EASYTALK: A TRANSLATOR FOR SRI LANKAN SIGN LANGUAGE

Final (Draft) Report

D. Manoj Kumar
IT17050272

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

Department of Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

September 2020

EASYTALK: A TRANSLATOR FOR SRI LANKAN SIGN LANGUAGE

D. Manoj Kumar
IT17050272

The dissertation was submitted in partial fulfilment of the requirements for the BSc Special Honors degree in Software Engineering

Department of Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

September 2020

DECLARATION

I declare that this is my work and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or institute higher learning and to the best of my 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 non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or another medium. I retain the right to use this content in whole or part in future works (such as article or books).

Name	Student ID	Signature
D. Manoj Kumar	IT17050272	

The above candidate is researching the undergraduate Dissertation under my supervision.

……………………………………..   ……………………..

Signature of the Supervisor: Date:

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the supervisor, Mr. S.M.B. Harshanath & the co-supervisor, Mr. J. Alosius for their continued support & advice for this research. They provided us with great insight & how to & what not to do when it comes to conducting professional research. Moreover, we would like to thank Dr Dharshana Kasthurirathna for his continuous encouragement & for providing useful information & tips regarding the research domain.

Above all, our extreme thanks go to our parents who actively encouraged us to excel in our academics & providing a great motivation to take this research as an opportunity to show our skills in academics.

ABSTRACT

A Language is a way of communicating thought between two parties. Effective communication exists when both parties actively engage in the session and respond. Just like speech and text that we use to communicate, Sign Language is a method used by the deaf and mute all over the world to communicate with each other and with other normal people. Not all of us are aware of the Sign Language alphabet and we do require a translation. When it comes to implementing a system to translate sign language to text and voice and vice versa, the process requires input in the form of a series of images and text respectively. This component mainly focusses on capturing a series of images of hands and processing them to identify hand gestures. A presenter who has ample knowledge in the English Alphabet[1]of the Sri Lankan Sign Language will be showing a series of sign gestures to the input device where the device will capture the full body of the presenter. The detection of hand gestures is done by applying Image Processing algorithms such as R-CNN[2]and with the aid of Computer Vision, the component will be isolating the hand gestures depicted by both hands of the presenter in real-time.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PWD – Persons With Disabilities

ASL – American Sign Language

LKR – Sri Lankan Rupee

2D – 2 Dimensional

ANN – Artificial Neural Network

ROI – Region of Interest

RCNN – Region-based Convolutional Neural Network

COVID – Corona Virus Disease

CV – Computer Vision

XML – eXtensible Markup Language

GUI – Graphical User Interface

SSL – Sri Lankan Sign Language

KB – Kilo Bytes

SDLC – Software Development Life Cycle

API – Application Programming Interface

IDE – Integrated Development Environment

OS – Operating System

VS – Visual Studio

SDK – Software Development Kit

AI – Artificial Intelligence

GoSL – Government of Sri Lanka

ICTA – Information Communication Technology Association

TTS – Text To Speech

Wi-Fi – Wireless Fidelity

JSON – JavaScript Object Notation

GPU – Graphical Processing Unit

REST – Representational State Transfer

1. INTRODUCTION

1.1. Background

In Sri Lanka, nearly 8.7% of the total population is considered to be PWDs [1]. They are affected by various forms of disabilities from hearing impaired to physical illnesses or disabilities. These people have always been a subject of special care among society. They get prioritized healthcare, state services and other concessionary services from the state and the fellow citizens [2]. The verbal and hearing-impaired people globally use Sign Language to communicate with each other. Each language in the world has its alphabets and syntax and likewise, different sign languages are used to communicate.

There isn't any universally spoken or a standardized, organized single sign language that is spoken globally. Instead, each country has its sign language which is based on the main language/s spoken in that region. The below list shows some of the sign languages spoken around the world [3]

- American Sign Language
- British Sign Language
- Indo-Pakistani Sign Language
- Sri Lankan Sign Language.
- Brazilian Sign Language
- New Zealand Sign Language, etc.

For each country, there is one or more sign language based on the languages spoken in those countries. There are about 135 such sign languages used. However, in most scenarios, the interchangeability is comparatively lower in between sign languages. For instance, a person communicating in British Sign Language might not understand any communications done in Sri Lankan Sign Language. This is a great barrier for local and overseas communications which are based on sign languages. In the next section, we will look at some of the sign languages spoken in various regions of the world.

American Sign Language (ASL)

This sign language has been spoken by the North American deaf and mute people. This sign language has the same properties as the other spoken languages. The origins of the language are unknown, but the usage of the sign language is been around for over 200 years. It is said to be originated from LSF – Langue des Signes Françoise (French Sign Language). The alphabet includes letters from A-Z each containing one hand sign for a letter and numbers from 0-9 [4]

British Sign Language (BSL)

This is the preferred language used by the hearing and verbally impaired people in the United Kingdom. According to the census of the United Kingdom and Wales, In 2011, over 15,000 people were using this language as the main language for communication [5]. Unlike ASL, this language uses spaces in between words. Using the sign language involves hand, body, head and face movements combined. The language origins go back to the 16th century. But at the beginning of the 20th century, the language split into many sign languages. [6]

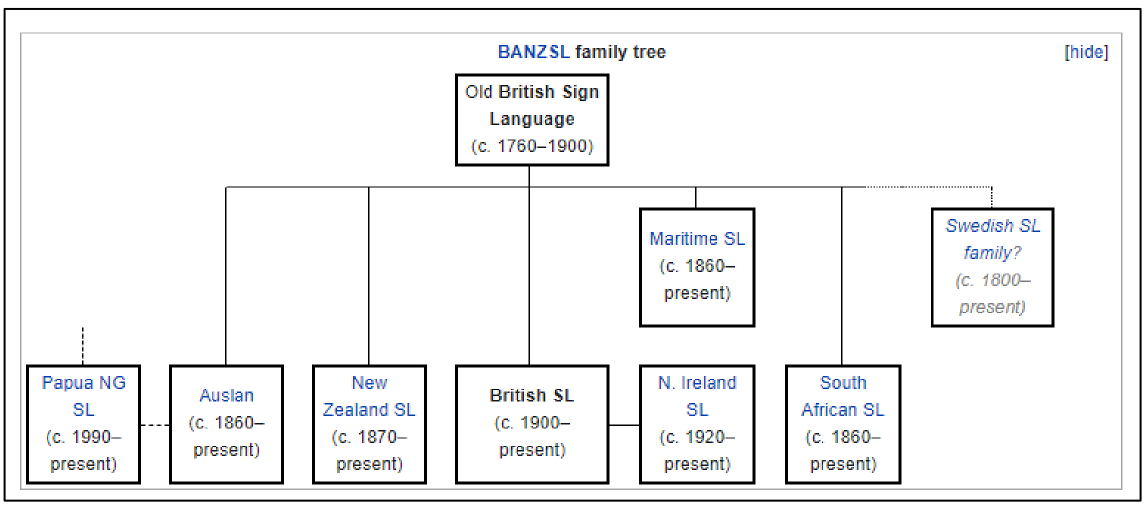


Figure 1-1: British Sign Language Family Tree

Indo-Pakistani Sign Language (IPSL)

This sign language is the most widely used in the south Asian region. The exact number of users is unknown but in Pakistan, around 0.24 million people which is 7.4% of their population [7]. Its usage has been so minimal in the Indian subcontinent and has not been privileged with official status. This means that the language cannot be spoken/taught in schools. [8]

Brazilian Sign Language (BSL)

The Brazilian Sign Language is spoken by the urban community of Brazil. Locally, the language is known as *Língua Brasileira de Sinais* (Portuguese). The language is in a well-organized form with multiple learning materials and instructive videos. The alphabet includes 44 individual hand signs. There are over 200,000 native speakers of the language in the country. [9]

New Zealand Sign Language (BSL)

This is the main language of the hearing-impaired community in New Zealand. In 2006, the language gained official status by law. This was to make the government services and information and services accessible to the hearing-impaired community. The language roots from the British Sign Language. The early British settlers of New Zealand brought the language with them when migrating to the country. Speaking of the language combines tiny lip movements and the two-headed manual alphabet. [10]

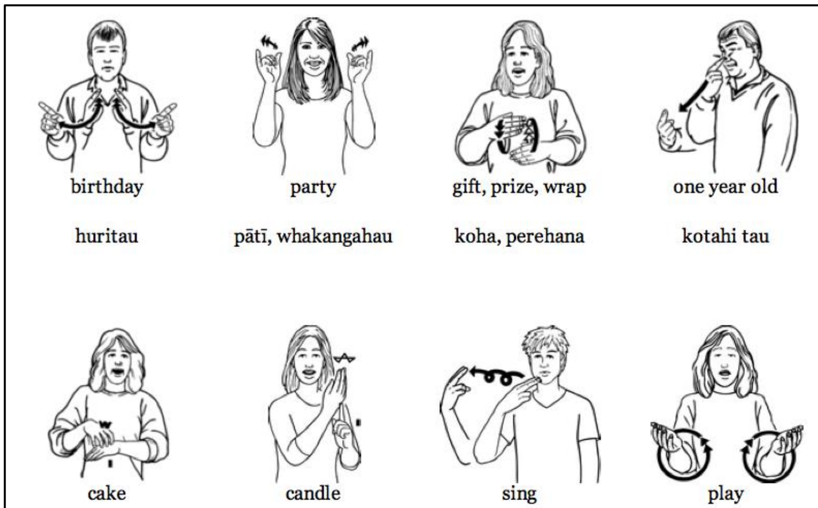


Figure 1-2: New Zealand Sign Language

In the Sri Lankan context, this research focusses on Sri Lankan Sign Language and its alphabet [11]. The signs consist of three different sets of alphabets for each language spoken (i.e. Sinhala, Tamil & English). The sign language also contains a set of common words/phrases that could be used in everyday activities. This includes phrases such as “Good Morning”, “Nice to meet you”, “Vanakkam”, “Ayubowan” etc.

With the introduction of free education in the 1950s, the literacy rate of the country has grown ever since and now it is at 96% [12]. This has led to a wide range of creative and innovative solutions for different problems faced by the society ever since. Numerous individuals have been shining in their respective field of engagement. However, such contribution from the verbal and hearing-impaired community is much less. They have contributed very little and this is mainly due to the barriers in communication. Although they get prioritized services, the proper funding for their education and career upliftment is not channelled and thus limiting their contribution. They are reluctant to participate in socio-cultural activities because they cannot speak linguistically or hear what someone’s saying. The only place we can see people using sign language is television news.

According to basic human rights, the needs and the rights of the verbal and hearing impaired should also be attended and fulfilled [2]. On the contrary, the practice is followed up to some extent. As mentioned above, the country has special schools, meal programs and government aids. These programs will keep them to lead a daily life. Their education levels are kept basic and in the rural areas, they’re reluctant to go to schools because of their disability. As a result, they get jobs that pay less wage [13]. Academically, their contribution is much less since the competition is with the people who can speak and hear. Because of these reasons, they are either being ignored or reluctant for active participation in academics beyond primary and secondary education [14].

The below table shows the population of hearing and verbally impaired community according to the census conducted by the Government of Sri Lanka in 2012.

Table 1.1: Population of hearing-impaired and inarticulate of Sri Lanka in 2012

Type of Difficulty	No. of Persons	% to Total Cases	No. of Persons not possible at all	% to Total Cases	Difficulty published in Census	% to Total Cases
Hearing	354,871	22.8	28,674	20.3	389,077	24.0
Communication	133,623	8.6	47,210	33.5	180,833	11.2

Problems faced by Hearing and Verbally impaired – Sri Lanka

The hearing and verbally impaired people of the country use hearing aids depending upon their level of hearing disability. This machine costs around LKR40,000. However, most of the people, mostly children cannot afford such equipment. Hence, hearing capability is lost forever. [15]

- Sign Language
    - Most of the people are not aware of how to speak or interpret sign language. This is the main barrier for education and team collaboration.
- Speech Therapy
    - Most Sri Lankan parents whose child is suffering from hearing or speech disability bring their children to school for the first time during their age of 6 or 7. In this age, it is hard for them to corporate and competes with other kids
    - Such kids must undergo speech therapy from a recognized therapist or a doctor before bringing them to school.
- Neglected
    - The hearing and verbal impaired community have many voices to raise for them. Despite all that, they are often considered to be a set of people with disabilities and ignored. They also have their minds on work and creativity to be put on the table. But often, the chances are declined, and they're neglected.

1.2. Literature Survey

According to the Census and Statistics report of 2012, the hearing and verbally impaired community comprise approximately 3% of the total population, who are above 05 years of age [16].

$$p = \frac{number\ of\ hearing\ and\ verbally\ impaired}{total\ population} * 100$$

$$p = \frac{(389077 + 180833)}{18615577} * 100$$

$$p = 3.061\%$$

There are various solutions developed by individuals to solve the problem of sign language translation. In this component, the main focus is on the feature extraction. In simple terms, the process of identifying hand signs among the most complex background. Below are some instances where feature extraction is done in similar systems/ solutions.

1.2.1 Sanwadha

Sanwadha is an instant messaging (IM) application mainly facilitated to help the hearing impaired. Here the application will get the text from the ordinary person in Sinhala and it converts to sign language [17]. This is achieved through the creation of 2D models for interpreting signs for a given text. It can be better understood by the below figure.

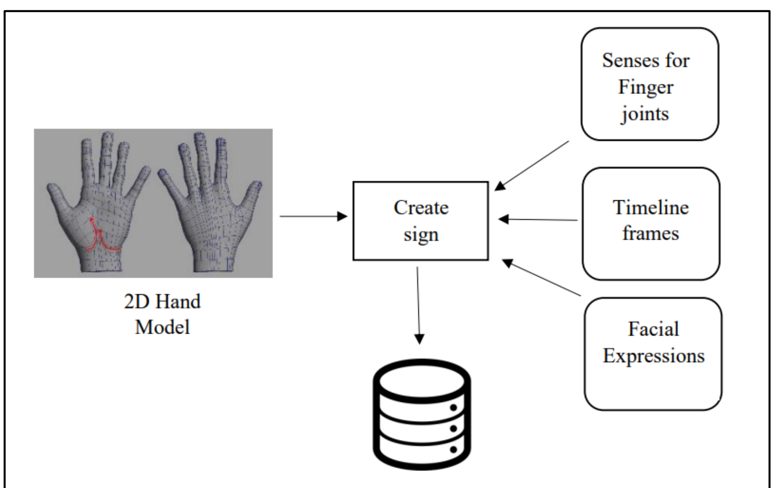


Figure 1-3: System Overview Diagram of Sanwadha Application

However, in the Sanwadha system, there is not any mechanism to capture or detect any live image and to translate it in real-time.

1.2.2 Sign Language Using Microsoft Kinect

This paper explains about identifying Sign Language using Microsoft Kinect [18] device. Using computer vision algorithms, they've developed a characteristic depth and motion profile for each sign language gesture. The generated feature matrix was trained using multi-class SVM classifier and the results are compared with existing techniques. [19]

Figure 1-4: Usage of Microsoft Kinect

1.2.3 Hand Sign Language Recognition for Bangala Alphabet based on Freeman Chain Code and ANN

Using ANN, a hand sign recognition framework has been proposed in this work. First, the image from the image is normalized and the skin area is extracted based corresponding to human skin color. The hand sign area is transformed into a binary-based image and the gaps in that are filled through the morphological operations. Then, the boundary edge of the hand sign area is extracted through the canny edge detector and extracts the hand sign region of interest (ROI). Finally, all the features are extracted from the hand sign ROI using Freeman Chain Code (FCC). The ANN is used for training and classifies the hand sign images. [20]

Unlike using ANN, in this research, however, it is about data acquisition and feature extraction using R-CNN and Computer Vision. Also, we're optimizing this for low- resolution images which would enable this system being used in any device of low camera power/web cameras

1.3. Research Gap

There is a communication gap between the hearing and speech impaired and other normal people. Most of the time, normal people use an interpreter or a sign language translator when they communicate with the hearing impaired and the mute. The real problem occurs when there is no interpreter, or a translator is available to find. In these situations, the normal people refrain from initiating a communication session with the hearing impaired and mute. This also leads to the disabled not being actively participating in social activities such as education, ceremonies etc. [15].

In countries like America and India, there are inventions to overcome this problem, but those solutions are very much limited to their country's context and languages. This means that there is no proper methodology or an application to overcome this problem in Sri Lanka. The main reason is that in Sri Lanka, the country has a different set of Sign Language alphabets in all three languages for the deaf and mute to communicate.

The second reason is that Sri Lanka is a 3rd stage developing country. Due to this title, the government cannot put much effort on developing a high technological system targeting the deaf and mute within the country as building a system of such complexity requires high labour and technical expertise which costs a lot. Therefore, my component will be acting as one of four pillars which would solve this problem

1.4. Objectives

1.4.A. Main Objective

The main objective of this research component is to detect hand gestures using Faster RCNN. The model will be trained on top of TensorFlow models [21] using Faster RCNN configurations. The model is trained through images taken by a low-resolution camera such as a laptop webcam and the detected output will be presented with a bounding box to the user.

1.4.B. Specific Objectives

Capturing Image

- The first objective of this research component is to capture the image using a low-resolution camera and creating a dataset. The data collection has to be done manually because of the COVID-19 pandemic [22]. Using web camera to capture images and train models would facilitate the system being used as a web application which could use the web camera of the respective device.

Detecting Hand Gestures

- With a model that is trained on top of TensorFlow models and Faster RCNN configuration, the system would identify hand gestures in real-time and show them in the bounding box with an assist text on how accurately the hand is detected.
- The system should detect hands irrespective of hand size, skin color, background etc. for optimal image recognition.

In Real-Time

- The above objectives will be accomplished in real-time. This is mainly because the next phase of the application will be a mobile application which will be portable and could be used anywhere. So, lags in detection and translation are of utmost importance to the system

2. METHODOLOGY

In this chapter, the focus is on how the research component – Hand Gesture Detection is implemented and tested. The implementations include how the data is a gathered, initial insights on how the methodology was thought to be implemented, how it was implemented and to a working component.

This part of the solution (research component) presents a hand sign detector to be used to classify different hand signs.

2.1 Requirements

- Should be able to identify hand gestures in low-resolution camera input.
    - The system should be able to detect and identify hands from a stream of web camera input. This should produce a bounding box around the detected hand and the accuracy of the detection.
- All the intended operations of the component should happen in real-time regardless of the device nature.
    - This aspect should consider how fast this part of the solution has performed its task. There are areas where internet connectivity is low and computers/ mobiles with less processing power. So, the component should be able to work fine given such situations

2.2 System Overview

To implement the system using Faster RCNN, we must first understand the concept of CV. CV is used for many applications and most of them include image classification. Features such as edge detection, Object Detection, Object segmentation and Semantic Segmentation are some of the key features of CV. In this research component, the main objective is to identify hands (hand gestures) to take imagery and send it to the next component for classification.

In brief, the model is trained on top of TensorFlow models with Faster RCNN configuration. The generated model is an inference graph and will be stored locally in the running system. Once given a stream of images through web camera’s live feed, the model will detect hands and if the hand matches one of the inputs, it will show a bounding box over the detected hand in real-time. The bounding box will show what the object is wrapped around and how accurately it has detected. Initially, the datasets are created using hand signs by the members.

The datasets are created in a way that for each English SSL, 05 images are taken manually. Those image sets were taken in the laptop camera which is of 0.3MP. Each of these images is of approximately 300KB – 600 KB depending upon the color content of each image. After that, they’re labelled in *PASCAL* format. After generating the labels for each image and the label are exported as .xml files. Using a small piece of code, the individual XML files are aggregated into two files both for training and testing respectively. Those training and testing label sets are then used by the TensorFlow trainer to train the model. It is a long process and requires time to train on average computers. It is possible to see how good or bad the model is being trained in tensorboard. Tensorboard is a web-based GUI where it is possible to have all insights about the training process in real-time.

Once the training process is completed and the loss level goes below 0.2, it is recommended to terminate the trainer. It saves an inference graph (TensorFlow model) which it refers while detecting objects. Once it is done, the next step is to take the Faster RCNN configuration from the TensorFlow model garden and edit the label information according to the project’s needs.

Here, it is only one class to be labelled at(hand). So, it is comparatively easier to train a model in this scenario rather than training a model for multi-class detection/classification purposes.

The below Figure 2.1 shows the system overview.

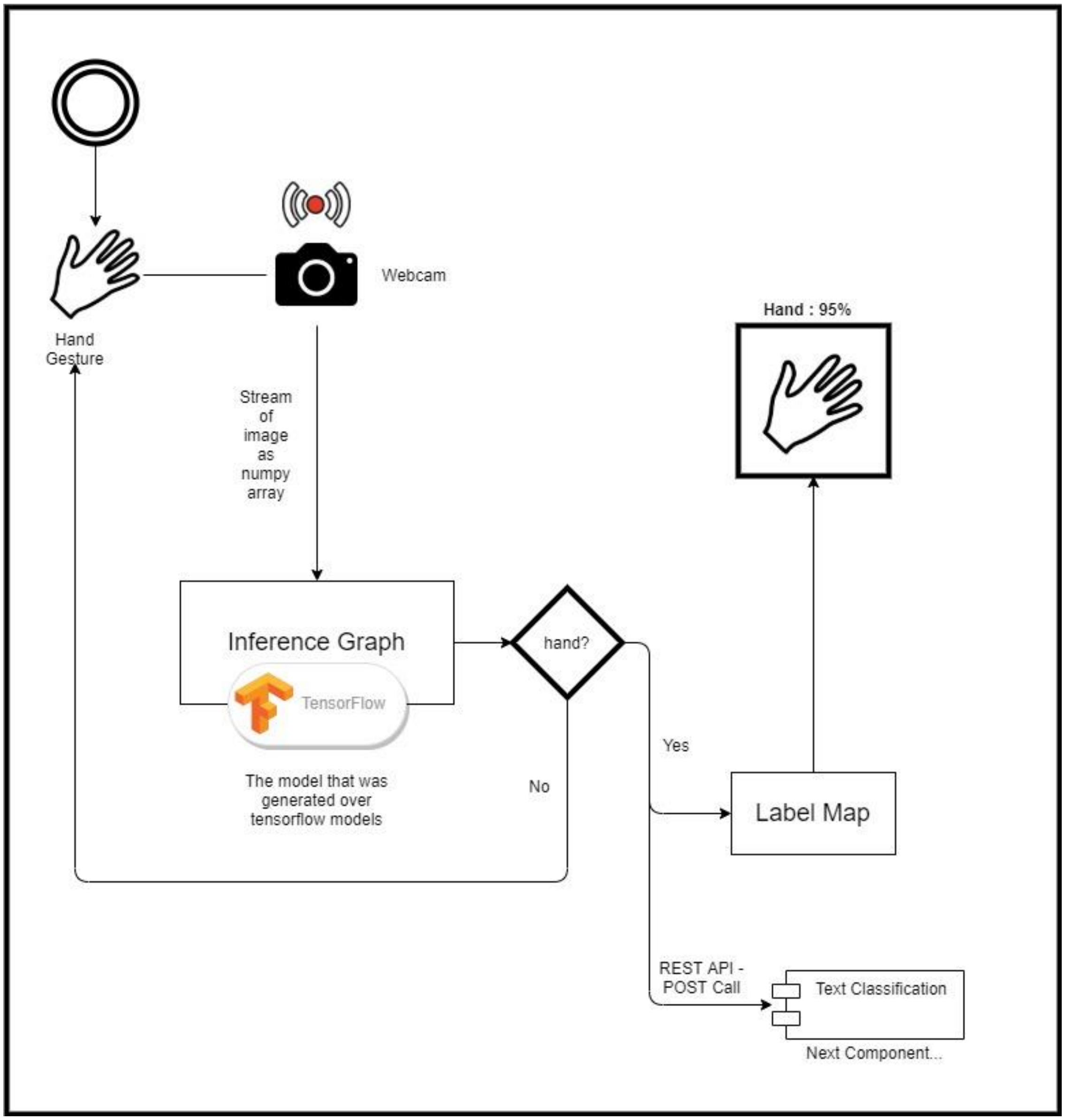


Figure 2-1: System Overview Diagram for Data Acquisition Model

2.3 Flow of Project

2.3.1 Feasibility Study

In this section, factors such as the estimation of the probability of the project, the problems and limitations faced, and other key constraints faced are discussed. It also discusses the technical and functional limitations that were hindering the project on its course.

Work Breakdown Structure (Microsoft Teams)

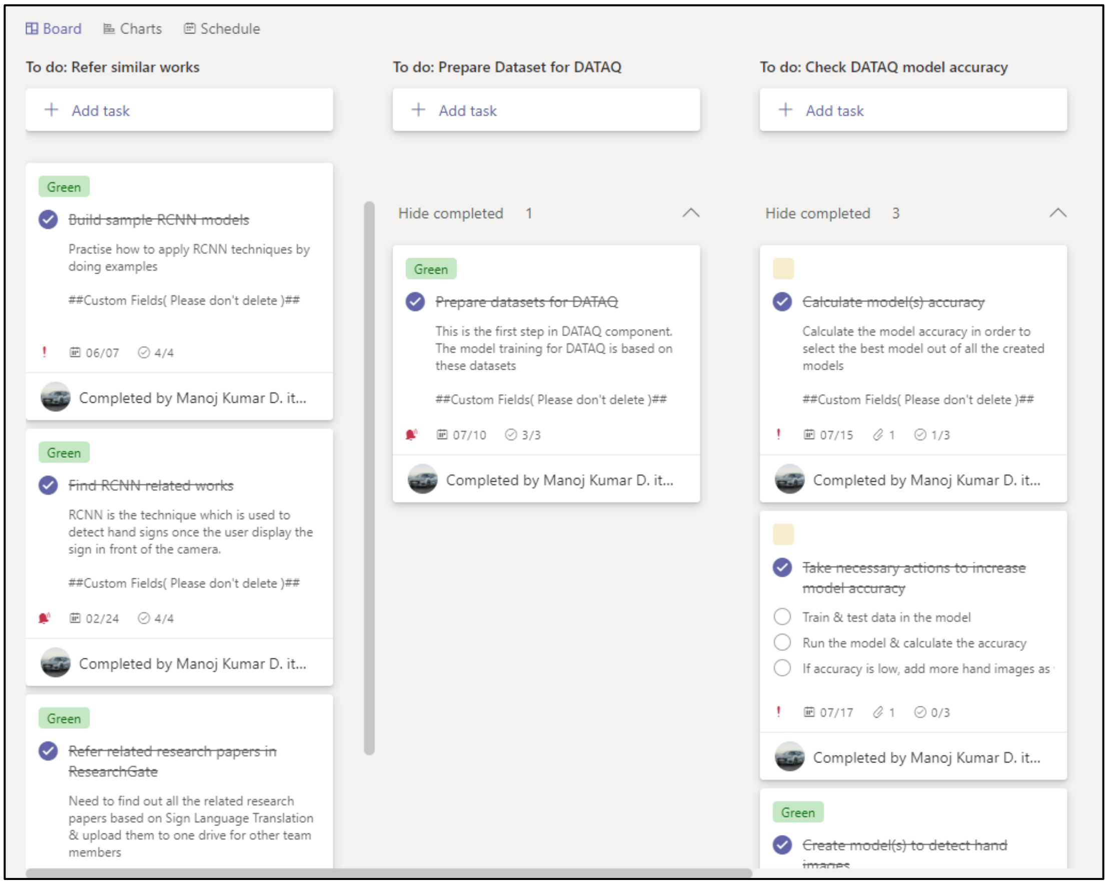


Figure 2-2: Work Breakdown Chart part I

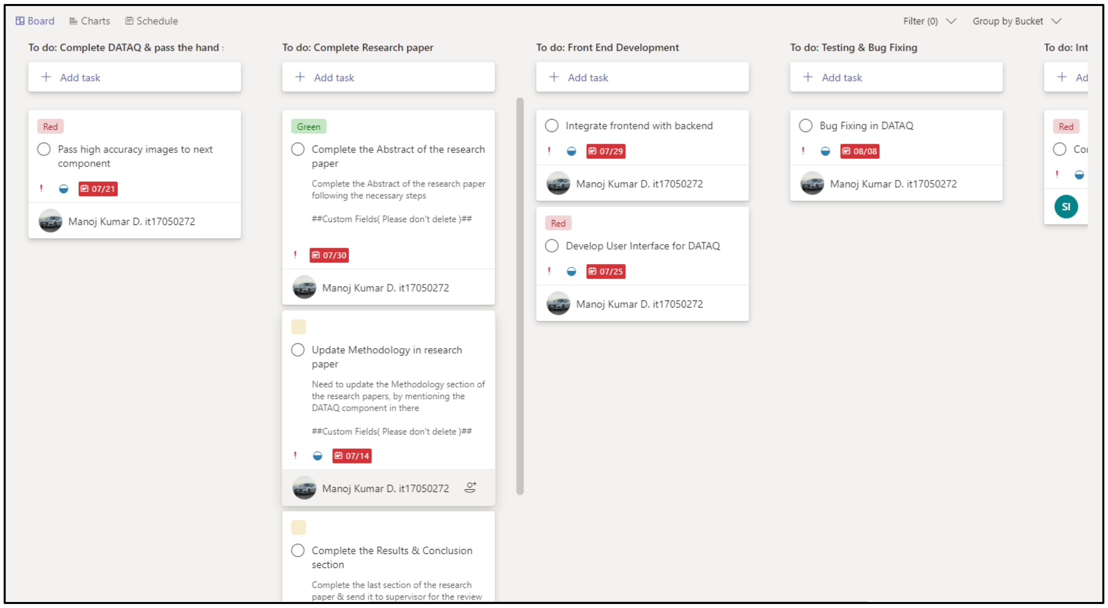


Figure 2-3: Work Breakdown Chart part II

Component Timeline

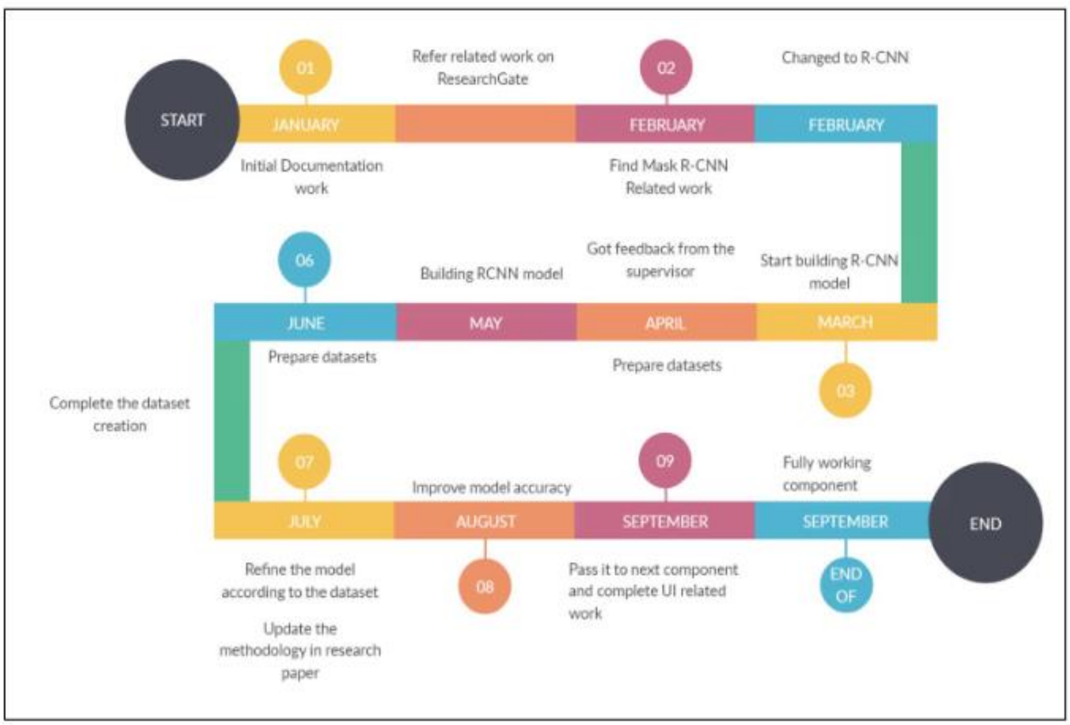


Figure 2-4: Component Timeline

Identification of the problem domain

In this competitive world where everyone tries to put their name on the top, the hearing and verbally impaired population of the country seemed to be left behind. The modern world creates so many platforms that everyone could shine in their desired domains. The accessibility of those platforms to the hearing and verbal impaired community of Sri Lanka is much less. In some scenarios, a family might have an only child where that child cannot hear or speak. In such cases, the future of the family is at risk because of the financial and academical incapability of the child. So, there is a need for the hearing and verbally impaired for their ideologies to be heard and spread.

Although they cannot speak or hear, their mind is at work. Without access to voice or audible sounds, their mind is in constant search for something new to be learnt. The internet has some great incidents and videos where such people get a chance to speak/ hear for the first time of their life and we can see how they're excited. The EasyTalk system will provide an interface for the hearing and verbal impaired community where they can use in day to day activities and also in academic applications. This can be a starter in their career to become a successful person in their lives. Using this system, they can talk with other people without hassle and assistance from anyone.

Data Collection

In the beginning stage of the project, it was decided to visit a deaf & dumb school in Batticaloa to collect useful data and to collect images for datasets. But due to COVID-19 pandemic [22], the Government of Sri Lanka implemented an island-wide curfew for months. Due to that, it was impossible to create and collect data for this project. However, after studying the basic alphabets of the SSL and created our datasets. The dataset included the hand of four different people. This was to recognize different sizes of hands and to identify hands despite skin color.

Financial Viability

The economic expediency estimation signifies to administer the complimentary monetary compensation the proposed application will contribute. Impersonating a cost/benefits interpretation based on meaningful experiences is similar to Fundamental Technologies, Transportation, and other resources and assistance. Subsequent studying everything the investments against the advantages, the profits of the application overweighed the spending. Accordingly, the purposed project container is estimated as nearly a financially viable scheme.

Technical Possibility

The Technical Hope emphasizes obtaining a considerate of the contemporary technological sources and their soundness to the thought application's feasible requirements. It determines whether the construction vessel dispatches the fashionable system requirements and can be contained to confirm that the idea is technically possible. In advanced application, essentially concentrated on Interaction with chat application, related APIs and Source codes, Machine Learning Algorithms, Compression, furthermore Extraction Algorithms throughout the Technical judging.

Operational Probability

Operational Expectation refers to how the application oversees how it meets the documented conditions in the collection and analysis required in the SDLC (Software Development Life Cycle). In the event of accurate identification of the dilemma, I proposed an explanation, having previously collected the conditions by which it was sufficient for the interpreter through conversations and inquiries. As a result, the current complex recommended solution approvals ensure that an expansion is necessary. Accordingly, he concluded that the intended application was a functionally available extension.

2.4 Resources Needed

2.4.1 Software Boundaries

- Visual Studio Code
  - Visual Studio Code is a great IDE for editing code and it is available on Windows, Linux and Mac OS. It has built-in support for JavaScript and TypeScript. Visual

Studio Code (also known as VS Code), could be customizable using extensions available. It is possible to write and edit code in any programming language.

- Initially, the idea was to use Jupyter Notebook. However, it was running on a browser-based IDE and could not be integrated with the web application. It is also difficult to build APIs using it and thus it was discarded from use.
- In this project, VS Code was used to write python code and implement backend logic.

Figure 2-5: Visual Studio Code Logo

Below are some screenshots of using VS Code in this project.

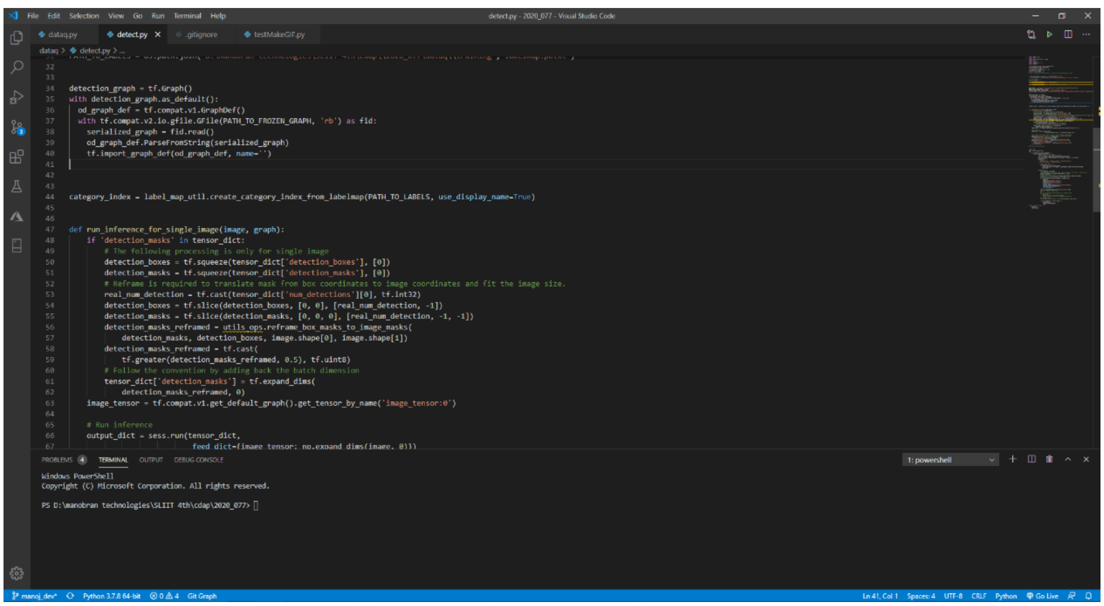


Figure 2-6: VS Code Coding - I

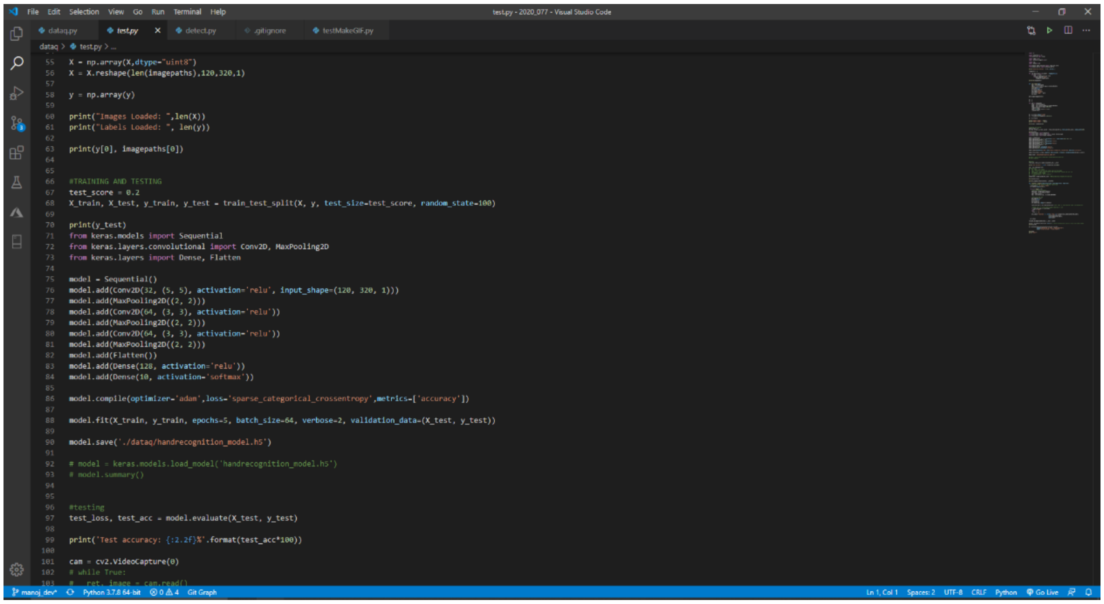


Figure 2-7: VS Code Coding - II

To write python logic in VS Code, it is necessary to install Python on the operating system and install Python Plugin [23].

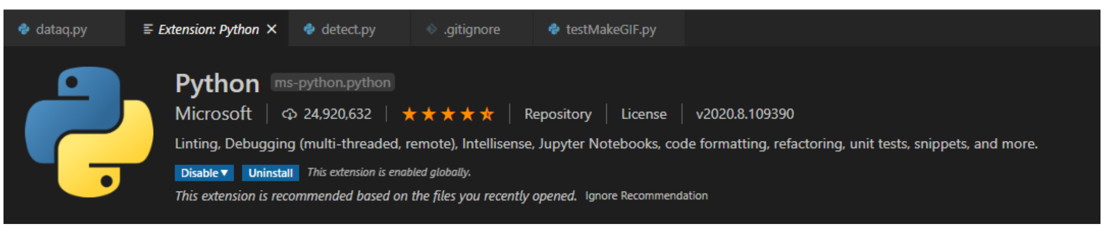


Figure 2-8: Python Plugin in the VS Code

- TensorFlow Models – TensorFlow Model Garden
    - The TensorFlow model garden contains some great solutions for modelling and users who use it for development. The model library contains various types and examples of TensorFlow models

Figure 2-9: TensorFlow Logo

- LabelBox – Labelling Tool.
    - Labelbox is an enterprise-grade data training solution with fast AI-enabled labelling software, automation of labelling, human resources, data management, development API & extensibility SDK.

    - In the initial stages of the project, LabelBox was used to label the dataset which was created. It had some amazing features. Although, the labels created through LabelBox were only capable of training images, not for a live stream input.

Figure 2-10: LabelBox Icon

Screenshots of LabelBox usage in this project.

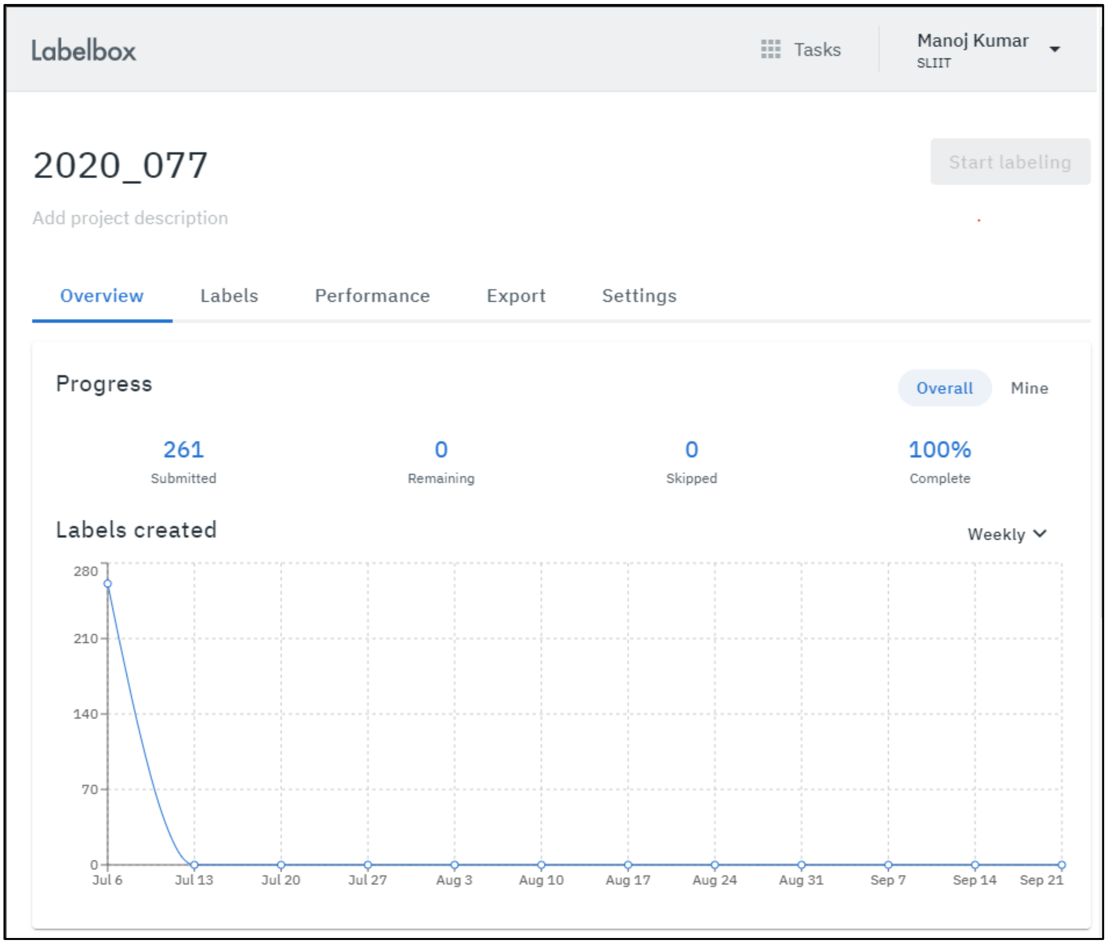


Figure 2-11: LabelBox Implementations for the Component

2.5 Commercialization Aspects of the Product

The EasyTalk application would be beneficial for the following set of people:

- Hearing Impaired
- Verbally Impaired
- Hearing and verbal impaired
- Anyone who wants to talk or learn sign languages.

The unavailability of a system being able to translate SSL into text and voice is a major point of marketing and pivotal in business strategy. Using the application, the users can

Translate Sign Languages to Text, Get the Sign corresponding to a given text, Get the output as text or voice in real-time.

2.5.1 Business Pitch

The user base of this application has no demographical or geographical boundaries or limitations. In the second stage of the application, the user will be able to use it anywhere and anytime. Because in the next iteration, the mobile app will be more helpful translating SSL to text on the run.

Sources of fund for commercialization include:

- Funds via the approval of GoSL and ICTA
- Direct Aid Program for Sri Lanka & Maldives
- Approval of funds from NSPD

The application can be packaged and delivered in a way that there will be a set of services where the user could use them for free up to a limit. If the user wants to get the full experience of EasyTalk, they could pay on increments meaning that the user could unlock certain features after paying price for each.

This is how the EasyTalk application will be priced.

- Free Access (available to all)
  - This level of subscription offers Sign Language to Text translation in all three languages where the output is obtained in text and voice.
- Paid Access – Level 01 – Learner
  - In this level of subscription, the user can get audio (TTS) output along with text and text output in all three languages. The user also gets access to Text-To-SSL where they could enter a text message and get an SSL sign language GIF as the response. All the unlocked features will be available for 12 months
  - Price: 500LKR per 100 translations. Priced yearly

- Paid Access – Level 02 – Pro
    - This is the topmost access level of the EasyTalk application. This unlocks all the features of the application. Apart from the features mentioned in the previous access levels, the user can create their account and sync their translations with devices. They can also save and translate the words that are often used and create lists of words. They also get text and voice output in all three languages. All these features will be free forever
    - Price: 5000LKR

Since the application will be mostly used by deaf and dumb schools and homes, it is thought to license the product for the organization and the organization can pay for the actual users (just like Microsoft 365). This can also result in a situation where the government could fund part of the licensing cost of the organization.

The app will also be available on the Web, Google Play Store and Apple App Store where normal users can also download and use it. They also can unlock all the features of the application at the desired level.

The EasyTalk application will also be available as extensions to the leading chat apps like Facebook Messenger, Microsoft Teams etc. This will attract foreign investments and it is possible to scope for the international market via enabling support for all sign languages used in the world.

- Google AdSense

Since the solution is a mobile and Web-based, it is possible to use ads in the grey areas of the UI which would generate income. Apart from Pro – Subscribed users, all other users will be seeing ads on the application.

Figure 2-12: Google AdSense Logo

2.6 Testing & Implementation

2.6.1 Testing

The testing for this component is to be done in two stages, Unit & integration tests. This component should detect, and transfer hand signs and the flow should be flawless.

Unit Testing

Unit testing is done for a single software component to check whether the intended task is done or not. Here, it takes the scope of this component. The hand was detected successfully with varying accuracies. The testing strategy is to show different hand signs and it should detect the appropriate ones that are instructed by the model to be detected.

Here the camera is given with 3 varying inputs whether the accuracy is low or high. For each image detected the relevant console output is shown.

Input #1:

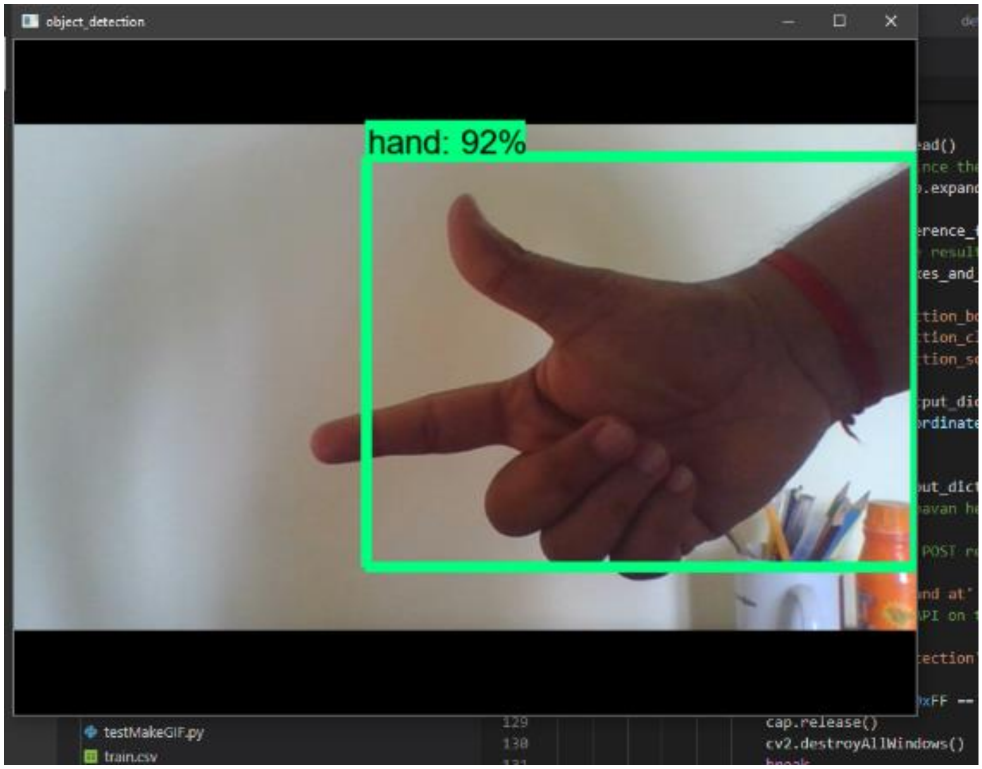


Figure 2-13: Hand Detection Input 1

Console Output for Input #1:

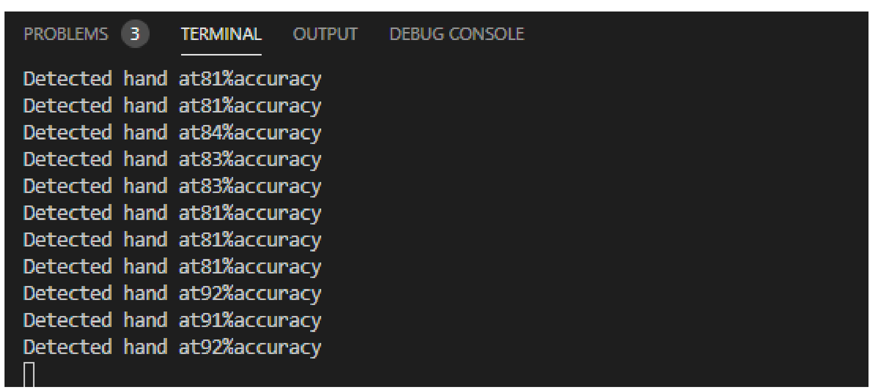


Figure 2-14: Accuracy Calculation for Input 1

Input #2:

This input is to detect the hand among a complex background. As seen in the below diagram, there is a mug, glue bottle, Wi-Fi router and a notebook. But the component filters out the hand sign among all of them.

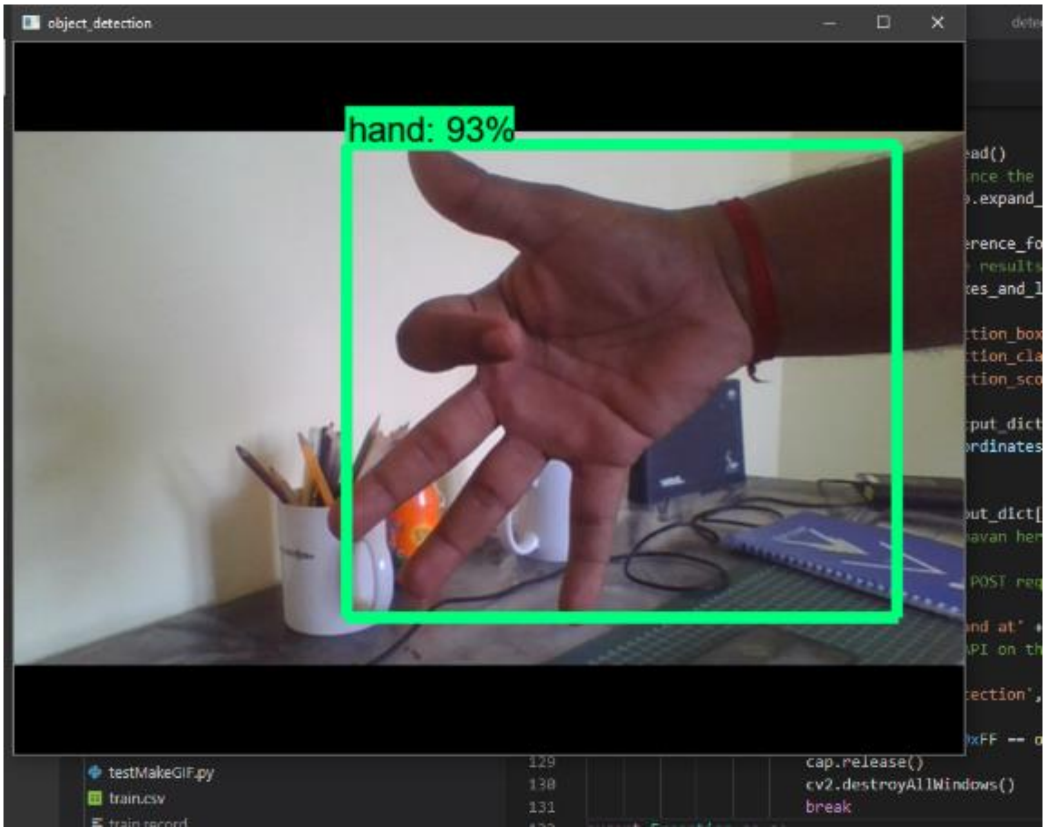


Figure 2-15: Hand Detection Input 2

Console Output for Input #2:

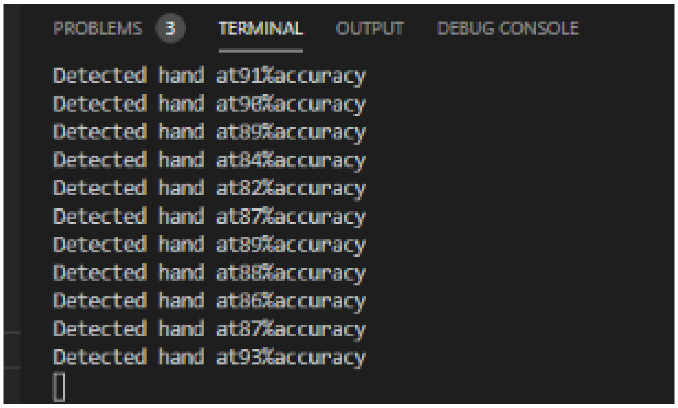


Figure 2-16: Accuracy Calculation for Input 2

Input #3:

This input was taken from another person's hand whose palm is slightly bigger than the previous one used.

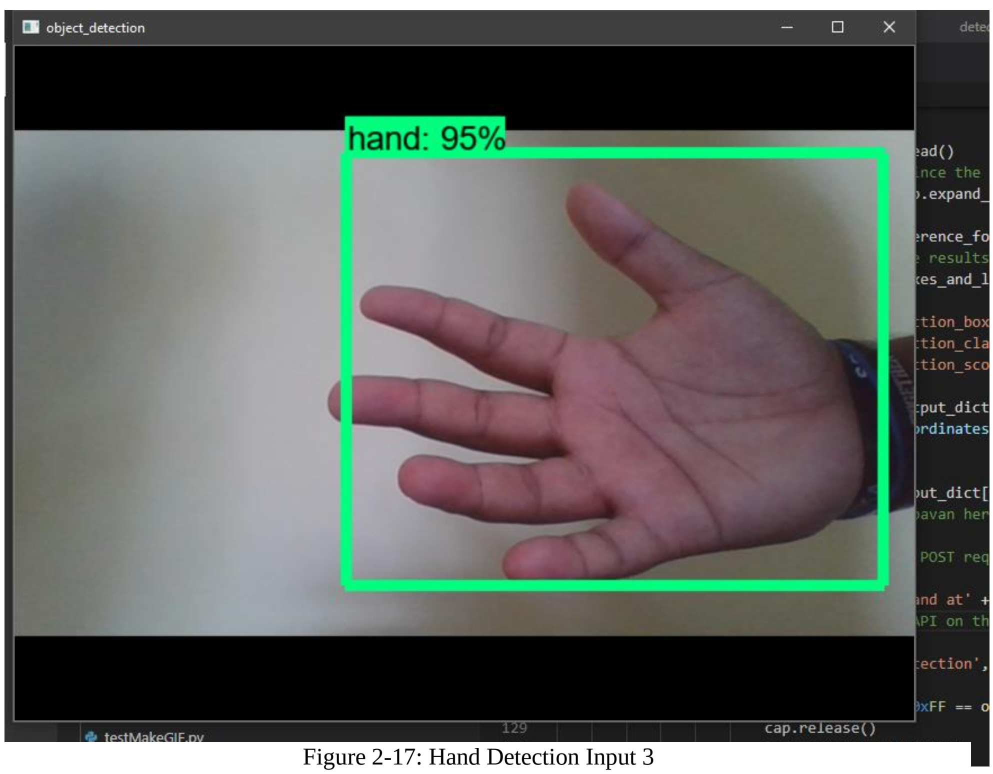


Figure 2-17: Hand Detection Input 3

Console Output for Input #3:

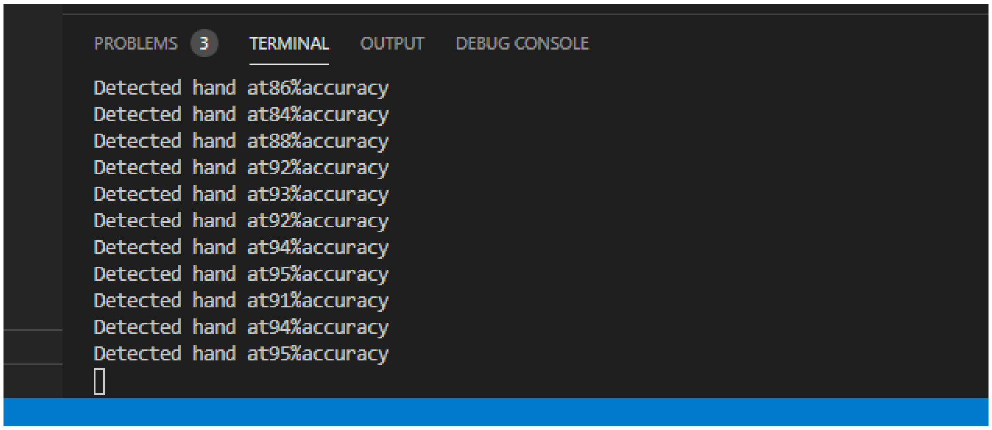


Figure 2-18: Accuracy Calculation for Input 3

Integration Testing

To integrate the component with the next component and with the whole system, it was decided to send the hand images through a REST API based POST call to the next component for text classification.

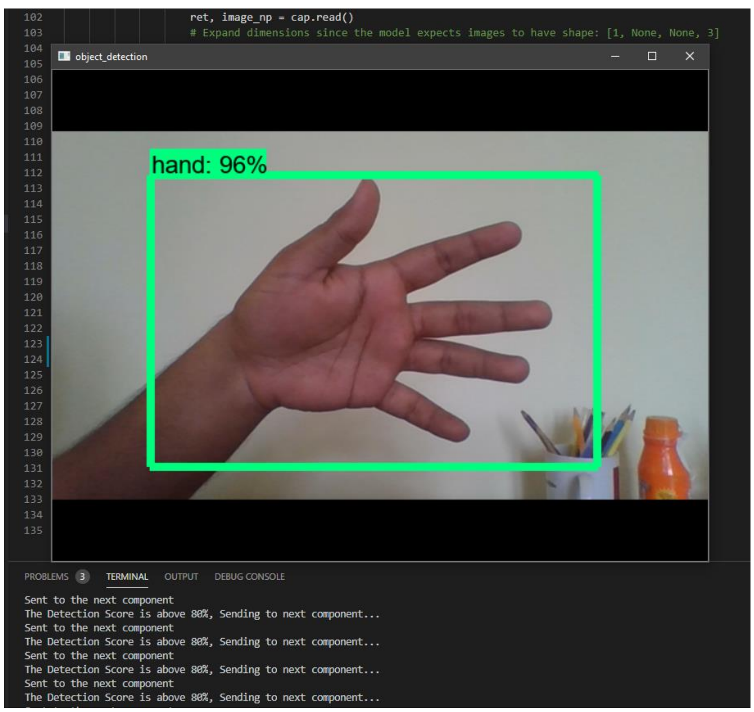


Figure 2-19: Integration Testing

3 RESULTS & DISCUSSION

3.1 Results

In this section, it is discussed about the results got detecting hand signs in EasyTalk web application. To get more understanding about the results, this chapter divided into two main subcategories namely Survey Results and Component Results. The survey results section defines the results got from the survey which was conducted on behalf of gathering requirements to implement the application. The Component Results section defines the results got from the execution of the research component, Data acquisition or hand detection.

3.1.1 Survey Results

In the preliminary stages of the application, a survey was conducted to gather user requirements and perspectives of the people on the upcoming application.

Since the “Data Acquisition” component mainly aims at both Sri Lankan ordinary community as well as hearing-impaired and inarticulate community, the survey was distributed among both communities. Among them, 50 has participated in the survey and this section describes the results got from them.

Among the results got from the survey, the following results were highlighted on behalf of implementing the component

- The most flexible language in daily lives
- The level of experience in SSL among Sri Lankan ordinary people
- The reasons for not having interesting on using SSL as ordinary people
- The reasons for not having interesting on using verbal languages as a hearing-impaired or inarticulate person
- The percentage of the familiarity of sign language translators
- The most preferable sign language translator type
- The most preferred language to handle the translator
- The most preferable way to an individual to capture an image from the translator

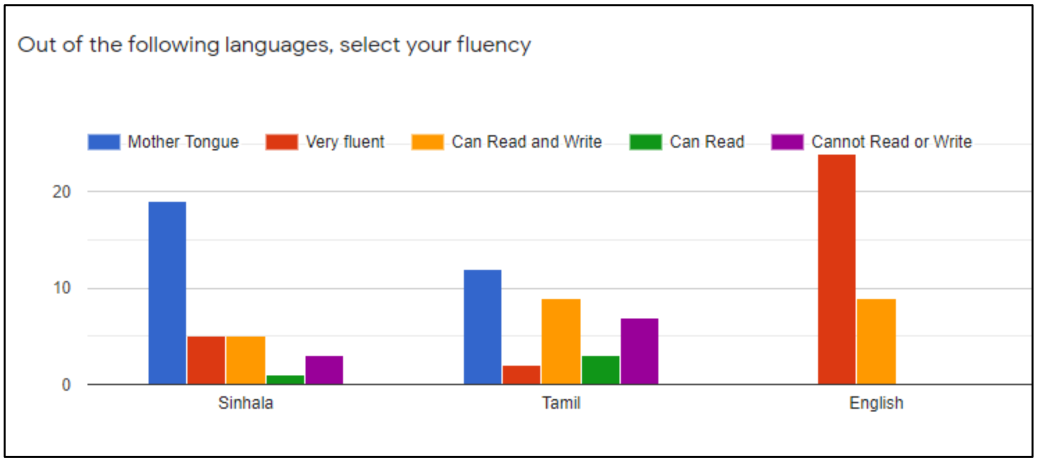


Figure 3-2: Verbal Language Fluency among Sri Lankan Ordinary Community

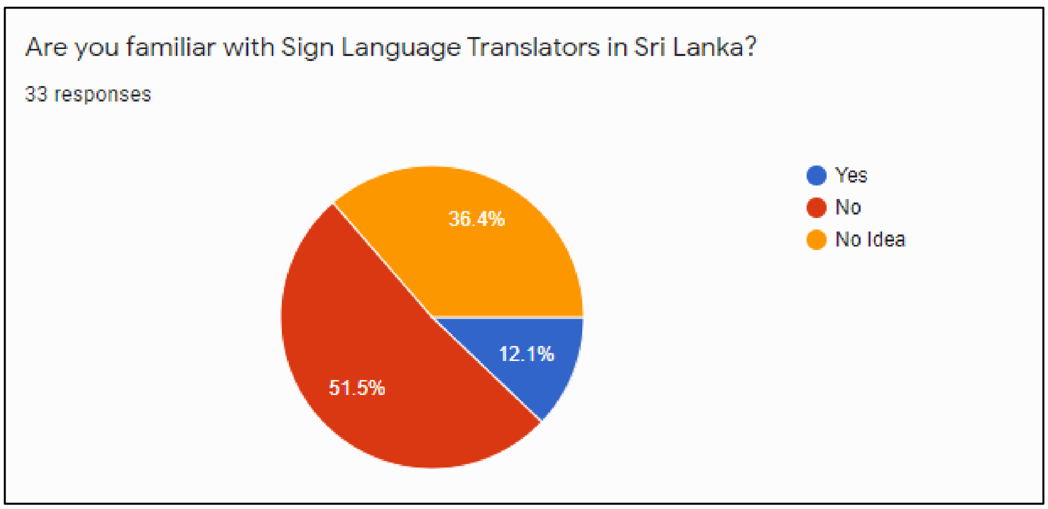


Figure 3-1: Familiarity of Sign Language Translators

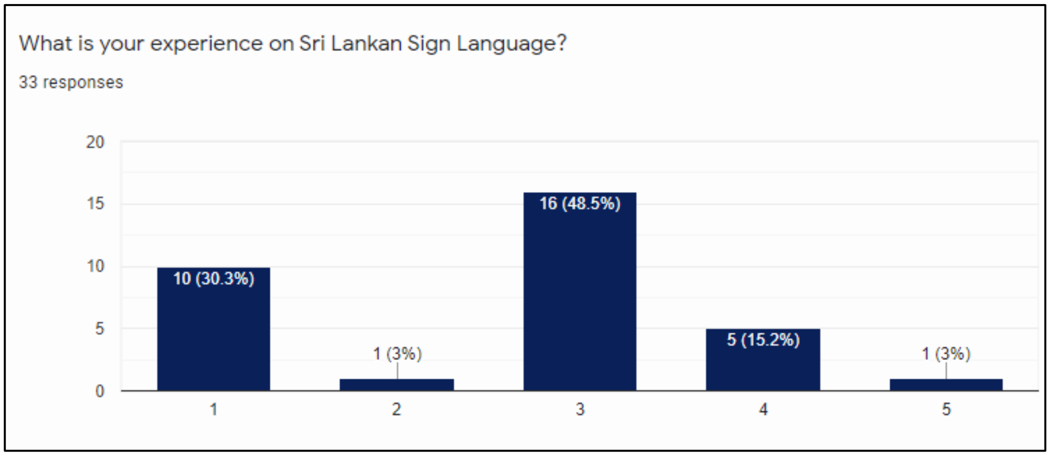


Figure 3-3: The experiencing level of SSL among Ordinary People

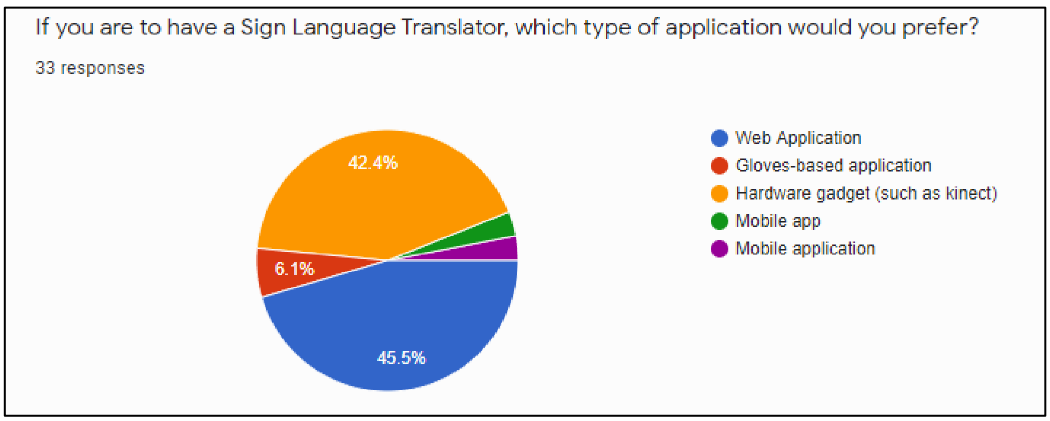


Figure 3-4: Preferable Sign Language Translator Type

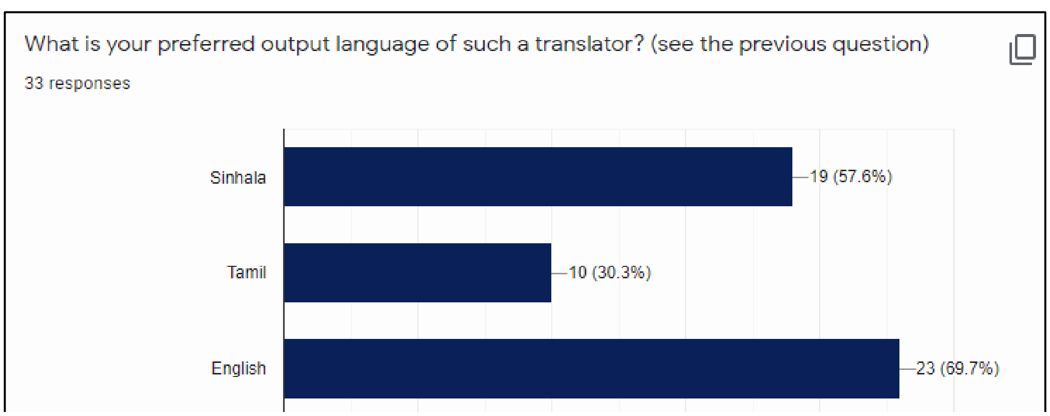


Figure 3-5: The preferable Language to Handle the Translator

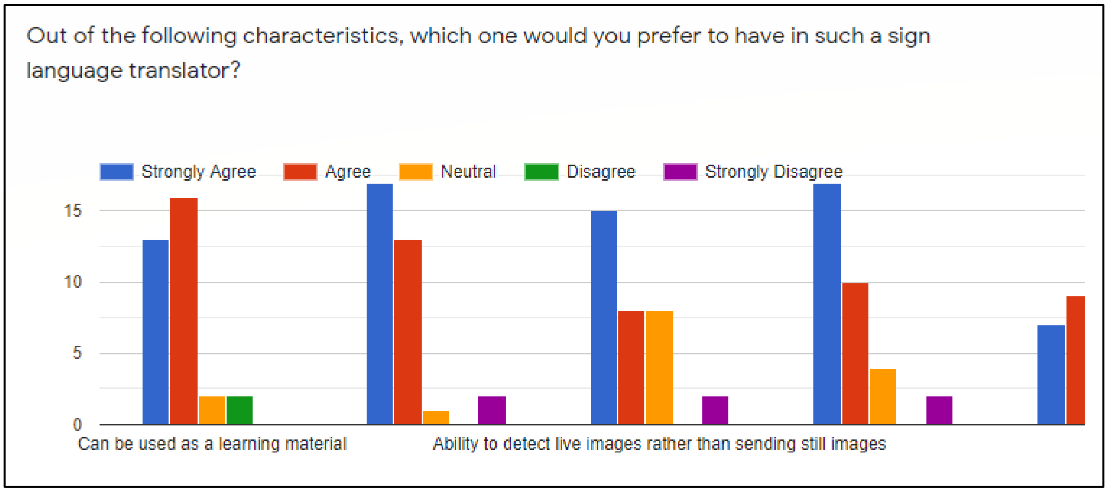


Figure 3-6: The Way of Detecting Signs

3.1.2. Component Survey

The Data Acquisition component is designed to capture the image of the user's hand in real-time. Since this is the initial stage of converting SSL in verbal language, the component should satisfy all the user requirements as it is. The main processes of this component include:

- Detect image captured by a high-resolution web camera
- Detect image captured by the low-resolution web camera
- Identify the shape of the hand from a stream of web camera images
- Calculate the accuracy of being hand
- Pass the identified handshape to Image Recognition and Translation component for further classifications

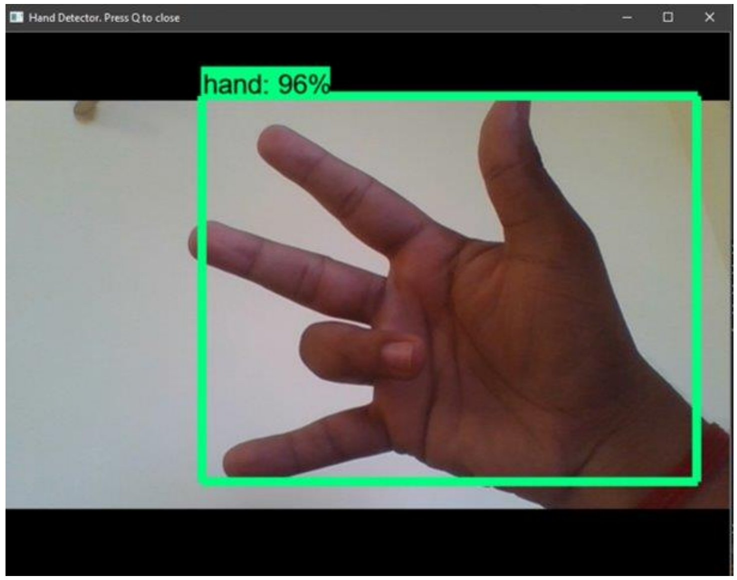


Figure 3-7: Hand Gesture Detection in Calm Environment – Accuracy: 96%

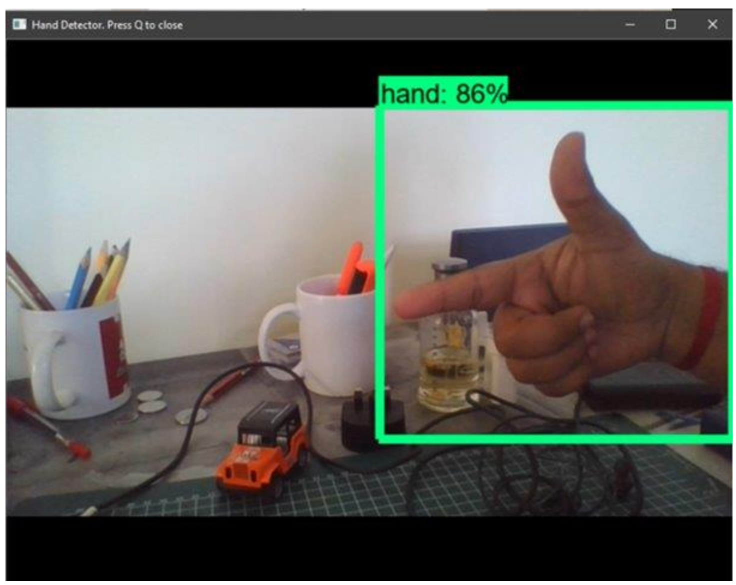


Figure 3-8: Hand gesture detection in a noisy environment – Accuracy: 86%

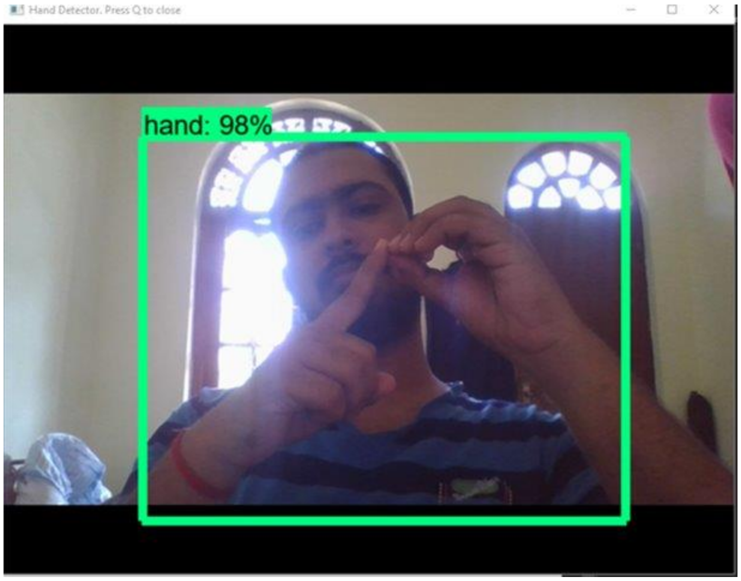


Figure 3-9: Hand gesture detection standing in front of web camera – Accuracy: 98%

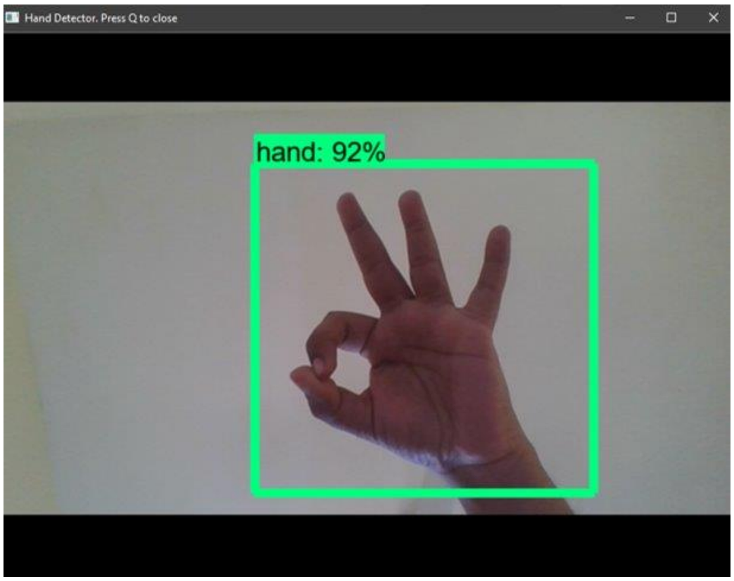


Figure 3-10: Hand gesture detection while turning a hand to right – Accuracy: 92%

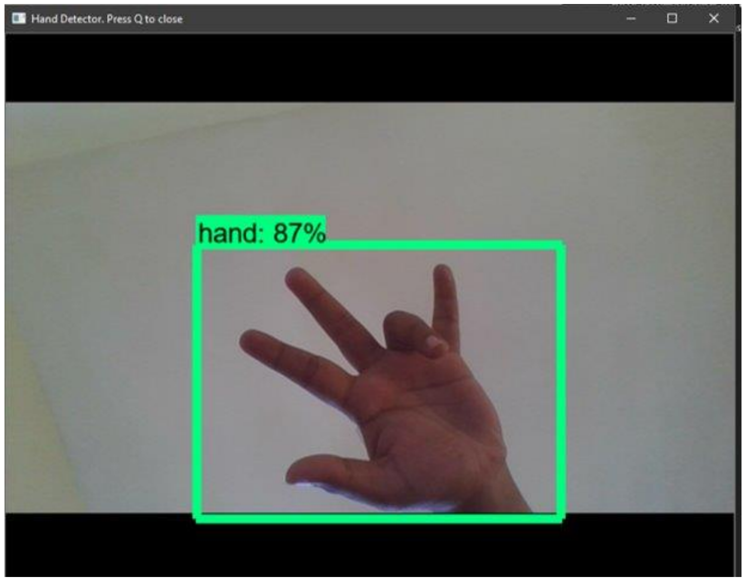


Figure 3-11: Hand gesture detection while turning a hand to the left – Accuracy: 87%

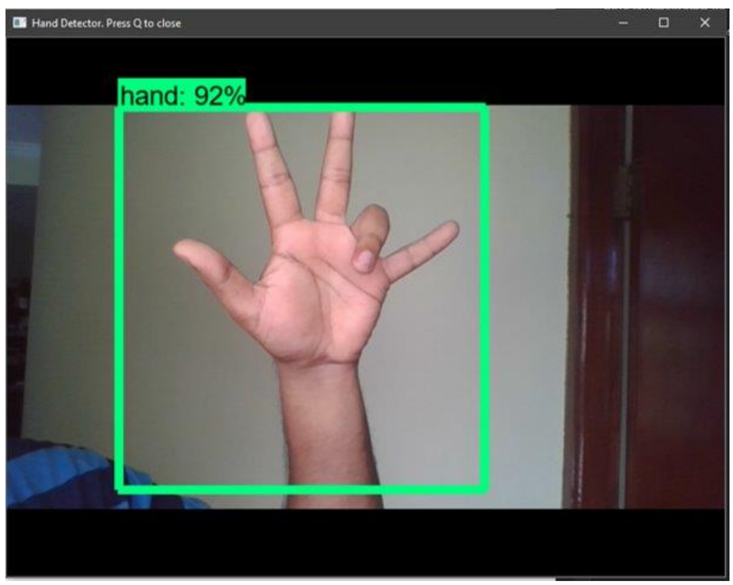


Figure 3-12: Hand gesture detection in a bright background – Accuracy: 92%

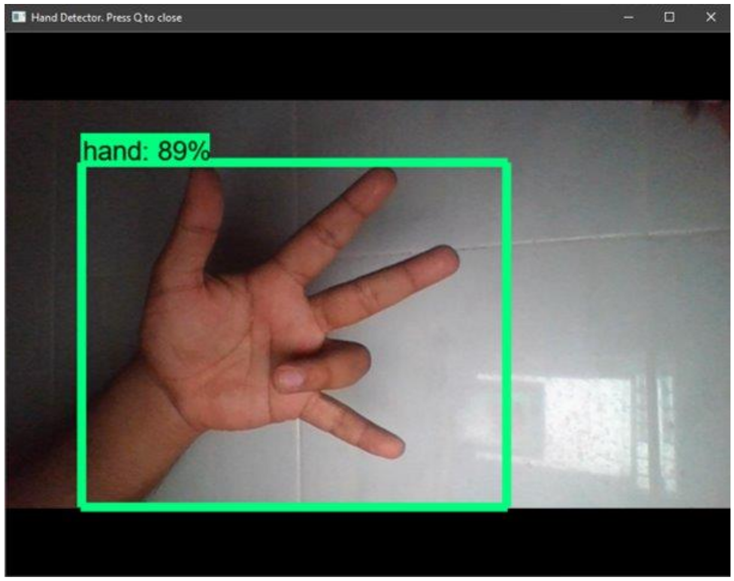


Figure 3-13: Hand gesture detection in the dark background – Accuracy: 89%

3.2 Research Findings | Application

From the results obtained after testing the Hand Detection component, the component was able to detect the hand shape even in a dark background and also in a noisy environment. The average accuracy for these difficult types of detection is 87%. This accuracy is getting for both frontal and non-frontal situations. Situations showing a full image of the hand, get the higher accuracy of hand detection compared to the situation which tries to detect hand gestures while not showing the full hand to the web camera. Also, the component can detect the shape of the hand while showing side pose (turning to left or right). The accuracy gets for these situations are varied from 85% - 95%. Even though the boundaries which frame the hand shape get smaller, it will not affect the hand gesture detection process. Since in the next stage, Image recognition and Translation component, the application starts to identify the gesture of the hand more clearly, this issue will not harm the process critically. The component is also able to identify the shape of the hand with 80% - 88% accuracy while testing the detection process in a noisy background with many unnecessary objects.

3.3 Discussion

From the research carried out for the Data Acquisition component and the test results, it is recognized that there is a drop in the speed of detecting hand shape due to the various several constraints such as GPU delay and obsolete TensorFlow backend version. This is because the dataset which was used as the base for building the model only supports TensorFlow version that is below 1.15.3. This is the next area of improvement and is heavily being worked on. The hand detection accuracy is already as expected. With having a faster detection, the component will be working at its full performance capability. Also, the component is now sending the detected handshape to the next component after the accuracy of a detected hand gesture gets 80% or above. Comparatively, it is an average accuracy that can calculate in between the process. This is taken as a result of testing activities. This issue will not affect the processes of the application in present days since the next component, Image Recognition and Translation does further classifications to reduce the noisiness of the detected hand sign. But the percentage of hand detection accuracy can be upgraded by retraining the model with better datasets and more advanced algorithms

The main objective of this research component is to capture and identify the shape of the hand in real-time especially using low-resolution web cameras. Also, most importantly the handshapes which captured at the first stage are the main user input for the conversion of SSL to Sri Lankan verbal language through EasyTalk, translator assistant. Therefore, the detection of the hand process has been tested under different environment targeting the location where the user can be used the application. Some of these background conditions can be the nosiness of the environment, the lighting conditions of the environment. When considering the lighting condition of the environment, the user gets the low percentage of accuracy as shown in Figure 3.13 i.e. below ninety(90%) percentage accuracy. This is a low accuracy compared to the accuracy of hand detection in a bright lightning condition as shown in Figure 3.12. When in the bright lighting conditions, the hand detection process completes the task with more than 90% accuracy. Therefore, to have the best-detected hand shape, the lighting condition of the background environment will affect directly. Therefore, it is recommended to have a bright lightning facility around the user while the application using. On the other hand, the nosiness of the background environment also affects the final result of the component. As shown in Figure 3.7, the accuracy of the handshape detection increases high when the background has less noisiness. Compared to this, the accuracy level of hand detection is low when there is a noisy background as shown in Figure 3.8. Therefore, it is easy to understand that the user can get a much better outcome when the background environment is calmer. Due to this, it is recommended to use the application in a location where there are a smaller number of unnecessary objects around the user to get the best result.

3.4 Summary of Contribution

Table 4.1: Summary of Contribution

Task	Description
Train Hand Detection Model	Train a model based on top of TensorFlow model zoo for hand detection from web camera
Detect Hand Gestures	Detect hand gestures and show a bounding box over the detected hand with the accuracy on how accurate the hand was detected.

Send detected model to next component	The detected hand image will be sent as a REST API Post call to the next component as NumPy arrays for ease in the transfer.

4 CONCLUSION

In a nutshell, it can be said that despite some technical downfalls, the component was able to detect hand sign at a considerably great rate of accuracy. Discarding the Mask RCNN came with a challenge to use another algorithm which is of less weight. The reason being that Mask RCNN is capable of doing more complex detection with multiple classes. Instead, the model is trained upon TensorFlow models with Faster RCNN configuration. However, this setup came with a technical downside. The TensorFlow model that is compatible with the above-mentioned model training configuration is obsolete. So, adjusting the model and the input was a bigger challenge. This would affect the overall system performance since the application is to be launched in the web.

This was a decision based on a fact that the system should cover all device types regardless of the operating system. This way, there is much less dependency on-device APIs and libraries. Another reason was that it is hard to develop a desktop application in this smaller time frame. So, it was decided to have a web-based client and an API to communicate and transfer information between components. Starting with the implementation, the model will take input from the live web camera live feed. For each of the frames in the feed, a scan is run and displayed in the camera output itself.

The output includes a bounding box and an accuracy level. The bounding box will be drawn around the detected hand and along with the accuracy level. This can be seen in Figure 3-7. After detecting the hand gesture, the frame will be sent as a NumPy array to the next component for translation and classification. The array will be sent through a JSON payload as a REST API Post call. On the other end, the next component (Sign Recognition and Translation) will be translating the identified hand gesture.

Identification of hand gesture has been successful but with some additional technical modifications, the component is now working at full performance regardless of background, lighting or any objects in the background.

REFERENCES

[1] Wikipedia, "Disability in Sri Lanka," 29 08 2017. [Online]. Available: https://en.wikipedia.org/wiki/Disability_in_Sri_Lanka. [Accessed 22 January 2020].

[2] CBM Australia, "Sri Lanka Disability Information Sheet," 2014, pp. 4-5.

[3] Wikipedia, "List of sign languages," 20 January 2019. [Online]. Available: https://en.wikipedia.org/wiki/List_of_sign_languages. [Accessed 29 01 2020].

[4] Wikipedia, "American Sign Language," 7 September 2020. [Online]. Available: https://en.wikipedia.org/wiki/American_Sign_Language. [Accessed 23 July 2020].

[5] Office For National Statistics - UK, "Natiional Archives," 2011. [Online]. Available: https://webarchive.nationalarchives.gov.uk/20160108132401/http://www.ons.gov.uk/ons/dcp171778_297002.pdf. [Accessed 23 July 2020].

[6] Wikipedia, "British Sign Language," [Online]. Available: https://en.wikipedia.org/wiki/British_Sign_Language. [Accessed 13 January 2020].

[7] Asia Media Information and Communication Center, "Pakistan Sign Language – A Synopsis," 15 March 2012. [Online]. Available: http://www.amic.org.sg/ict/external/awards/0202a2_l59attachment2.pdf. [Accessed 22 March 2020].

[8] Wikipedia, "Indo Pakistani Sign Language," [Online]. Available: https://en.wikipedia.org/wiki/Indo-Pakistani_Sign_Language. [Accessed 22 March 2020].

[9] Wikipedia, "Brazilian Sign language," [Online]. Available: https://en.wikipedia.org/wiki/Brazilian_Sign_Language. [Accessed 22 March 2020].

[10] V. U. o. W. Deaf Studies Research Unit, "New Zealand Sign Language Alphabet," 2020. [Online]. Available: https://www.nzsl.nz/. [Accessed 01 April 2020].

[11] Rohana Special School, "An Introduction to Sri Lankan Sign Language," 2007. [Online]. Available: http://www.rohanaspecialschool.org/wp-content/uploads/2010/08/18Alphabets.pdf. [Accessed 15 01 2020].

[12] Wikipedia, "Education in Sri Lanka," 15 February 2018. [Online]. Available: https://en.wikipedia.org/wiki/Education_in_Sri_Lanka. [Accessed 26 January 2020].

[13] Roar Media, "The Relationship Between Disability and Poverty," Roar.media, 7 July 2020. [Online]. Available: https://roar.media/english/life/features/the-relationship-between-disability-and-poverty. [Accessed 01 August 2020].

[14] Y. Ellepola, "Sri Lanka's Invisible Children: The need for inclusive education for children with special needs," Talking Economics, 25 April 2016. [Online]. Available: https://www.ips.lk/talkingeconomics/2016/04/25/sri-lankas-invisible-children-the-need-for-inclusive-education-for-children-with-special-needs/. [Accessed 22 May 2020].

[15] A. Ismail, "Disability is Not Inability," Daily Mirror Sri Lanka, 6 December 2016. [Online]. Available: http://www.dailymirror.lk/120267/Disability-Is-Not-Inability. [Accessed 02 February 2020].

[16] Department of Census and Statistics, "Census of Population and Housing," Ministry of Finance and Planning, Government of Sri Lanka, 2012.

[17] S. Perera, J. Jayalath, W. S. Tissera and A. Bandara, "Sanwada-App," 18 November 2017. [Online]. Available: http://project.sliit.lk/redmine/documents/206. [Accessed 23 January 2020].

[18] A. Agarwal and M. K. Thakur, "Sign Language Recognition using Microsoft Kinect," IEEE, pp. 181-182.

[19] "1.4 Support Vector Machines," [Online]. Available: https://scikit-learn.org/stable/modules/svm.html. [Accessed 02 02 2020].

[20] M. M. Hasan, M. Khaliuzzaman, S. A. Himel and R. T. Chowdury, "Hand Sign Language Recognition for Bangla Alphabet," in *Proceedings of the 2017 4th International Confer*, Dhaka, 2017.

[21] Tensorflow, Google, "Models & datasets," [Online]. Available: https://www.tensorflow.org/resources/models-datasets. [Accessed 01 April 2020].

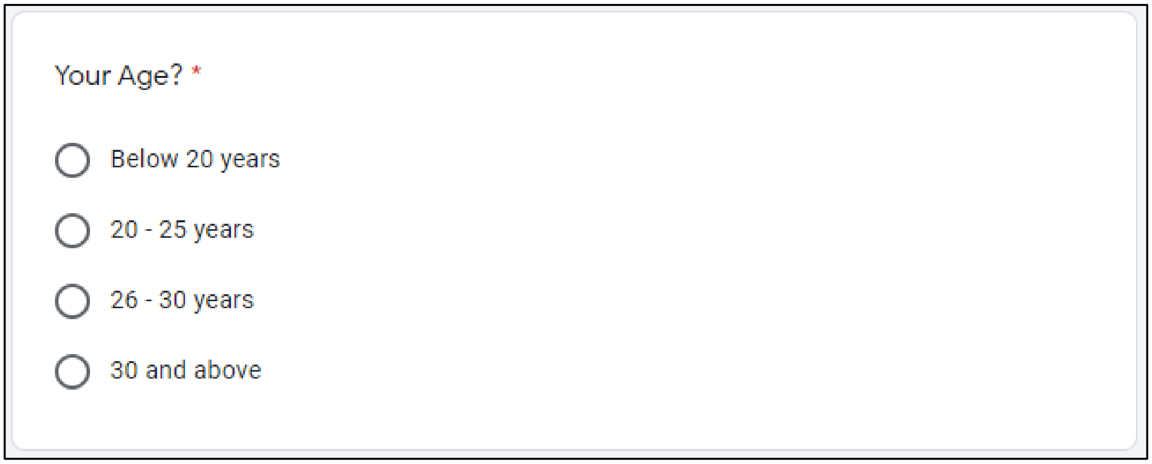
[22] Wikipedia, "COVID-19 pandemic," 2020. [Online]. Available: https://en.wikipedia.org/wiki/COVID-19_pandemic. [Accessed 22 July 2020].

[23] Visual Studio Marketplace, "Python Extension," [Online]. Available: https://marketplace.visualstudio.com/items?itemName=ms-python.python.

APPENDICES

Appendix A. Survey Questions

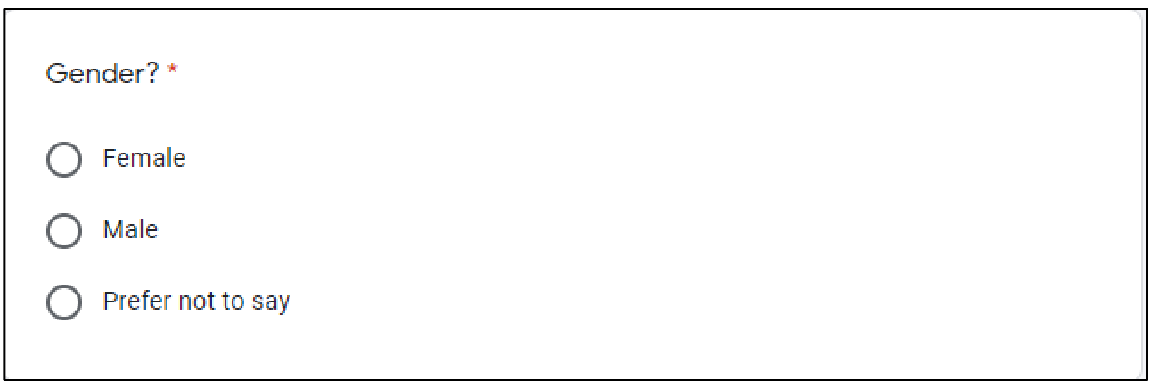
Question 01:



Your Age? *

- Below 20 years
- 20 - 25 years
- 26 - 30 years
- 30 and above

Question 02:



Gender? *

- Female
- Male
- Prefer not to say

Question 03:

Out of the following languages, select your fluency *

	Mother Tongue	Very fluent	Can Read and Write	Can Read	Cannot Read or Write
Sinhala	○	○	○	○	○
Tamil	○	○	○	○	○
English	○	○	○	○	○

Question 04:

What is your experience on Sri Lankan Sign Language? *

	1	2	3	4	5	
No Idea what is Sri Lankan Sign Language	○	○	○	○	○	I am very fluent in Sri Lankan Sign Language

Question 05:

How often do you communicate in Sri Lankan or any sign language in the following places? *

	Always	Often	Sometimes	Never	Never heard of a Sign Language
School / Universities	○	○	○	○	○
Workplace	○	○	○	○	○
Out in the road	○	○	○	○	○
Religious and cultural places	○	○	○	○	○
Social Media	○	○	○	○	○

Question 06:

If you were to learn Sign Language of some sort, what could be the possible barrier in learning them? *

Your answer

Question 07:

According to your opinion, why do you think the hearing and verbally impaired community is reluctant to talk to the ordinary people? *

Your answer

Question 08:

Are you familiar with Sign Language Translators in Sri Lanka? *

- ○ Yes
- ○ No
- ○ No Idea

Question 09:

If you are to have a Sign Language Translator, which type of application would you prefer? *

- ○ Web Application
- ○ Gloves-based application
- ○ Hardware gadget (such as kinect)
- ○ Other: ______

Question 10:

What is your preferred output language of such a translator? (see the previous question) *

- [ ] Sinhala
- [ ] Tamil
- [ ] English

Question 11:

What is your preferred output type of the translator? *

- [ ] Output should be displayed on the screen
- [ ] Output should be presented through audio

Question 12:

Out of the following characteristics, which one would you prefer to have in such a sign language translator? *

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Can be used as a learning material	○	○	○	○	○
Can be used as a mode of communication	○	○	○	○	○
Ability to detect live images rather than sending still images	○	○	○	○	○
Ability to convert sign language into text or speech	○	○	○	○	○
Specific to one domain (health care, industrial, educational etc.)	○	○	○	○	○