significant light absorption and scattering, and directly produce the energy attenuation carried by the light rays reflecting from the surface of the item in the picture.

The ambient air light is also scattered along the entire optical path of light rays to the imaging device so that it can take part in the imaging of scene objects. Together, the two elements lessen the image contrast and resolution. Depending on the size and shape of the particles, aerosol particles can have different effects on light scattering. In addition, while a scene is being imaged in the environment, the particles themselves take part in the imaging as well. This is known as image noise, and it leads to the loss of picture detail information.

Along the light route to the camera lens, the light reflected by the scene's object surfaces is dispersed. In addition, after multilevel scattering, light that is reflected by other surfaces of objects in the scene also contributes to imaging, causing hazy images.

The brightness of the image captured in haze weather is insufficient because the radiation strength of lights will be attenuated due to cloud cover, reducing the light intensity in the environment.

In conclusion, a lot of suspended particles in the air cause substantial light scattering, which is the main cause of image blur in haze environments.

**Traffic signs Recognition Process**

TSR has three main parts. The first section is for picture preparation, which typically includes operations like image augmentation, image scaling, and others. The second part, traffic sign detection has three important steps.

1. Region candidates extraction.
2. Determine whether a sign is a traffic sign or not to classify it as such.
3. Specialized category of traffic signs, which is a detailed classification of traffic signs..

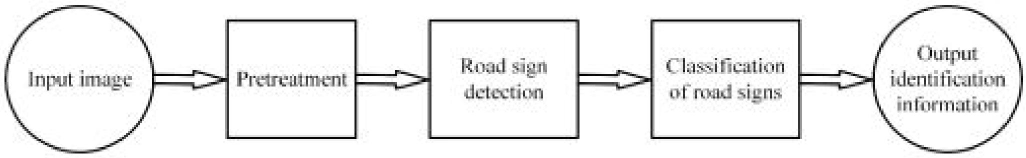



*Figure 20: Workflow of TSR systems*

### 5.1.1. Image Preprocessing

Compared to general object recognition, TSR is different. External influences including haze, rain, snow, and all types of lighting will inevitably have an impact on the traffic sign image we receive. These influences will cause the image brightness to be too low or noisy, which is not good for TSR. Preprocessing an image efficiently raises its quality, lowers or even eliminates interference elements that interfere with object recognition, and increases the effectiveness and precision of traffic sign detection. Therefore, the image needs to be preprocessed before the sign recognition in order to increase the recognition accuracy of traffic signs.

**Traffic Sign Detection**.

Traffic sign detection part has two main processes.

1. Extraction of vindicate frames
2. Classification

The extraction of candidate box is the process of segmenting the original image's area containing traffic signs based on the pertinent characteristics of those signs. The properties of 34 signs are extracted for the second categorization, which determines if the indicators fall into the relevant categories. Some of the methods used to identify traffic signs include shape detection, color segmentation, feature extraction combined with classifiers, and deep learning detection algorithms.

### Classification of Traffic Signs.

The third component, traffic sign categorization, is used to establish the precise category for the area with segmented traffic signs. Multi-object classification is essentially what classification of traffic signs is. The feature vector and classifier, as well as template matching, are combined in conventional classification methods. With the ongoing advancement of artificial intelligence, deep learning's use in TSR has produced superior outcomes. Unlike traditional methods, the deep learning classification method may automatically learn characteristics without human input. The performance of classification improves as training data volume increases. Convolution neural networks in particular have improved the capability of feature extraction in neural networks. As one of the most representative deep learning networks, it is frequently used for object classification.

### 5.1.2. Image Defogging Preprocessing Algorithm

The preceding chapter's introduction and analysis informed us that the TSR has a wide range of applications. The development of deep learning has led to significant advances in all TSR, but since most research has focused on TSR in ideal visual conditions, more work is still needed to understand how to identify signs in foggy weather. This chapter will examine the TSR under foggy conditions from the viewpoint of a fog image. In the midst of the fog, there are four steps to roadside identification.

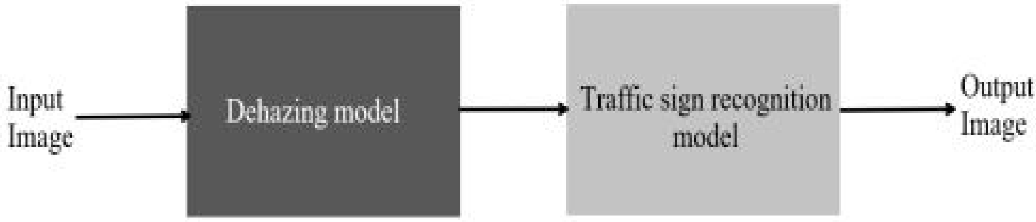



*Figure 21: TSR process in foggy weather*

When it's foggy outside, TSR has two processes.

1. To obtain a clear image free of fog, the defogging model is used to remove the fog from the image.
2. After dehazing the image in the previous phase, we retrieved it, trained our deep learning model on it, and then our model was able to identify the traffic signs in the image.

### 5.1.3. YOLOv5 Model for Traffic Signs Recognition

The fuzzy issue with traffic sign photos will cause the identification accuracy to drop in hazy weather, endangering the autopilot's safety standards. The influence of the traffic sign's size and angle will also have a negative impact on recognition accuracy. The model's identification speed must meet demanding criteria for the speed of real-time detection. We improve the YOLOv5 model, which has greater advantages in small object detection and considers both accuracy and speed, in order to more effectively recognize and detect traffic signs in hazy weather. In order to have a better outcome, we also enhance the YOLOv5 model for satellite photos, another auxiliary perspective of traffic sign detection.

### Development of YOLO Family

One of the top regression algorithms for deep learning object detection and a representative of single-stage algorithms is the You Only Look Once algorithm. In 2016, the YOLO algorithm was originally proposed. The YOLO algorithm was enhanced over the course of the next two years, and YOLOv2 and YOLOv3 were consecutively released. The functionality of YOLOv4, which was created in April 2020 with the help of more academics, is expected to greatly increase. YOLOv5 was made available on the GitHub platform more than a month after YOLOv4 and went live as YOLOv5.1.0 in June.

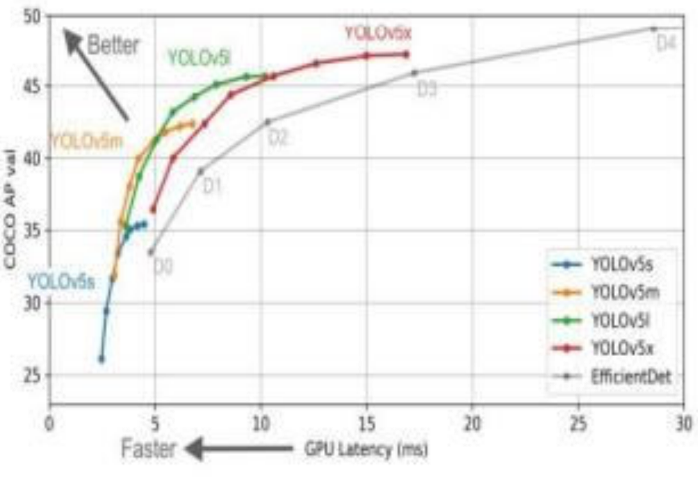



*Figure 22: comparison of YOLOv5 EfficientDet on coco dataset*

In YOLOv5, there are four net topologies with different depths and widths. On the COCO dataset, the chart above compares the map and reasoning speeds of YOLOv5 with EfficientDet. We can observe that YOLOv5 prioritizes both speed and accuracy, and its performance is superb.

### YOLOv5 Algorithm.

The skeleton structure of YOLOv5 and YOLOv4 are extremely similar. YOLOv5 has three aspects:

- Backbone
- Neck
- Prediction

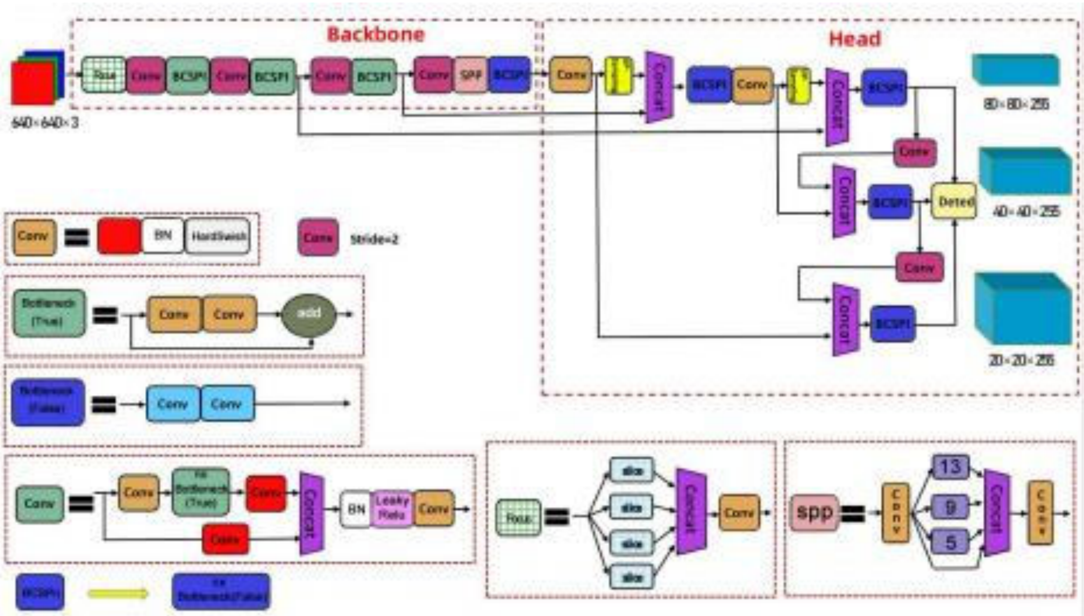



*Figure 23: Network structure of YOLOv5*

The PyTorch platform, which is broken down into submodules like the focus module, spatial pyramid pool SPP module, and two bottleneck modules, is used to build YOLOv5. Only the depth multiple and width multiple parameters are used to build networks of various sizes. We describe each module's philosophy and purpose.

Three components make up the input side:

- adaptive anchor frame computation,
- image size processing, and
- mosaic data improvement.

Both YOLOv5 and YOLOv4 use the mosaic approach to supplement their data. This approach fulfills the requirements for small object detection in this thesis and is appropriate for small object detection. The YOLO algorithm, which was used for training, requires that the size of the input image be converted into a fixed size.

The typical image size for this thesis is 460 x 460 x 3. We must first set the initial anchor frame before doing network training. To obtain the prediction frame and compare it to the actual frame, the network model is trained using fundamental anchors. The model parameters are updated in reverse order and changed iteratively in response to the difference.

The critical step in YOLOv3 and YOLOv4 is cut as indicated in the graphic below because they lack a focus. For instance, the focus structure is connected to the original picture 416 x 416 x 3, which is then transformed into a 208 x 208 x 12 feature map by slicing operations and and using a 32 convolution kernel operation, scaled to a 208 x 208 x 32 feature map..

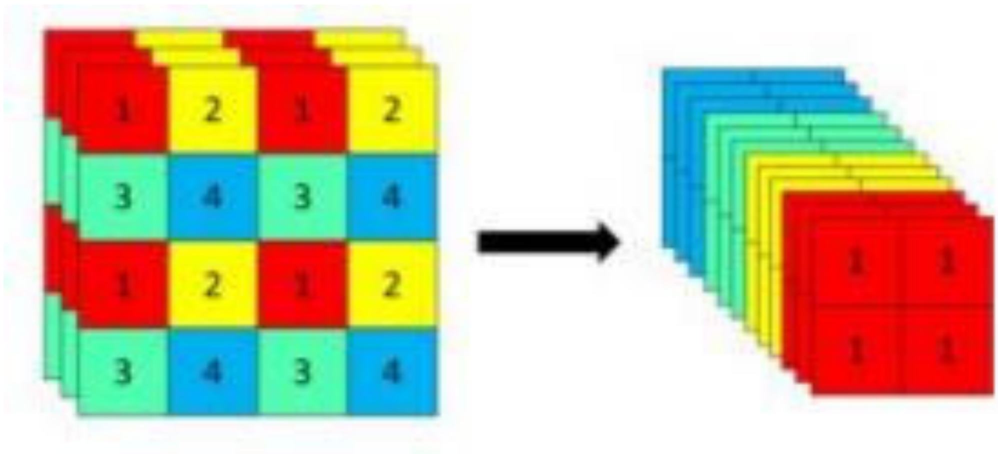



*Figure 24: slicing operation*

In the neck, the FPN+PAN (Feature Pyramid Network + Pixel Aggregation Network) structure is used. Since the FPN runs top to bottom, the optimal feature map before prediction can be created by perfectly combining high level features and low-level features. This guarantees that high-level, robust semantic features will pass while solely enhancing semantic information. Information about the user's location is not transmitted. Therefore, PAN uses a bottom-up feature pyramid to support FPN in order to improve semantic expression on various scales. The figure below illustrates the FPN + PAN structural layout.

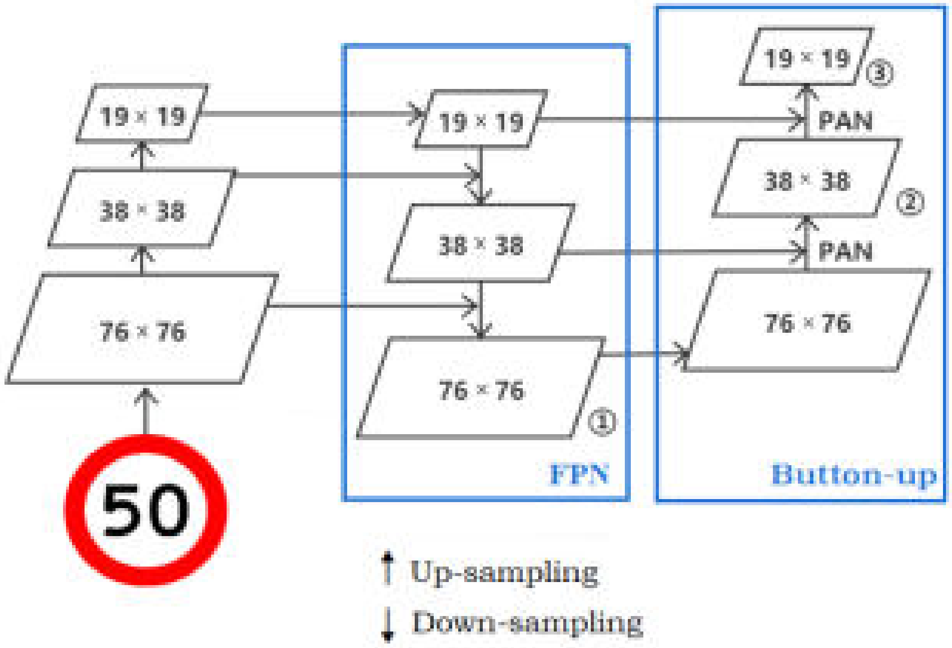



*Figure 25: FPN + PAN structure*

The loss function of bounding box regression and non-maximum suppression (NMS) are included in the prediction. Since YOLOv5 employs IoU Loss and the bounding box does not usually overlap, we use CIoU Loss in this article. Weighted NMS operation is used to select the best object frame from among the many frames that are screened during the object detection and prediction stage.

### Improved YOLOv5 Model.

Here, we contrast the experimental outcomes of the enhanced YOLOv5 with the enhanced Faster R-CNN. These two approaches' datasets, experimental setting, and hyperparameters are identical to those used in the earlier studies.

After processing, the tested video had 2,590 frames. Each frame is processed by YOLOv5 in 0.009 seconds, while each frame is processed by Faster R-CNN in 21 seconds. YOLOv5 processes each frame in 0.009 seconds, while Faster R-CNN processes each frame in 21 seconds. The recognition speed of YOLOv5 is faster while maintaining the same accuracy rate. YOLOv5 is a great fit for TSR because it is frequently used for real-time detection and has strong criteria for recognition speed. The two approaches' recognition results from the FRIDA dataset are shown below.

*Figure 26: the TSR results based on FRIDA dataset with Faster R-CNN*

*Figure 27: The TSR result based on FRIDA dataset with YOLOv5*

We can see from the photographs above that these two upgraded approaches successfully locate traffic signs in the background of intricate fog circumstances. The method of YOLOv5 is ideal for recognizing signs, but one of the identification results reveals that Faster R-CNN has two traffic signs that have not been detected. At the same time, thanks to the quicker identification time and smaller model size. The enhanced YOLOv5 performs better overall.

**Analysis**

In this experiment, we discover that Faster R-CNN receives a smaller loss score while earning a higher accuracy-related score. For instance, although the recognition speed is slow, the value obtained by mAP is essentially the same as that derived by precision and recall. YOLOv5 has a quick identification rate and good levels of recall and precision.

Let's use YOLOv5 as an example since it performs better. The next image displays how each indicator has changed as the number of iterations has increased. At this point, the bounding box value has decreased as the iteration has increased, while the mAP has increased. It demonstrates that as iteration times are increased, the detection result of the network in this thesis improves. The repetition of the network parameters also improves recall and precision. This suggests that as the number of iterations increases, so does the number of accurate positive samples find. In general, the YOLOv5 approach in this thesis improves and becomes more stable for TSR as the number of repetitions increases.

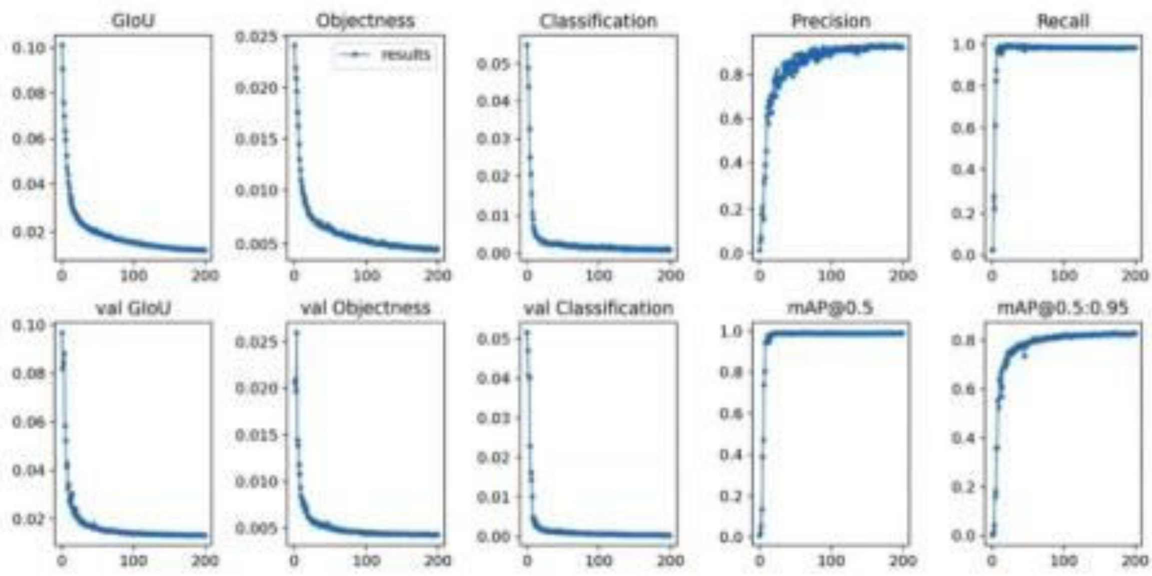


Figure 28: TSR results by using YOLOv5

The accuracy rate is plotted on the y-axis below, and the recall rate is plotted on the x-axis. Our model is working well, as can be seen by the PR curve's proximity to the upper right corner. As a result, TSR built on YOLOv5 has good performance and was developed well.

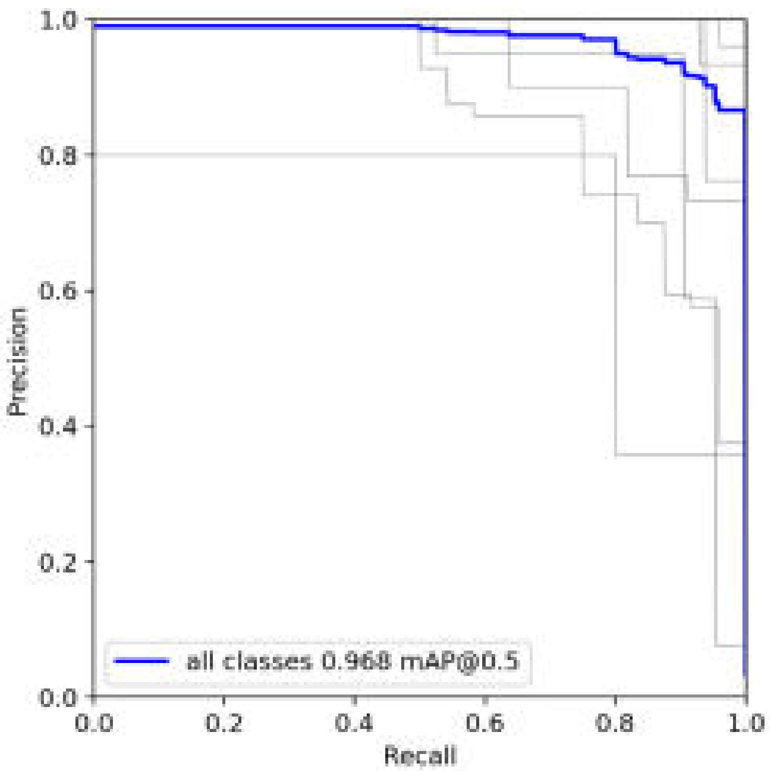



*Figure 29: The PR curve of YOLOv5*

*Figure 30:Image guided filtering is not applied, (b) by applying image guided filtering*

The experiment results utilizing YOLOv5 to identify items with and without directed picture filtering are shown in the above image. There are some situations where the model cannot recognize the traffic signs in the distance when there is dense fog since it only detects the signs in the photos without clearing the fog. After defogging, our model can recognize traffic signs from a distance using the original as a reference, which reduces the likelihood of missing and erroneous detection. This highlights the requirement for a picture defogging method, which strengthens our YOLOv5 model and produces superior results.

# 6. TESTING

**Test Plan and test strategy**

Test planning is required for creating a baseline plan with tasks and accomplishments to track the project's progress. It's described the test approach, goals, agenda, estimation, deliverables, and assets required to perform trying out for a software product. Furthermore, the functions to be tested are chosen based on the importance of the functions and the risks they pose to users. The test cases were then written under the available use cases. They were handled manually, and the results were recorded.

Making a test plan document has many advantages.

• Help customers, business managers, and other non-test team members understand the specifics of testing.

• Test Plan directs our thought. It appears to be a set of requirements that must be followed.

• The Management Team can evaluate the Test Plan and apply its key details, such as test estimation, test scope, and test strategy, to further projects.

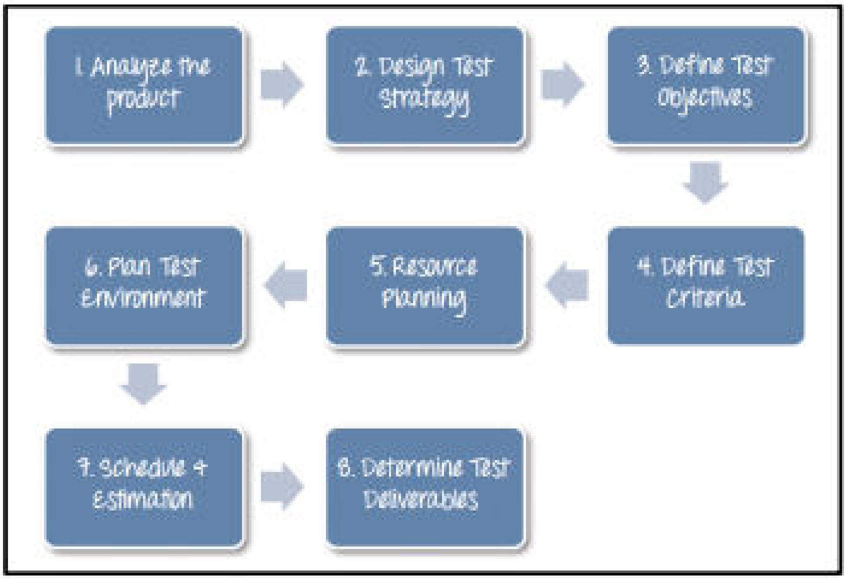



*Figure 31: Test planning steps*

# 7. RESULT AND DISCUSSIONS

The use of deep learning models is the most recent development in traffic sign identification. With CNN, deep learning models can approximate 98% accuracy. But there are following problems in deep learning models.

- The computational complexity is considerable and expensive resources are needed.
- Unable to run in parallel.
- It takes a long time to classify.

The proposed solution in this work has low accuracy when compared to deep learning models, however it offers the following benefits.

1. Low computational complexity allows for implementation on devices with limited resources.
2. Algorithm parallelization is possible.
3. Because only a straightforward distance computation procedure is required, categorization takes less time.

Even if the accuracy of the suggested solution is only 72%, it can be increased by building a sturdy mesh for feature extraction. Accuracy is enhanced in this manner. The proposed solution's ability to be implemented in parallel processing mode with less computing demand makes it look promising.

## 8. FUTURE SCOPE

- We are hoping to use raspberry pi4 module for the future implementation of the component.
- Tune the model to those data samples and get higher results.
- Integrating all the four components to a single unit.
- Our data set needs to be continually improved.
- Right now, all we have is a sunny dataset. We will gather data for the upcoming project in a variety of lighting situations, including cloudy and wet days.
- In TSR, contrast additional object recognition and detection techniques.
- Last but not least, we will use additional evaluation techniques to assess our model, allowing us to intuitively identify its flaws and improve it.

## 9. REQUIREMENTS

### 9.1. Functional Requirements

**Table 2*: functional requirements.***

| Requirement ID | The Requirement | Addressing the Requirement |
|---|---|---|
| 1 | Extract features | Frames will be separated from the video. The signs, whether indoor or outdoor, will then be found using the photographs. |
| 2 | Identify signage | Using real time images for detect the signage using by CNN model + Keras. |
| 3 | Identify closest signage. | Using real time images for detect the signage using by CNN model + Keras. |
| 4 | Make a decision for close signage. | According to obstacle distance give user a message about signage. |
| 5 | Generate a voice notification about signage. | After identifying the signage, convert them into voice command. |

## 9.2. Non-Functional Requirements

**Table 3*: non-functional requirements***

| Requirement ID | The non-functional Requirement |
|---|---|
| 1 | **Security**- This non-functional criterion makes sure that the entire device, or at least a piece of it, is secure against ransomware attacks and unauthorized access. |
| 2 | **Performance** - Describes how quickly a software system or one of its essential elements reacts to certain users' activities during a specific workload. |
| 3 | **Reliability** - This quality feature calculates the possibility that the gadget or a part of it will function properly under particular conditions for a predetermined period of time. It's usually represented as a percentage of the possibility. For instance, there is an 85% likelihood that the system won't go out during that month if the gadget has an 85% dependability under typical operating conditions. |

The non-functional comes under following category.

1. Product Prerequisites

2. Administrative Requirements

3. User specifications

4. Operationally Essential Requirements

**Product Requirements**

Portability: All types of traffic signs should be compatible with the designed software.

Correctness: It's important to recognize traffic signs accurately.

Ease of Use: The system is created in a way that makes it possible for users to engage with it easily.

Modularity: A modular system is required. It is possible to add or remove any module.

Robustness: The program is being created so that its overall performance is maximized, and that the user can anticipate results in a short amount of time that are almost relevant and accurate.

Usability: An eye-catching color should be used for the visual alert.

**Organizational Requirements**.

Process Standards: The majority of standard programming designers around the world construct their applications using IEEE models, which are the industry standard.

Design Methods: The project is being designed using a modular design methodology.

**Basic Operational Requirements**

Mission profile or situation: Recognizing the traffic sign is the project's goal.

Use situations: Any image of a traffic sign can be used with it.

Operational Life Cycle: The system can be used for classification as many times as necessary after being loaded with labelled images.

# 10. DESCRIPTION OF PERSONAL AND FACILITIES

Visual information abounds in our world, which a sighted person uses on a regular basis. Unfortunately, the visually challenged are denied access to such information, limiting their mobility in open spaces. To aid with this, we're working on a wearable device that can detect and recognize indicators in natural settings. The system's two primary components are sign detection and identification. The sign detector finds regions in an image that correlate to a sign using a conditional maximum entropy model. The sign recognizer looks for sign images in a database that match the hypothesized sign regions. The system assesses whether the hypothesized sign region is a part of a database sign or if the most likely sign is accurate. Our data sets cover a wide range of variables, such as lighting, orientation, and viewing angle variations. We offer an overview of the system as well as the performance of its two main components in this work, with a focus on the recognition phase. Unlike most past research, our approach is not restricted to recognizing a single type of indicator, such as text or traffic. In this use, a "sign" is just any physical device that displays data that could be beneficial to the blind. This system has a number of difficulties, the majority of which derive from the environment's high fluctuation. This variability may be influenced by a variety of factors, including the vast range of lighting conditions, different viewing angles, occlusion, and clutter, as well as the wide variety in text, symbolic structure, color, and shape that signs can have.

# 11. CONCLUSION

The Smart Navigation system has been primarily developed using the concept of Image Processing and also uses the concept of IoT for the smart cane implementation. The application has been developed with a user-friendly interface to effectively aid the visually impaired community. Due to multiple features such as face recognition coupled with emotion detection, walk-lane detection, road sign detection, obstacle detection, and pothole detection using the smart cane, the current system presents a unique opportunity for blind people to become more active in a cost-effective manner. The features detected through the camera and its image processing results are converted into audio/vocal commands and relayed to the user. Compared to the other systems in the market, this system consists of several unique features that cater to the multiple needs of blind people and encourage them to lead a more active social life. However, there are several shortcomings in the application. Therefore, overcoming these shortcomings and making this application more usable for the blind community is essential. Road damage identification is a significant issue, and numerous types of study have been conducted to overcome this difficulty. By using data from Sri Lankan roads to train the model, we applied a YOLO-based technique to deep learning to detect road deterioration. Smartphones are used to collect the dataset. We examined numerous dataset settings, and the results revealed some intriguing details. One pre-trained weight only requires 42MB of memory for our YOLOv5-based solution, and its inference performance is very quick. In the context of actual road damage identification issues, not only accuracy but also inference speed and, in some cases, frame rate are key considerations. As a result, this approach might be a good choice for blind people to detect road damage.

We have devised a computer-vision system that may help those who are blind or visually handicapped read signage inside buildings. This system is still a work in progress and has its limitations. It demands that the signs be precisely positioned on plain backgrounds that differ in color from the signs. Users must walk slowly—less than 3 kmph—and the mechanism only handles a small amount of tilt. Only a real-time system will allow visually impaired persons to adjust to it. The systems' main usability objective was to achieve this. We emphasized the significance of expressing the sign's position in relation to the user in order to help the user with pathfinding. The proposed hypothesis will be verified if the system can obtain a user rating score from a bigger survey of at least 80% and a reliable recognition rate of approximately 90% on average. Future studies will focus on developing methods for including text recognition and detecting more indoor objects. To ensure that the system operates in real-time despite the additional complexity, we will need to significantly increase its efficiency. In order to train a convolutional neural network and improve classification rates over the extracted signs, we might consider enhancing the dataset to make it adequately vast. Each sign will be enhanced to a huge number of samples for the augmentation by translations, small rotations, scaling, and even blurring. The application of CNNs might increase the number of supported indications. Hough transform on edges could be used to further enhance the shape detecting process. Then, using weaker classifiers, we might find hypothesized ROIs that are very likely to contain a sign. Hough Transform's shape detection accuracy will need to be compared to the approach employed in this paper. By reading outdoor tags and notice boards, text recognition could increase the dependability of such systems, as can be observed from the examined literature [22].

It is clear from this proposed system's present performance that people could benefit greatly from computer vision technologies who are visually impaired with their daily routines. These technologies can also help with other related applications, such automating industrial job flow.

# 12. REFERENCES


[1] https://www.iapb.org/learn/vision-atlas/

[2]Mugdha Surve. (2019). *Pranoti Mane | Head of department | Phd,researchgate*.Research Gate Retrieved October 21, 2022, from https://www.researchgate.net/profile/Pranoti-Mane

[3] http://www.cmta.lk/downloads/Motor%20Traffic%20Act%20-%20Road%20Signs.pdf

[4] G V S Murthy1,2, C Gilbert2 , E Schmidt3 , P G Mahipala4 , K M K Gamage5 , C Banagala5 , A P Abeydeera6, The Sri Lanka National Blindness, Visual Impairment and Disability Survey, ©2017 IEEE

[5] https://www.iapb.org/news/sri-lanka-releases-latest-blindness-survey-data/

[6] Electronic Travel Aids: New Directions for Research Working Group on Mobility Aids for the Visually Impaired and Blind. Committee on Vision, National Research Council, National Academy Press, Washington, D.C., 1986.

[7] Po-Cheng Shih, Chi-Yi Tsai and Chun-Fei Hsu, An Efficient Automatic Traffic Sign Detection and Recognition Method for Smartphone's, 2017 10th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI 2017), 978-1-5386-1937-7/17/$31.00 ©2017 IEEE

[8] Feng Lin, Yan Lai, Lan Lin, and Yuxin Yuan, A Traffic Sign Recognition Method Based on Deep Visual Feature, 2016 Progress In Electromagnetic Research Symposium (PIERS), Shanghai, China, 8– 11 August

[9] Ayoub Ellahyani, Mohamed El Ansari, Complementary features for traffic sign detection and recognition, 978-1-5090-4320-0/16/$31.00 ©2016 IEEE

[10] P. Hewlett. Mobile speak pocket, 2006. http://www.hp.com/hpinfo/abouthp/accessibility/ partnerships/mobilespeak.pdf.

[11] L. Tischler. Intel's read-aloud reader: Breakthrough for dyslexics, vision for the blind, 2010. http://www.fastcompany.com/1520074/ intels-read-aloud-reader-breakthrough-dyslexics-vision-blind.

[12] G. S. en Foo. Grocery shopping assistant for the blind/visually impaired, 2009. http://grozi.calit2.net/files/TIESGroZiSu09.pdf.

[13] Visual impairment and blindness, World Health Organization, 2014. http: //www.who.int/mediacentre/factsheets/fs282/en/. [accessed 29-May-2014].

[14] C. Rahman, W. Badawyand, and A. Radmanesh. A real time vehicle's license plate recognition system. 2001.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[16] T. Ubukata, M. Shibata, K. Terabayashi, A. Mora, T. Kawashita, G. Masuyama, and K. Umeda. Fast human detection combining range image segmentation and local feature based detection. In Pattern Recognition (ICPR), 2014 22nd International Conference on, pages 4281–4286. IEEE, 2014.

[17] F. Fanga, C. Fuhb, P. Yena, S. Cherngc, and S. Chend. An automatic road sign recognition system based on a computational model of human recognition processing. volume 96-2. Computer Vision and Image Understanding, 2011.

[18] R. Cuingnet, C. Rosso, M. Chupin, S. LehAl'ricy, ˜ D. Dormont, H. Benali, Y. Samson, and O. Colliot. Spatial regularization of {SVM} for the detection of diffusion alterations associated with stroke outcome. Medical Image Analysis, 15(5):729 – 737, 2011. Special Issue on the 2010 Conference on Medical Image Computing and Computer-Assisted Intervention.

[19] J. Jean and M.-J. Hsieh. Object recognition in household environment using fusion of shape and haar-like features. 10.9, 2011

[20] K. Juang, Li-Hong, M.-N. Wu, and Z.-Z. Weng. Object identification using mobile devices. Measurement, 51, 2014.

[21] M. Roberto, S. Kurniawan, and H. Bagherinia. Blind guidance using mobile computer vision: a usability study. 2010.

[22] X. Yang, T. Tian, C. Yi, and A. Arditi. Context-based indoor object detection as an aid to blind persons accessing unfamiliar environments. Proc. International conference on Multimedia Understanding. ACM, 2010.

[23] G. P. Fajarnes, L. Dunai, V. S. Praderas and I. Dunai, "CASBLiP- a new cognitive object detection and orientation system for impaired people," Proceedings of the 4th International Conference on Cognitive Systems, ETH Zurich, Switzerland, 2010.

[24] Manikanta K S, Phani T S S and Pravin A. 2018. Implementation and Design of Smart Blind Stick for Obstacle Detection and Navigation System International Journal of Engineering Science and Computing 8 (8).

[25] Amjed S. Al-Fahoum, Heba B. Al-Hmound and Ausaila A. Al-Fraihat. 2013. A Smart Infrared Microcontroller-Based Blind Guidance System Hindawi Publishing Corporation Active and Passive Components 2

[26] Ayat Nada, Samia Mashelley, Mahmoud A Fakhr and Ahmed F Seddik. 2015. Effective Fast Response Smart for Blind People Second International conference in Bioinformatics and environmental engineering.

[27] Vijayalakshmi Badre, Romba Chhabria, Tanmay Kandam and Kritika Karamchandani. 2016. Ultrasonic Blind Walking Stick with Voice Playback International Research Journal of Engineering and Technology 03 (04).

[28] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart terHaarRomeny, John B. Zimmerman, KarelZuiderveld, Adaptive histogram equalization and its variations, Computer Vision, Graphics, and Image Processing, Volume 39, Issue 3, 1987

[29] Abhishek&LakshmeshaK.N,"THINNING APPROACH IN DIGITAL IMAGEPROCESSING",International Journal of Latest Trends in Engineeringand Technology,Special Issue SACAIM 2017, pp. 326-330