

ONLINE LEARNING PLATFORM FOR HEARING IMPAIRED PEOPLE

Project ID: 2022-059

Final Report

Wanasinghe W. A. D. B. – IT19201160

Maddugoda C. D. – IT19153346

Munasinghe M.W.S.S.T.M.B – IT19162706

Ramawickrama H.N – IT19174686

Bachelor of Science (Hons) Degree in Information Technology
Specializing in Data Science

Department of Information Technology
Faculty of Computing

Sri Lanka Institute of Information Technology
Sri Lanka

March 2022

ONLINE LEARNING PLATFORM FOR HEARING IMPAIRED PEOPLE

Project ID: 2022-059

Project Proposal Report

Wanasinghe W. A. D. B. – IT19201160

Maddugoda C. D. – IT19153346

Munasinghe M.W.S.S.T.M.B – IT19162706

Ramawickrama H.N – IT19174686

Supervised by – Dr. Lakmini Abeywardhana

Bachelor of Science (Hons) Degree in Information Technology

Specializing in Data Science

Department of Information Technology

Faculty of Computing





Sri Lanka Institute of Information Technology

Sri Lanka

March 2022

DECLARATION

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

Name	Student ID	Signature
Wanasinghe W. A. D. B.	IT19201160	
Maddugoda C. D.	IT19153346	
Munasinghe M.W.S.S.T.M.B	IT19162706	
Ramawickrama H.N	IT19174686	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

.....

Signature of the supervisor

(Dr. Lakmini Abeywardhana)

.....

Signature of the Co-supervisor

(Mr. Yashas Mallawarachchi)

.....

Date

.....

Date

ABSTRACT

Sign language is the primary means of communication for the hearing-impaired community. Introducing a learning platform can result in many ways to make learning more accessible for the hearing-impaired community of Sri Lanka. Although there are many approaches that are being made to build such systems, the learning platform “Hastha”, is aimed to provide a more interactive outcome by introducing a component that converts YouTube videos to sign language and a Chatbot component that acts as an intermediate between a hearing-impaired user and a Google Search Engine. Furthermore, it includes a game-based learning platform and a gesture translation component from Sri Lankan to American Sign Language while the results are displayed to the users in the form of an animation. The proposed methodology is achieved by using Natural Language Processing, speech recognition, and machine learning techniques. This web-based application has been effective in increasing interaction between the student and the system making it an effective learning environment for the hearing impaired.

Keywords: *Hearing Impaired, Sign Language, Sri Lankan sign language, Online learning, Natural Language Processing*

TABLE OF CONTENTS

DECLARATION	i
ABSTRACT.....	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF ABBREVIATIONS	viii
1. INTRODUCTION	1
1.1 Hearing impaired and Sign Language	1
1.2 Area of Research	2
1.3 Component Overview	2
1.3.1 Teaching SSL and evaluation	2
1.3.2 Sign Language Capturing module	2
1.3.3 SSL to ASL translation.....	3
1.3.4 Search module for SSL.....	3
2 LITERATURE REVIEW	4
3 RESEARCH GAP	8
4 RESEARCH PROBLEM	10
4.1 Teaching SSL and evaluation.....	10
4.1 Sign Language Capturing module.....	11
4.3 SSL to ASL translation	13
4.4 Search module for SSL	14
5 OBJECTIVE.....	15

5.1 Main Objective.....	15
5.2 Specific Objective	15
5.2.1 Teaching SSL and evaluation	15
5.2.2 Sign Language Capturing module	16
5.2.3 SSL to ASL translation.....	17
5.2.4 Search module for SSL.....	17
6 METHODOLOGY	18
6.1 System Architecture Diagram	18
6.1 Teaching SSL and evaluation.....	19
6.2.1 System Architecture Diagram	19
6.2.2 Sign Language Answer Detection	20
6.2.3. Feedback Generation	22
6.3 Sign Language Capturing module.....	25
6.3.1 System Architecture Diagram	26
6.3.2 Audio extraction and splitting	27
6.3.3 Converting to sign language	30
6.3.4 Identifying emotion using speech recognition.....	30
6.3.5 Identifying emotion using text analysis.....	31
6.3.6 Identifying emotions using the facial expression	32
6.4 SSL to ASL translation	35
6.4.1 System Architecture Diagram	35
6.4.2 Conversion of SSL to text	35
6.4.3 Conversion of text to ASL.....	37
6.4.4 Training the avatar model.....	37

6.5 Search Module for SSL	37
6.6 Tools and technologies.....	38
7 TESTING.....	39
8 RESULTS AND DISCUSSION.....	41
8.1 Results	41
8.2.1 Teaching SSL and evaluation	41
8.1.2 Sign Language Capturing module	42
8.1.3 SSL to ASL translation.....	43
8.2 Discussion	44
8.2.1 Teaching SSL and evaluation	44
8.2.2 Sign Language Capturing module	46
8.2.3 SSL to ASL translation.....	46
9.CONCLUSION	51
REFERENCES.....	52
LIST OF APPENDICES	56

LIST OF FIGURES

Figure 1 :Response-real-time feedback on mistakes.....	10
Figure 2 : Response-feedback on the answer 1	11
Figure 3 : Response-feedback on the answer 2	12
Figure 4 :Response-feedback on the answer 3.....	14
Figure 5: System Architecture Diagram	18
Figure 6: Teaching SSL and evaluation-architecture diagram.....	19
Figure 7: folder structure.....	22
Figure 8: Comparison Algorithm.....	24
Figure 9 : Sign language capturing module-architecture diagram	26
Figure 10 : Splitting audio code.....	29
Figure 11 :Text Preprocessing Code.....	31
Figure 12: Lemmatization and Stemming Code	31
Figure 13: Stop Word Removal Code.....	32
Figure 14: Text Analysis Model 1	32
Figure 15: Text Analysis Model 2	32
Figure 16: Text Analysis Model 3	32
Figure 17: Facial Emotion Identification Code.....	33
Figure 18: Frame Trimming Code	34
Figure 19: Facial Emotion Detection Model.....	34
Figure 20: SSL to ASL-System architecture diagram	35
Figure 21: LSTM training model	44

LIST OF TABLES

Table 1 :Comparison between existing studies and the proposed system	9
Table 2: Test case 1	39
Table 3: Test case 2	40
Table 4: Months in a year	40
Table 5: Comparison of Accuracies from Category 1 and Category 2	42
Table 6: SER Results	42
Table 7: Text Emotion Analysis Results.....	43
Table 9 :Accuracy comparison for each category of words.....	44

LIST OF ABBREVIATIONS

SSL	Sri Lankan Sign Language
HI	Hearing Impaired
ASL	American Sign Language
UNICEF	United Nations Children's Fund
ISL	Indian Sign Language
BISINDO	Indonesian Sign Language
ArSL	Arabic Sign Language
API	Application Program Interfaces
MFCC	Mel Frequency Cepstral Coefficients
SER	Speech Emotion Recognition
MLPC	Multi-Layer Perceptron

1. INTRODUCTION

According to the ministry of health [1], it was evidential that 9% of Sri Lankans suffer from some sort of a hearing disorder. Sign Language is the main and official medium of Communication of such hearing-impaired individuals in the world as mentioned in [2]. As per [3], Sign language varies significantly depending on the factors as country, region, and nation. Calling attention to the same reference [3], there exist several sign languages with slightly different regional accents even within the countries where one language is spoken. Similarly, several sign languages with minor variations can be found in Sri Lanka, as per the research paper [4]. Nevertheless, Sri Lankan Sign language (SSL) can be recognized as an elementary attempt to standardize island wise sign language.

As per the findings of [5], solid foundation mother tongue helps in improving literacy skills and acquiring academic skills easier than in any other language. Application of this finding to sign language depicts the importance of helping to learn SSL to Sri Lankan hearing impaired students. Hence although there exist few learning systems for ASL and BSL, developing a learning system to learn SSL correctly from ground level lays the foundation to improve skills of Sri Lankan hearing impaired students immensely.

1.1 Hearing impaired and Sign Language

Hearing impairment is defined as the inability of an individual to hear sounds adequately. This includes people with any degree of hearing loss and individuals who have lost hearing entirely by birth or later in life. The primary communication method of these individuals is using sign language. Sign language is the mean of communication using hand gestures and body movements. One of the main misconceptions regarding sign languages is that it is the same wherever you go. Nevertheless, it is not the case. According to an article published by Richard Brooks [6], there are 138 to 300 sign language variations used around the world today. Among these variations, the variation used in Sri Lanka is known as Sri Lankan sign language [4].

1.2 Area of Research

A lot of research has been done in breaking the barrier between HI and who are not hearing impaired. Many chat applications [7], translating applications [8] and sign language teaching applications [7], [9] ,[10] has been developed over the years. This research focuses on studying and developing an automated learning platform for the hearing impaired taking some issues that they face to give them the opportunity to learn using online tools adapting to the new normal situations due to the pandemic.

1.3 Component Overview

1.3.1 Teaching SSL and evaluation

Developing an interactive online learning system for HIP in SLSL is focused as the end product of the research. Hence, teaching content based SLSL in level-based game for kids with hearing impairments, using a 3D avatar, and evaluating the child with questions at the end of each level is focused on this research component. The evaluation will be conducted by detecting the answer given by the child using the normal camera in the device and comparing the correct answer against the given answer. In case of a partially correct answer, the correct percentage as well as what must be corrected to make the complete answer correct will be provided to the user as feedback through the avatar with expressive emotions. This is expected to encourage the user to learn SLSL with interest and to increase user engagement.

1.3.2 Sign Language Capturing module

The focus of this research component is to develop a way for the HI to gain knowledge using video resources. Although some video resources provide sign language translations along with the video for accessibility, they are more focused on ASL or BSL rather than SSL. This component will provide a solution so that the HI will be able to understand the content of a video in you tube through the embedded avatar who will be translating the content to sign language with facial expressions.

1.3.3 SSL to ASL translation

The focus of this component is to translate SSL into ASL by giving the HIC a more exposure on a different language. There are systems that have been implemented to translate a sign language into a spoken language or vice versa. However this platform is created considering the importance of learning a new language for the HIC

1.3.4 Search module for SSL

This research component focuses on developing a chatbot to user where user can find anything through Google. Therefore, this component works as an intermediate between user and Google search. User can ask the question using SLSL and Chatbot translates it into a Google. Then first ten results of the Google search will be converted into SLSL using the component and present it to the HIP user. If user selects one component reads and present the content. Other than that user can ask for more results or retry the Google search in other way. That is the basic behavior of the component.

2 LITERATURE REVIEW

Learning platform for hearing impaired kindergarten kids has been developed previously for Indonesian sign language with augmented reality in the research [11]. However, this application media was available only in offline mode and hence large RAM capacity and processor speed was required for fast video access. Mitigating this downside, an interactive online learning media for hearing impaired kindergarten children has been developed for Indonesian sign language [12]. Despite the speed access to the resources without requiring higher RAM capacity, this platform has only used flashcards for the teaching component, hence real-time user interaction is observed to be limited.

Static sign language detection mechanism has been implemented for SLSL based on conversational signs in an application called “Wadhan” [13]. In another work [8] Sign language is recognized with template matching technique. In this study gestures are captured by the camera as images and are subjected to feature extraction with analysis, background removal and image smoothing. Then the image is compared against the dataset and respective sign alphabet symbol is displayed as the final result. Since the work is image based, interactivity in the system is observed to be limited.

Augmented reality-based system for Arabic sign language focusing on literacy development of hard hearing children is proposed at the conference [14], anyhow the research is still marked as ongoing, hence real implementation of the proposed system is not found to be available online.

According to J.R. Liddell [15], sign language can be described as a combination of three components.

1. Shape of hand
2. Position of hand
3. Movement of hand

Therefore, analysis of the above three components must be conducted in order to provide accurate feedback to the user. Considering these three factors to provide feedback, an interactive system to teach Irish sign language has been developed according to the research

paper [16], Signs are demonstrated to the user using a virtual teacher and real-time feedback on the sign is given to the user evaluating user performance using colored gloves. The system is available as a software.

In [21], an online learning platform for HIP in ASL is developed. The teaching component in this platform is based on uploaded videos for teaching, and the platform can capture low light videos and enhance them using low light enhancement strategies. This enhancement is approached by converting the video to grey scale from RGB. Background removal and feature extraction methodologies have been followed in order to achieve the best results in capturing. The system is objected to teach basics of ASL such as ASL alphabet. Upon completion of each lesson, the user is asked to repeat the lesson as displayed in the tutorial and depending on the correctness of the answer, the user will be asked to repeat the task or will be given a new task. However, the user will not be given any feedback on the fault part in the answer in this approach.

The mobile-based application “Sanwadha” provides real-time communication between the ordinary and the hearing-impaired community [18]. The application consists of functionalities such as the conversion of text into Sign language, voice to sign language, and GIF conversion. In text conversion, once a text is entered in English or Sinhala, the set of strings will be translated into sign language and will get transformed into a GIF format. Similarly in voice conversion, the application introduces a 2D model or an animated sticker for the deaf user to input the sign and the result will be delivered to the normal user through a voice output.

EasyTalk is another similar application that translates SSL into text and audio formats as well as verbal language into the SSL [17].

This application captures the hand signs through a hand gesture detector using pre-trained models and using an image classifying component it classifies and translates the detected hand signs. The identified hand signs produce a text or an audio formatted output with the aid of text and voice generator components. The users also mention Hand Gesture Detector uses pre-trained models to capture hand gestures. The detected hand signs are classified and translated by the Image Classifier. For detected hand signs, the Text and Voice Generator generates a text or an audio structured output. Finally, the Text to Sign Converter converts

entered English text into animated graphics based on sign language. Many such research have been done on improving the accuracy and improving the ability to interpret sign language to text or speech.

UTalk is an SSL converter developed using Computer vision and Machine learning techniques [19]. This mobile application focuses on interpreting static as well as dynamic signs that are expressed in SSL. The system captures the video of the user performing sign language as the input while extracting the frame segment and removing the background out of those frames with the aid of image processing techniques. These frames that are being preprocessed are made to go through two separate machine learning models stated as static sign classifier and dynamic sign classifier, after which the output will be fed to a language model and presented in a text format.

Another research introduced static sign language recognition using deep learning [13]. The system was based on a skin color modeling technique where the predefined skin color range will extract the pixels from non-pixels, in other words, the hand is separately recognized from the background. The images are then fed into the model Convolutional Neural Network (CNN) for image classification and are later trained using Keras. The system obtained an accuracy of 93.67%

A Japanese team of researchers developed a CG-Animation system for sign language communication between different languages [14]. The system analyzes the image of the sign language gesture, and the image is described through text and programs. CG animation is generated through these texts and programs. Apart from sign language to text conversion, it's been noted that there are a few other systems that have implemented text conversion into the relevant sign language.

Considering text-to-sign language conversion, a study was conducted by a group of researchers from the University of Pennsylvania where they implemented a system that converts English to ASL [15]. The implementation is done based on two approaches where initially the English text which is taken as the input is converted into an intermediate representation after which it is further converted to a set of quantitative parameters which control an articulated 3D human model to produce the ASL.

Similar research was done using Russian sign language. In this research, a semantic analysis algorithm is developed and introduced. The aim of semantic analysis is to model the meaning of the words in the sentence [16].

In [20] the authors have developed an android based learning application for the hearing impaired. According to authors this application contains videos and materials posted by teachers on a variety of subjects. If necessary, it is possible to download it. It means that users can study whenever and wherever they want, without having to worry about time constraints or internet availability. Quiz, set a schedule, event, chat, and a memory game are among the app's other features as well as additional tools to help hearing-impaired people get the most out of online learning. This system is a teacher-based teaching environment which enables the hearing-impaired students to chat with each other if necessary. This system does not address the issue that majority of the HI (deaf) do not know how to read written text. Many such research has been done and has developed learning platforms that can be used by HI but many of these do not address the issue that majority of HI are having trouble reading.

3 RESEARCH GAP

According to the literature survey done above the following issues were found as research gaps,

When we consider existing implementations in this area:

- Although the application “Wadhan” in the research [13] is based on SLSL, the system is limited to static sign language detection based on images. Since interactivity is a key feature in teaching sign language as discussed above, implementation of dynamic capturing is important.
- In the study [21], online learning system with dynamic capturing through visual computing is implemented. However, this application has been developed for ASL and although there exists a teaching component, knowledge evaluation to check if the user has learnt properly is not implemented in this system.
- In the work of [16], feedback enabled learning system is implemented for Irish Sign language as discussed in the literature review. Anyhow, the system requires additional colored gloves to be worn in order to detect the symbol. Further, it is not capable of providing feedback on the answer and is only capable of detecting if the answer is correct or wrong. In case of partially correct answer, user will not be educated on where they got wrong and what part of their answer is correct or wrong. Hence correcting the mistake from the user’s side is tedious.
- In Research A [7], the authors have presented a 3D virtual reality environment where HIP can communicate with each other and to assist the learning and teaching process of the Brazilian sign language.
- In Research B [9], the authors have proposed a teaching environment for sign language which gives live gesture feedback for Irish sign language.
- In Research C [10], the authors have proposed a system for teaching Sri Lankan sign language which would be beneficial for primary school students to learn the basics without any help or guidance from their parents or teachers.

- In Research D [21], the authors have proposed an e-learning platform for the HIP where a lecturer uploads the lesson video, and the user can see the content of the video as caption along with the video. This system also facilitates communication between students and the teacher and teaching of sign language. This system has also proposed a video enhancing feature which would enhance low light videos uploaded by the lecturer without any third-party involvement to produce a clear enhanced lecturer video to the students. This system is based on American sign language.

Table 1 :Comparison between existing studies and the proposed system

Feature Research h	Base d on SSL	Detect if the answer given is correct	Provide s feedbac k on the answer	Generatin g captions in the form of sign language	Analysin g of emotion in each video	Conversio n to sign language	Translatio n of one sign language to another
Research A [7]	✗	✗	✗	✗	✗	✓	✗
Research B [9]	✗	✓	✗	✗	✗	✗	✗
Research C [10]	✓	✗	✗	✗	✗	✓	✗
Research D [21]	✗	✗	✗	✗	✗	✓	✗
Research E [19]	✓	✗	✗	✗	✗	✗	✗
Proposed system	✓	✓	✓	✓	✓	✓	✓

4 RESEARCH PROBLEM

4.1 Teaching SSL and evaluation

HIP Users are lacking resources to connect with the world. It can be Google or YouTube or in any platform. Correcting mistakes is a key area in learning [22]. Unfortunately, out of the limited number of available learning platforms for SSL, none provides the user with the opportunity of knowing the mistakes in their answers in sign language as justified in the literature survey. Further, it is possible to have partially correct answers in SSL as it requires a series of gestures to represent one word as explained. Hence in case of the answer being partially correct, the user should be able to know what part of their answer is correct, where they got wrong and what to improve. Considering these factors, developing feedback enabled learning system for SSL starting from basic signs for hearing impaired children is crucial. The survey results as per the survey conducted by the research team confirm this request, as can be seen in figure 5.

Is it useful if the hearing impaired people are able to get real time feedback on the mistakes in gestures when interpreting words in sign language? (ශ්‍රවණාබාධිත පුද්ගලයන්ට සංඥා භාෂාවෙන් වචන අර්ථකථනය කිරීමේදී අභිනයන්වල ඇති වැරදි පිළිබඳව තත්‍ය කාලීන ප්‍රතිපෝෂණ ලබා ගත හැකි නම් එය ප්‍රයෝජනවත්ද?)

27 responses

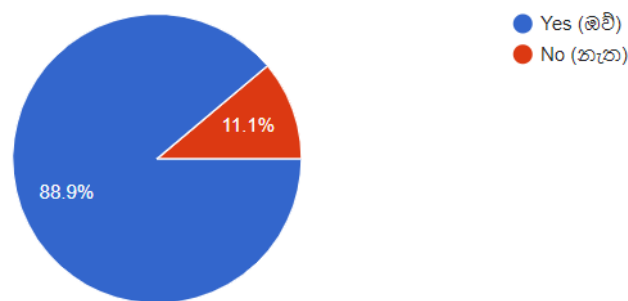


Figure 1 :Response-real-time feedback on mistakes

As per the above survey results, more than 89% have stated that it is useful if the hearing-impaired people are able to receive real time feedback on their mistakes in the gestures when interpreting words in sign language. Hence this requirement will be addressed through the research.

In line with the survey results, it is evidential that most hearing-impaired people expect to state what percentage of their answer is correct or wrong along with the mistake they have done and the correct answer, since more than 77% of the responses have voted for that option. This can be seen in the figure 6

How would you expect the system to provide feedback on the answer they have given?

(ඔවුන් ලබා දී ඇති පිළිතුර පිළිබඳව පද්ධතිය ප්‍රතිපෝෂණ ලබා දෙනු ඇතැයි ඔබ අපේක්ෂා කරන්නේ කෙසේද?)

27 responses

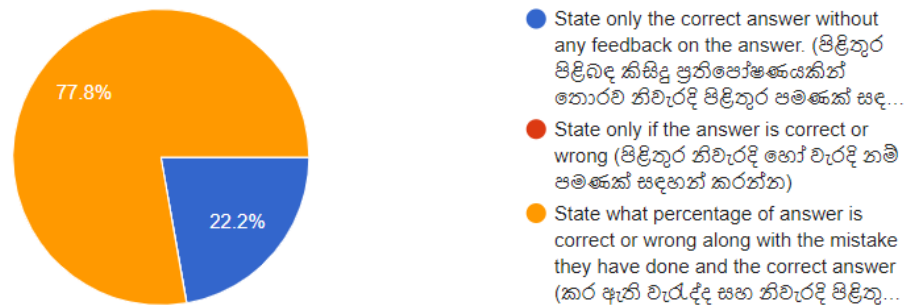


Figure 2 : Response-feedback on the answer 1

Moreover, it is difficult for the teacher to pay individual attention to each student in the physical classroom. Feedback enabled learning systems bring individual attention to each child and hence provide a guaranteed quality education. However, Children learning SSL lacks this opportunity as learning systems with proper feedback is not available for SSL.

4.1 Sign Language Capturing module

A survey was carried out to identify what solutions can be provided to the Hearing-Impaired community of Sri Lanka that would benefit them. This survey was carried out at a deaf school in Sri Lanka and 27 responses were collected.

Do you think it would help the hearing impaired if content of youtube videos to be translated to Sri Lankan sign language to gain knowledge? (ගුවණාබාධිත අයට දැනුම ලබාගැනීම සඳහා යු ටියුබ් වීඩියෝවල අන්තර්ගතය ශ්‍රී ලංකාවේ සංඥා භාෂාවට පරිවර්තනය කළහොත් එය උපකාරයක් වේ යැයි ඔබ සිතනවාද?)

27 responses

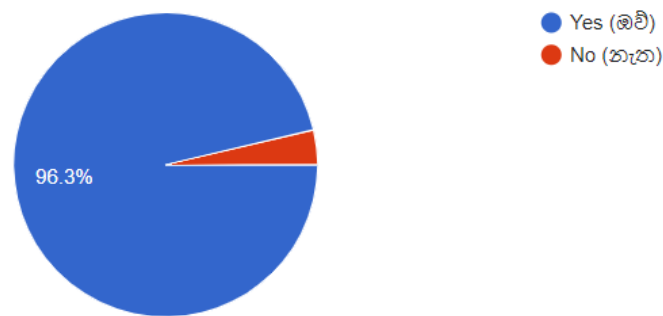


Figure 3 : Response-feedback on the answer 2

According to figure 2.3, 96.3% of the responses suggested that a feature where the content of a video captioned using sign language would benefit them.

In [21] The authors have proposed a platform that makes use of sign language to help students and tutors communicate more effectively while also providing sign language learning resources, practice opportunities, and Q&A sessions. The system includes a low-light enhancement module that enhances the videos submitted by the instructor, as well as a module that converts the uploaded lessons to American Sign Language and converts the sign language questions to the text. In the paper, the authors mention that for the uploaded lessons to be converted to sign language they have used the caption of the video as a text and a text to ASL conversion has been done. The author expects to provide a user with the ability to get the content of the uploaded lesson on ASL. In this system using text as a caption, the method is not the ideal method according to [23] where the author mentions that majority of HI does not read better than elementary level meaning that they have trouble reading long sentences. Also, the systems are based on ASL and do not include emotion analysis. Facial

expressions are a crucial factor in sign language. According to [24] Facial expressions play a vital role in sign language and are used to express emotions when communicating. For example, if a person is happy the greeting “Good morning” would be said with a smiling face and if sad it would be said with a sad facial expression just like people who are able to speak greet with a happy tone when happy and with a sad tone when sad.

4.3 SSL to ASL translation

Most of the existing systems were implemented to facilitate the necessity to translate from a sign language to a spoken language or vice versa. However, a system that is implemented to understand one particular sign language cannot be used to understand another foreign sign language. Due to the unavailability of such interpreters, there is a need for a platform that can provide the HIC to familiarize themselves with various other sign languages. Therefore, this application reaches out to the HIC in Srilanka who needs support with getting SSL translated to a non-native sign language which would further enhance their knowledge of a sign language that they are not familiar with while expanding their communication skills. A survey that was conducted with 27 participants presented that it's beneficial to have a platform that can be used to learn the American Sign language. Figure below shows a diagram of responses to the conducted survey.

Is it useful if the hearing impaired people are able to learn American sign language by the use of Sri Lankan sign language using the platform? (ශ්‍රී ලංකා සංඥා භාෂාව භාවිතා කිරීමෙන් ඇමරිකානු සංඥා භාෂාව ඉගෙන ගැනීමට හැකි නම් එය ප්‍රයෝජනවත්ද?)

27 responses

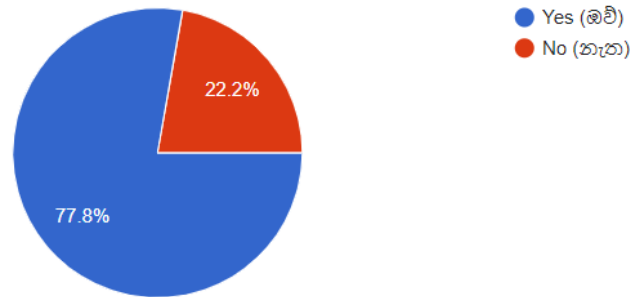


Figure 4 :Response-feedback on the answer 3

A 77.8% of a notable number of responses were given for the above questionnaire which indicates that this platform will be useful for a significant number of people. As a result the approach is carried out as a beginner level translating system that can translate some of the words that are used regularly in the day to day activities into ASL.

4.4 Search module for SSL

HIP Users are lacking resources to connect with the world. It can be Google or YouTube or in any platform. Therefore, within this pandemic and the evolution of the world to a digital world, it is suitable to go for online learning platforms to enhance education system of humans. Then comes to the special learning platforms to the HIP and we identified even in learning platforms not all the platforms focus on everything. We introduced a special feature to connect with google and search anything so HIP users can learn from multiple ways, and it is considered to be a better achievement because most of the obstacles will be destroyed.

5 OBJECTIVE

5.1 Main Objective

The main objective of the research is to develop an interactive learning platform to help learning SSL for hearing impaired children and to guide learning ASL for users who already know SSL in Sri Lanka. The system is expected to provide additional learning recommendations that will be provided to outside videos by converting them into SSL as well as to clear user doubts with an interactive Chabot.

5.2 Specific Objective

The following sub objectives should be fulfilled to achieve the specified objective.

5.2.1 Teaching SSL and evaluation

- **Clearly Teach sign language**

One objective is to teach SSL to the child with hearing impairments in an interesting way as a level-based game, using the gif avatar model. The content in the game is tallied to SSL syllabus and hence the HI kids gets a chance to learn the SSL content interestingly

- **Identify sign language Answer to the quiz question**

The level-based game is consisted of a quiz where at the end of each level knowledge evaluation is conducted. One of the objectives is to detect the answer given by the user/child with a higher accuracy in order to provide effective feedback on the answer. This objective is accomplished using the TensorFlow MediaPipe holistic by detecting the correct gesture using the LSTM -trained model.

- **Provide Feedback on the answer**

One of the important objectives is to compare the answer given by the user/child with the actual answer and provide feedback on what percentage of their answer is correct and what must be improved in order to make the answer completely correct. A comparison algorithm to check the similarity between the given answer and expected

answer is built in order to achieve this objective. Attached to this objective, the system encourages child/user to provide completely correct answers by using the gif avatar

5.2.2 Sign Language Capturing module

- **An algorithm to identify SSSL and converts it to text.**

This model is the very first phase of the proposed Chatbot service. This model will identify user movements through the cv2 OpenCV library. We have created our own dataset of images to identify Signs of Sri Lankan Sign Language. The next step is to train a CNN model with the captured data set. For the whole process, TensorFlow is used for the back-end and Keras is used for image preprocessing. The user gestures will be identified with the use of data sets, through the system. After identification, all the captured footage gets converted into the predicted dataset. Therefore, for the google search model, this text output is used as the input. The specialty of the module is the SSSL signs gets converted into English words directly without getting converted to Sinhala words in Sinhala letters resulting better search results. This is achieved through datasets as, signs in the data-set are in Sri Lankan Sign Language and the meaning of the signs is in the English language.

- **An algorithm to Google the search query**

Extracted output from the Sri Lankan Sign Language to text module is used as the input to this module. Python library is used for the process of Google Search , and BeautifulSoup is used for Web Scraping. The text output from the previous model is used as a search query and applied through the Google Search engine. After the module identifying the most relevant ten search results using arrays the text is given to the Sri Lankan Sign Language module to present search results through sign language. If the user wants more results with the search query that was used at the beginning, user can inquire more and that will be identified as a reply to the first search query. Then the module delivers the next ten search results. The model is capable of successfully

returning the most relevant search results. Therefore, most of the time users will find the best solution in the first set.

- **Text to SLSL algorithm**

This module takes all the search results given by the Google Search results module and translates them to Sri Lankan Sign Language using a python server. The system can easily translate search results into English words since the datasets are in English words representing Sri Lankan Sign Language signs. Then the final output gets presented through an animated avatar model.

5.2.3 SSL to ASL translation

- Creating an interactive and efficient learning platform for the HIC by enhancing their communication aspects.
- Ensure that an accurate response is given by the system according to the user requirements.

5.2.4 Search module for SSL

- **An algorithm to identify SLSL and converts it to text then Google**

This objective focus on converting it to a text and search through Google in the next step. Then component identifies search results and identifies title of first ten search results.

- **Presenting search results using 3D Model.**

This objective aim on presenting SLSL using a 3D avatar model which we identified search results and content of selected search result.

6 METHODOLOGY

The objective as mentioned of this component is to provide the HIP the ability to get knowledge through Google search by converting the content to sign language and vice versa. To achieve this first the real time video is captured using Open CV. Then converted to a text and search through Google. Then system will identify search results in bundles of ten. Then title of first ten research bundles will be displayed using 3D avatar model.

6.1 System Architecture Diagram

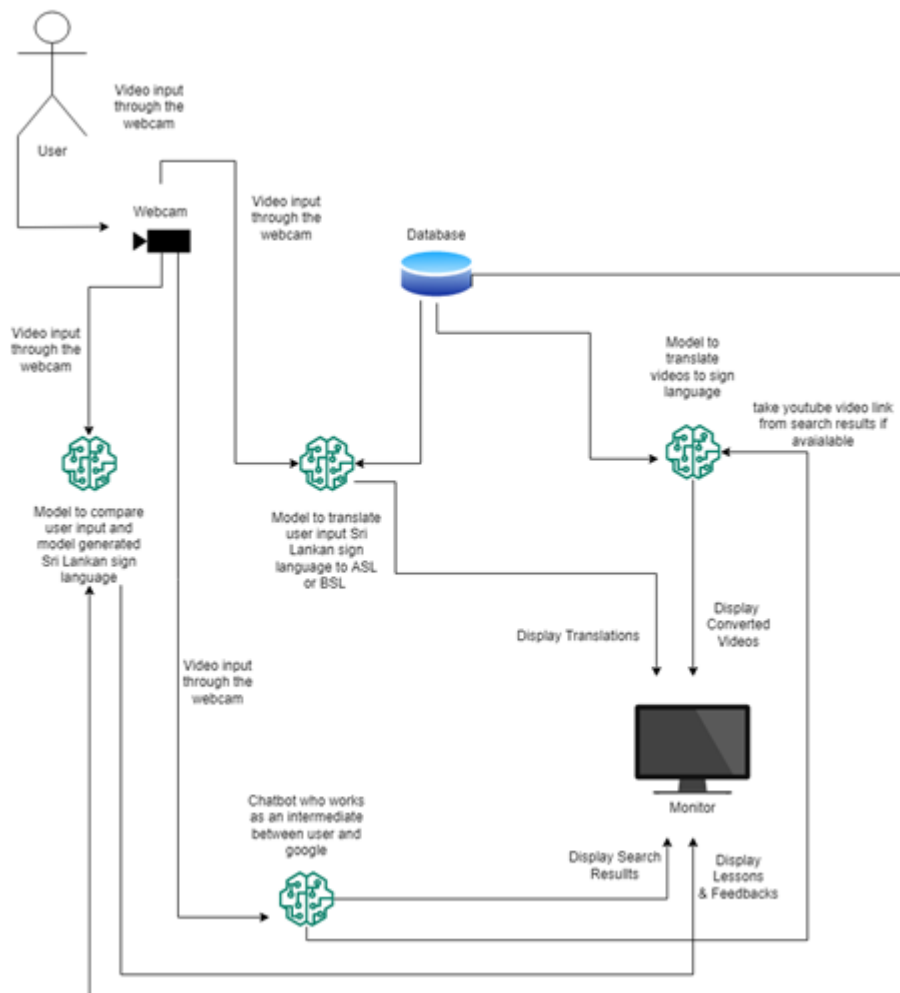


Figure 5: System Architecture Diagram

6.1 Teaching SSL and evaluation

This component is objected to help HI (specifically hearing-impaired children) in learning SSL. As discussed above this is achieved as a level base game where customized evaluation is carried at the end of each level. The SSL content is taught using the designed avatar. The tedious task here is to provide feedback on the performance of the child against the questions being asked by the avatar at the end of each level in the game. Each answer given is compared with the available answer in the database and the correct percentage is displayed. In case of a partially correct answer, the mismatched section should be conveyed to the user, using the avatar.

6.2.1 System Architecture Diagram

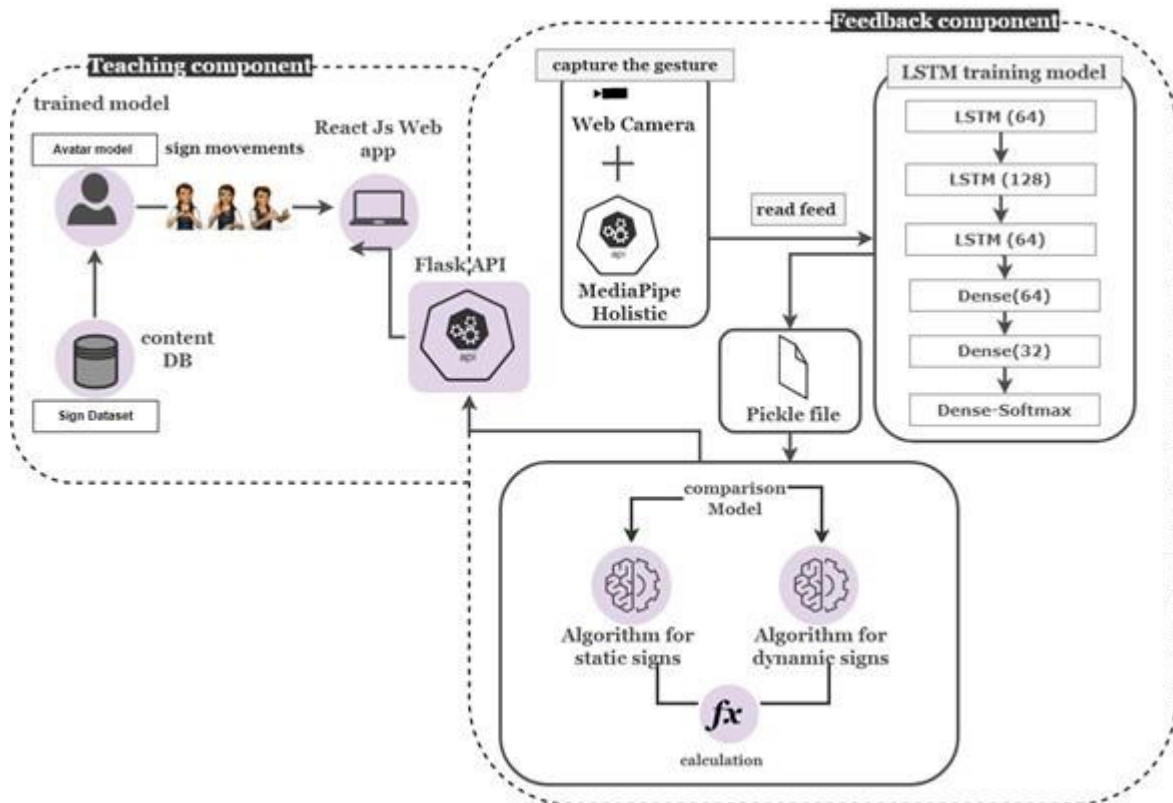


Figure 6: Teaching SSL and evaluation-architecture diagram

As demonstrated in the architecture diagram in figure 6, the content stored in the database is taught using gif avatar model. Upon completion of learning users can take quizzes to check their knowledge in SSL. Answer for each question in the quiz are captured using web cam and MediaPipe holistic. The feed is then inserted to LSTM training model where the model gets trained using 3 LSTM layers and 3 dense layers. The trained model is saved to a pickle file for future reference. Once the answer is detected, it is checked with respect to saved content in pickle file and correctness of the answer is checked using a comparison algorithm. Then the obtained results with evaluation is passed to the frontend through Flask API for the display to the user.

6.2.2 Sign Language Answer Detection

6.2.2.1 Importing Libraries

The following libraries are imported for the below mentioned functionalities.

Tensorflow,opencv	-Access the webcam
Mediapipe holistic	-Extract keypoints
Numpy	-To structure the arrays
Sklearn	-For evaluation marix
Matplotlib	-To make visualization easier

6.2.2.2 Importing Models

The following models in Keras library is imported for the training.

Sequential	-To Build the sequential neural network
LSTM, Dense	- For action detection
Tensorboard	-To monitor and traise model

6.1.2.3 Dataset Preparation

The detections are captured using OpenCV library and feed is read as frames. The stacked frames are equivalent to a video feed. Then the holistic model and drawing utilities are

defined to make the detections and draw the key points so that the user can observe the detected key points. Landmarks are detected using the drawn key points. Landmarks are represented in terms of below visibility values

1. x: X axis position
2. y: Y axis position
3. z: Relative distance to the camera

flattened array of these x,y,z visibility values are used for keypoint extraction.

6.1.2.4 Color Format

The original capture from the camera is in the form of BGR but for mediaPipe holistic the color format of RGB is required hence the color format is converted from BGR to RGB using `cvtColor` function in OpenCV. After detection, colors are converted back to BGR for display.

6.1.2.5 Landmark Visualization

Landmarks and connections are formatted defining color, thickness and circle radius for better display of landmarks.

6.1.2.6 Training

Sign Language input is captured using a Media Pipe Holistic and Tenser-flow object detection model. The necessary key points, pose landmarks, and hand landmarks are detected and marked using the media pipe holistic method. Then the detected landmarks are extracted, concatenated, and saved to a NumPy array.

The training of the model is separated into two selections based on below Category.

1. Category 1: SSL words that can be represented using one/few gestures (static and simple dynamic signs)
2. Category 2: SSL words that involve a series of gestures (dynamic signs with multiple gestures)

Category 1:

A group of words in the SSL syllabus are arranged into a single NumPy array, and 40 sequences each of which is 30 frames in length are trained for each word in the array and saved into a NumPy array within a folder. A collection of such folders containing NumPy arrays with frames are saved to another folder. A collection of such sequence folders is saved into another folder (one folder for each word/gesture).

Folder structure used is given below figure 7.

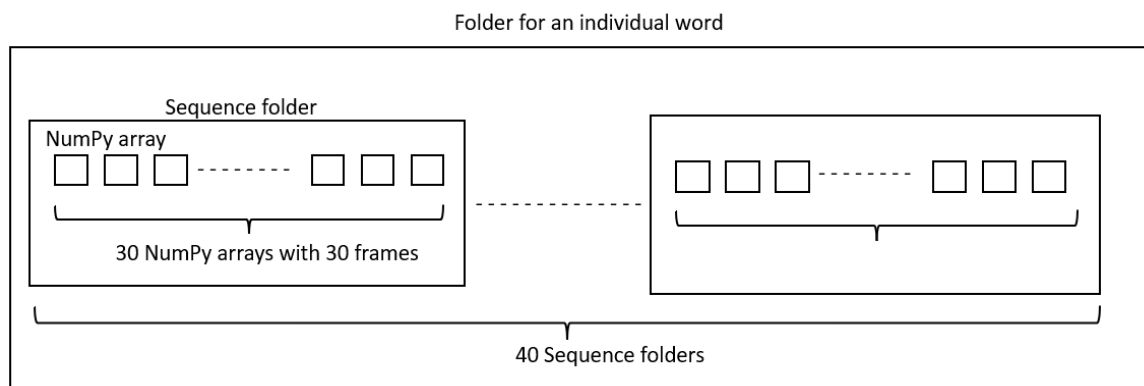


Figure 7: folder structure

Category 2:

Training steps are similar to Criteria 1. However, the NumPy array in these criteria consists of words each of which is involved with several gestures. Hence each word is saved in separate NumPy arrays. The array size is limited to several gestures.

6.2.3. Feedback Generation

Feedback is generated through a comparison algorithm. Once the series of gestures are detected, they get saved to a list in a pickle file for comparison with the expected answer. The saved model with the expected answer is compared with the detected answer saved in the pickle file. The algorithm is designed to vary according to the above-mentioned categories. For category 1, the answers in the two lists are compared element wise and different score counts are given for providing correct answers with a different number of

attempts. If the attempts exceed 3, the answer for that specific word is marked as incorrect and the loop is set to break.

The algorithm built for comparison category 2 is designed to evaluate each gesture in one word. Therefore, a score value is not calculated for each element in the array since one element is corresponding to a gesture. Nevertheless, each gesture is evaluated separately, and the correctness of the gesture is marked for each word. In the case of a partially correct answer, the sequence of gestures is marked as correct and wrong individually hence the user is able to know which gesture correct, and which gesture is wrong. Elements in the two arrays are compared using the zip function and if both elements are found to be correct, the gesture gets marked as correct.

The algorithm developed for comparison is shown in the below figure 8.

```

i=[]
j=[]
def arrays_equal(a, b, c):
    for ai, bi in zip(a, b):
        if len(j)>=len(b) or len(i)>=len(a) or c==1:
            print('breaking')
            break

        if a[len(i)]==b[len(j)]:
            print('correct')
            ans.append('correct')
            i.append(1)
            j.append(1)

        else:
            j.append(1)
            print('Try Again')
            trial=[]
            while a[len(i)]!=b[len(j)]:
                if len(trial)>1:
                    print('Incorrect')
                    ans.append('Incorrect')
                    break
                else:
                    print('Try Again')
                    trial.append(1)
                    j.append(1)
                    print(a[len(i)],b[len(j)])

            if len(trial)>1:
                if a[len(i)]==b[len(j)]:
                    print('correct')
                    ans.append('correct')
                    i.append(1)
                    j.append(1)
                if len(j)==len(b):
                    c=c+1
                    break
                else:
                    i.append(1)
                    j.append(1)
                    continue

            else:
                if a[len(i)]==b[len(j)]:
                    print('correct')
                    ans.append('correct')
                    i.append(1)
                    j.append(1)

                if len(j)>=len(b) or len(i)>=len(a):
                    break

    return ans

```

Figure 8: Comparison Algorithm

6.3 Sign Language Capturing module

The main functionality of this component is to capture the content of a provided video and to convert the content to sign language to display as a caption while identifying and displaying the emotion along with it. HI can provide a YouTube video link and the system will take the link and will download the video. Then the system will extract the audio segment from the video and will split the audio file into chunks based on the silences. These audio chunks are used to identify the emotions using speech recognition as well as to identify the emotion after converting to text. Then the video is also broken into parts using the time stamps of the audio chunks to be used to identify the emotion using facial expressions if facial expressions are available in the video.

6.3.1 System Architecture Diagram

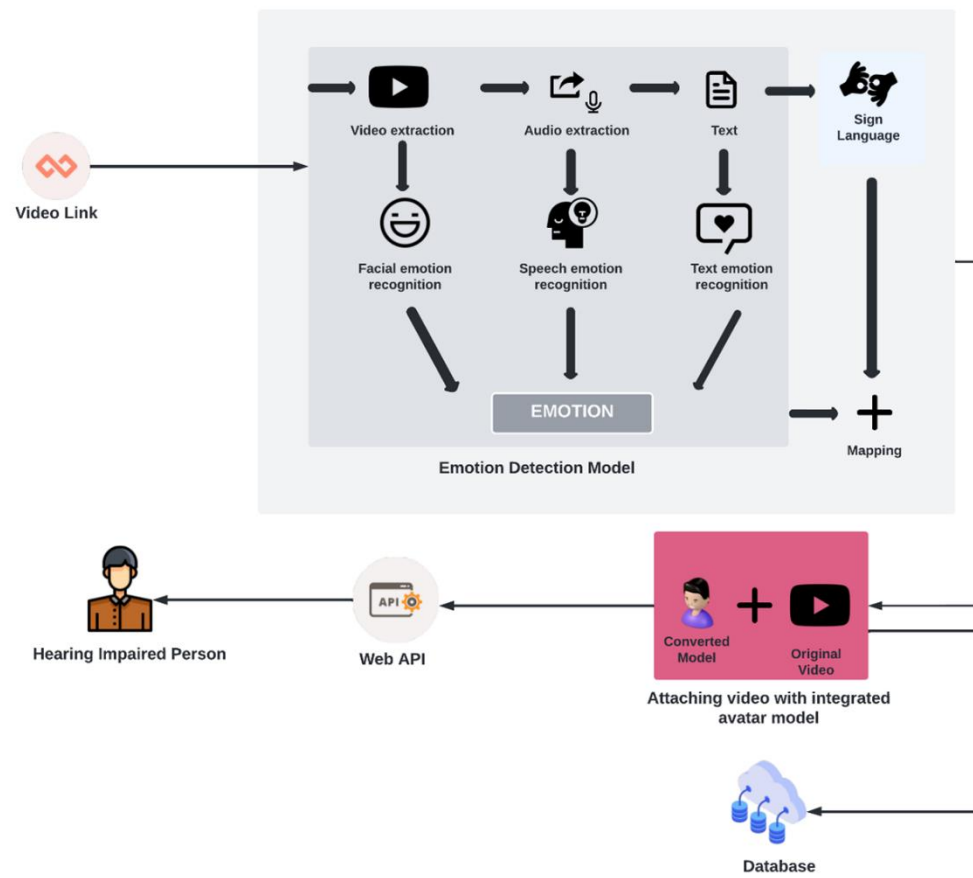


Figure 9 : Sign language capturing module-architecture diagram

The above functional diagram shows the functionality of emotion identification and conversion of video content to SSL.

6.3.2 Audio extraction and splitting

6.3.2.1 Downloading video

The video is downloaded using the youtube dl library of python where the video link is provided by the user and the video will be downloaded in mp4 format.

```
# Downloading given youtube video for processing
def download_video(link):

    link_split = link.split('/')
    end = link_split[3]

    ydl_opts = {}
    with youtube_dl.YoutubeDL(ydl_opts) as ydl:
        ydl.download([link])
        info_dict = ydl.extract_info(link, download=False)
        video_title = info_dict.get('title', None)

    return video_title , end
```

Figure 7 - Download Video Code

6.3.2.2 Audio extraction

After the video is downloaded this will be used to extract the audio to be used for emotion identification using SER and also to extract the text to convert content to sign language. This is done using ffmpeg and the subprocess module of python which is a module that allows you to spawn processes, connect to their input/output/error pipes, and obtain their return codes. The extracted audio is saved as a wav file.

```
def convert_video_to_audio_ffmpeg(video_file, output_ext="wav"):
    """Converts video to audio directly using `ffmpeg` command
    with the help of subprocess module"""
    filename, ext = os.path.splitext(video_file)
    subprocess.call(["ffmpeg", "-y", "-i", video_file, f"{filename}.{output_ext}"],
                    stdout=subprocess.DEVNULL,
                    stderr=subprocess.STDOUT)
```

Figure 8 - Audio Extraction Code

6.3.2.3 Audio and Video splitting by silences

Since it is necessary to analyze the emotion sentence by sentence it is essential that we split the video and the audio by sentences as well. This is accomplished using the pydub python library for audio splitting and using ffmpeg tools for the video.

The audio is first split based on the silences where if there is no audio for 5 seconds the system will consider this to be an end of the sentence and will split the audio accordingly. A silent threshold is defined as -16 dBFS to let the system identify beyond what amplitude level is to be considered as silence.

```

# Function which splits given audio in to chunks by sentences
def break_chunks(path):

    # open the audio file using pydub
    sound = AudioSegment.from_wav(path)

    # split audio sound where silence is 500 milliseconds or more and get chunks
    chunks = split_on_silence(sound,
        min_silence_len = 500,
        silence_thresh = sound.dBFS-14,
        keep_silence=500,
    )

    # create a directory to store the audio chunks
    folder_name = "audio-chunks"

    if not os.path.isdir(folder_name):
        os.mkdir(folder_name)

    # process each chunk
    # export audio chunk and save it in
    # the `folder_name` directory.
    for i, audio_chunk in enumerate(chunks, start=1):
        chunk_filename = os.path.join(folder_name, f"chunk{i}.wav")
        audio_chunk.export(chunk_filename, format="wav")

    return chunks

```

Figure 10 : Splitting audio code

After audio is broken into parts based on the timestamps of the audio segments the video is also broken into parts with ffmpeg tools.

```

def break_video(video):
    #Convert wav to audio_segment
    audio_segment = AudioSegment.from_wav("path.wav")

    #normalize audio_segment to -20dBFS
    normalized_sound = match_target_amplitude(audio_segment, -20.0)
    print("length of audio_segment={} seconds".format(len(normalized_sound)/1000))

    #Print detected non-silent chunks, which in our case would be spoken words.
    nonsilent_data = detect_nonsilent(normalized_sound, min_silence_len=500, silence_thresh=normalized_sound.dBFS-14, seek_step=1)

    times = []
    #convert ms to seconds
    print("start,stop")
    for chunks in nonsilent_data:
        print( [chunk/1000 for chunk in chunks])
        times.append([chunk/1000 for chunk in chunks])

    required_video_file = video

    for time in times:
        starttime = time[0]
        endtime = time[1]
        ffmpeg_extract_subclip(required_video_file, starttime, endtime, targetname="video/"+str(times.index(time)+1)+".mp4")

```

Figure 10: Splitting Video Code

The process of the captioning module contains 4 sub parts.

- Converting to sign language.
- Identifying emotion using speech recognition.
- Identifying emotion using text analysis.
- Identifying emotions using the facial expression in the video

if there are any.

6.3.3 Converting to sign language

Through this component, the extracted text is converted into sign language. First, the text goes through a tokenization process where the sentences are broken into words choosing only the words that are necessary for conversion. Then the stop words are removed which eliminates common words like articles, prepositions, models, conjunctions, etc. Then the words are stemmed to their root. After this process is completed, the remaining words are taken, and the respective sign related to the word will be used to display the output.

6.3.4 Identifying emotion using speech recognition.

This component is focused on identifying the emotion based on the extracted audio from the provided video. Here the emotions are segmented into three basic parts negative, positive, and neutral.

6.3.5 Identifying emotion using text analysis.

This component is focused on identifying the emotion based on the extracted text from the provided video. Here the emotions are segmented into three basic parts negative, positive, and neutral.

The dataset is initially cleaned to remove any characters and punctuations.

```
def cleantext(data):
    data = re.sub(r'@[A-Za-z0-9]+', '', data) # remove @mentions
    data = re.sub(r'#', '', data) # remove # tag
    data = re.sub(r'RT[\s]+', '', data) # remove the RT
    data = re.sub(r'https?:\/\/\S+', '', data) # remove links
    data = re.sub(r'(\u{1}[\u{1}a-z][\u{1}0-9])+', ' ', data) # remove unicode characters
    data = re.sub(r'\"', '', data)
    data = re.sub(r':', '', data)
    data = re.sub(r'=', '', data)
    data = re.sub(r'`', '', data)

    return data
```

Figure 11 :Text Preprocessing Code

Next, the text is lemmatized and stemmed for classification.

```
def pre_process(text):
    text_1 = text.split()

    lemmatizer = WordNetLemmatizer()
    ps = PorterStemmer()
    lemmatized_words=[]
    for w in text_1:
        w = ps.stem(w)
        lemmatized_words.append(lemmatizer.lemmatize(w))
```

Figure 12: Lemmatization and Stemming Code

The words were sorted and finally were taken through the process of removing stop words.

```
vectorizer = CountVectorizer(max_features=1500, min_df=5, max_df=0.7, stop_words=stopwords.words('english'))
X = vectorizer.fit_transform(data['Text']).toarray()
tfidfconverter = TfidfTransformer()
X = tfidfconverter.fit_transform(X).toarray()
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 13: Stop Word Removal Code

Finally, the text was classified using the model.

```
from sklearn.ensemble import RandomForestClassifier
classifier = RandomForestClassifier(n_estimators=100, random_state=0)
classifier.fit(X_train, y_train)
```

Figure 14: Text Analysis Model 1

```
lr=LogisticRegression(max_iter=1000, multi_class='multinomial')
lem=lr.fit(X_train, y_train)
```

Figure 15: Text Analysis Model 2

```
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
analyser=SentimentIntensityAnalyzer()
```

Figure 16: Text Analysis Model 3

6.3.6 Identifying emotions using the facial expression

This component focuses on detecting any facial expressions that are in the video and if any. Facial expressions are detected by the system will capture this frame using OpenCV.

```

# Function to extract frames
def FrameCapture(path , i):

    cap = cv2.VideoCapture(path)
    count = 0

    paths = 'images/'+str(i)
    # Check whether the specified path exists or not
    isExist = os.path.exists(paths)

    if not isExist:
        # Create a new directory because it does not exist
        | os.makedirs(paths)

    while cap.isOpened():
        ret, frame = cap.read()
        if ret:
            # Convert into grayscale
            gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)

            face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades + "haarcascade_frontalface_default.xml")
            # Detect faces
            faces = face_cascade.detectMultiScale(gray, 1.1, 4)
            for (x, y, w, h) in faces:
                cv2.rectangle(frame, (x, y), (x+w, y+h),
                    | (0, 0, 255), 2)

                faces = frame[y:y + h, x:x + w]
                cv2.imwrite('images/'+str(i)+'/'+'frame{:d}.jpg'.format(count), faces)

            count += 30 # i.e. at 30 fps, this advances one second
            cap.set(1, count)
        else:
            cap.release()
            break

```

Figure 17: Facial Emotion Identification Code

After facial expressions are detected and frames are captured these frames will go through a process where the images are cropped and enhanced.

```

def trim():
    x = 0
    paths = 'images/'+str(x)

    # Check whether the specified path exists or not
    isExist = os.path.exists(paths)

    if not isExist:
        # Create a new directory because it does not exist
        os.makedirs(paths)
    for i in range(0,1000):
        # Read the input image
        img = cv2.imread('images/frame'+str(x)+'.jpg')

        # Convert into grayscale
        gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)

        face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades + "haarcascade_frontalface_default.xml")

        # Detect faces
        faces = face_cascade.detectMultiScale(gray, 1.1, 4)
        for (x, y, w, h) in faces:
            cv2.rectangle(img, (x, y), (x+w, y+h),
                           (0, 0, 255), 2)

            faces = img[y:y + h, x:x + w]
            cv2.imshow("face", faces)
            cv2.imwrite('faces/face'+str(x)+'.jpg', faces)

    cv2.imshow('img', img)
    cv2.waitKey()

```

Figure 18: Frame Trimming Code

Finally, the facial emotion along with its percentage will be identified using the DeepFace library in python.

```

def detect_emotion(img):
    img1 = cv2.imread(img)
    result = DeepFace.analyze(img1 , actions = ['emotion'] , enforce_detection=False)
    # happy = result["happy"]
    # sad = result["sad"]
    # neutral = result["neutral"]

    return result

```

Figure 19: Facial Emotion Detection Model

6.4 SSL to ASL translation

6.4.1 System Architecture Diagram

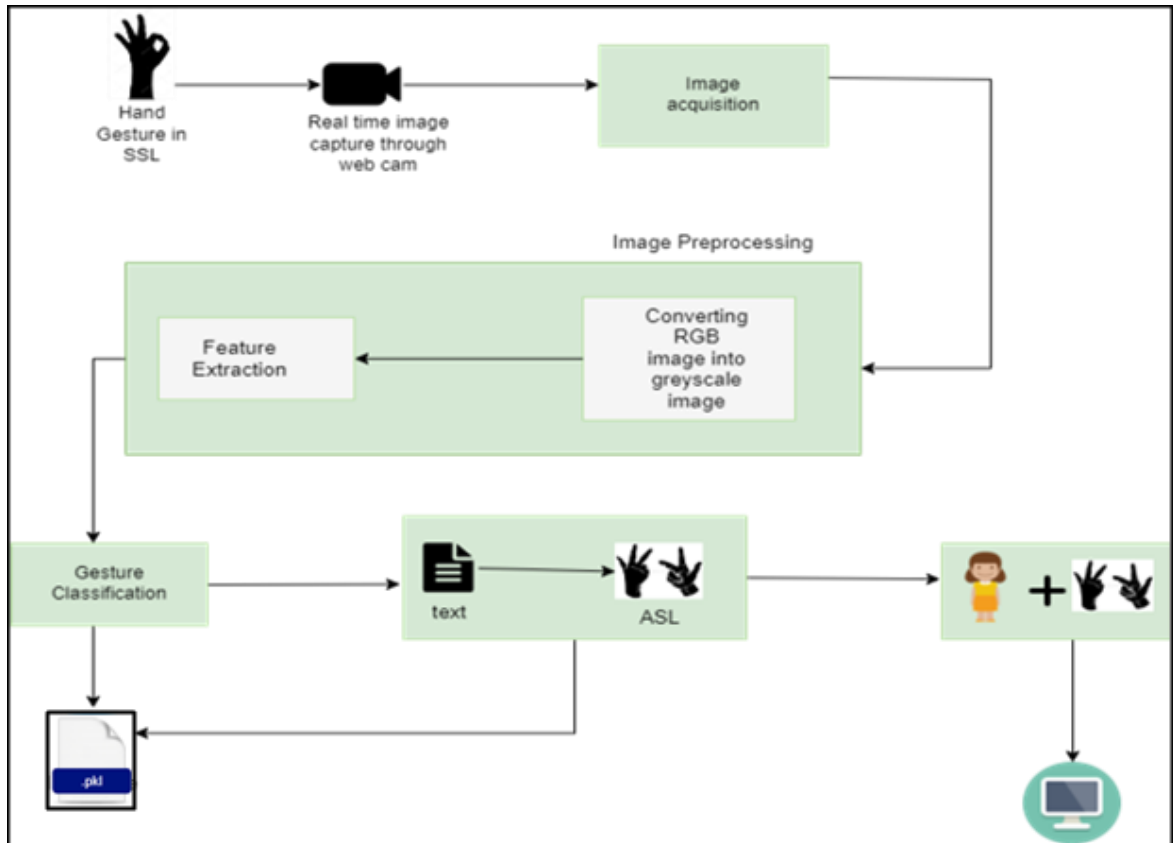
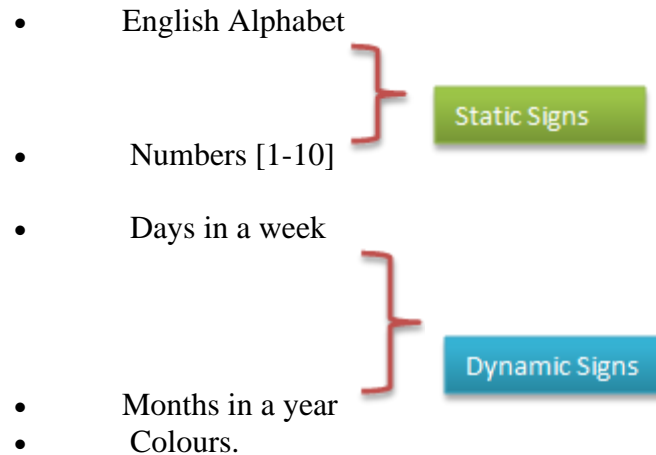


Figure 20: SSL to ASL-System architecture diagram

6.4.2 Conversion of SSL to text

As the first half of the component, a selected set of categories of gestures was chosen to convert into text. Considering the difficulties in finding datasets in Sri Lankan sign language, a dataset was generated by collecting images of the hand gestures from the web camera and passed to train. A collection of words was selected to translate initially belonging into the following categories:



Considering sign language, gestures mainly fall under two main categories as static and dynamic signs. Therefore, the English alphabet and Numbers [1-10] which contains static gestures are trained using a CNN model through image processing while dynamic gestures belonging to Days of the week, Months, and Colours are trained with the use of an LSTM model through video processing.

When considering the static signs an image count of roughly around 300-400 was collected per class from the web camera, summing up to 5000+ images overall. Considering that the original images that were collected come in an RGB color format, they are being converted into a greyscale color composition. Through this color arrangement, the amount of data in a single image could be further reduced by displaying the image in black and white colors. This will eventually make it easier to recognize the gesture explicitly and enhance the accuracy of the overall system. The captured images were converted into a grayscale image and are further resized. The collected data is fed into a model for the training process. Upon obtaining a considerable accuracy the weights of the models are saved in a file. Based on the trained weights the detection will take place where the given gesture in SSL will be translated into a text form.

6.4.3 Conversion of text to ASL

The latter part of the component is to translate the converted text to the relevant American sign. The overall text results are saved into a pickle file and are mapped with the corresponding American gestures. The results will be sent through a Flask API. Moreover, an avatar model is trained to denote the translated words. Ultimately the user will have a button click option to select which category needs to be translated. After the selection, the relevant Sri Lankan sign can be given as a camera input for translation. The translated gesture will finally be displayed as an avatar to the user.

6.4.4 Training the avatar model

The avatar models were trained using ‘DeepMotion’ and ‘3D Hand Draw’ applications. DeepMotion is a motion tracking application where it tracks bodily movements including hands and facial gestures. When a video clip of performing gesture was provided to the application, a prebuilt avatar model imitates the movements of the given video. The avatar movements were then recorded and taken into small video clips for each relevant gesture. 3D Hand Draw application was used to operate the hand movements of the gestures. Therefore, this was mainly used for the sign language gestures that were only indicated using hand movements. A pre-generated avatar was provided by the application. Using the provided avatar model small video clips were obtained by operating its’ hand and finger movements.

6.5 Search Module for SSL

6.5.2 Datasets

The RAVDESS emotional audio dataset will be used for the process of identifying the emotion. Since the dataset is too large initially a part of the dataset will be used to build the model and afterwards the full dataset will be used for final implementation.

For the conversion of text to SSSL a dataset will be created after gathering information from a few organizations and will be using a dataset which is already available on Kaggle.

6.5.3 Conversion to sign language

The audio of the given video will be extracted and the text that is extracted will be used as the source to convert the content to SLSL using NLP. The required data for the conversion of the SLSL will be taken from the Kaggle as they already exist a dataset about SLSL. It is planned to gather additional data from certain organizations. These data will be gathered by the team as it is essential to be used in every component of this research.

6.5.4 Speech Emotion Recognition

The extracted audio from the above mentioned will be used here to analyze the emotions using speech recognition. It is planned to use the RAVDESS emotional audio dataset as the data for emotional analysis. MFCC will be used for the feature extraction along with MLPC for this purpose of identifying the emotions.

6.5.5 Plan on Mapping emotion with SLSL

The plan on mapping the analyzed emotion with the converted SLSL would be to break the extracted text from the audio by sentences. Afterwards the text can be converted sentence wise and will be stored afterwards. The audio will be then broken into parts by sentence by sentence so it will be easier to map the identified emotion with the text. Finally, the identified emotion will be mapped with each sentence respectively.

6.5.6 Final Output

The final output will be presented as a 3D avatar signing in SLSL with the identified emotions integrated to the avatar as facial expressions. The plan is to achieve this by using Maya Autodesk.

6.6 Tools and technologies

Technology stack:

- For Object detection -Tensor flow
- For Video processing -OpenCV

- For Version controlling-GIT
- Frontend-HTML,CSS
- API-Flask

Programming Languages:

- Python –libraries: NumPy, sklearn, matplotlib,Tenserflow,cv2, Librosa, NLTK, VADER Sentiment, DeepFace, PyDub, MoviePy , SpeechRecognition

Tools:

- Google Colab
- Cuda Toolkit
- Media pipe

7 TESTING


Table 2: Test case 1


Test Case No 1		
Description	Check if the testing results are highly accurate as the accuracy score	
Input	Input 1	poses[np.argmax(res[1])] & poses[np.argmax(y_test[1])]
	Input 2	poses[np.argmax(res[1])] & poses[np.argmax(y_test[1])]
	Input 3	poses[np.argmax(res[1])] & poses[np.argmax(y_test[1])]
Expected output	Results should be equal for most cases	
Actual output	Result 1	Twelve & Twelve
	Result 2	Sixteen & Sixteen
	Result 3	Nineteen & Nineteen
Status	Pass	

Table 3: Test case 2

Test Case No	02
Description	Testing text to SSL conversion
Input	I Love Going to School
Expected Output	I school love go
Actual Output	<div> Input : I Love Going to School Output : ['i', 'school', 'love', 'go'] </div>
Result	Pass

Table 4: Months in a year

Test Case No	03
Category	The Alphabet
Description	Testing the conversion of a given letter in sign language to text
Input	
Expected Output	Detect the hand gesture and display the name as 'C'

Actual Output		
Test Status	Pass	

8 RESULTS AND DISCUSSION

8.1 Results

8.2.1 Teaching SSL and evaluation

Category 1:

The results are captured using the Tensor flow media pipe model. The multi-class classification models are defined level wise and number of words for each level is varied. Accuracy of each model is based on number of input features(classes) and overlap /uniqueness of gestures. The system identifies the gestures with higher testing accuracy if the number of classes/input features are lower and distinguishable.

Category 2:

Each dynamic sign is with involved several numbers of gestures. Hence unlike in category 1, one specific sign is identified as a series of gestures. This is a key consideration when developing the comparison algorithm for evaluation of each gesture involved for a specific sign.

Different Accuracy obtained for Trained models in Category 1 and Category 2 are listed in the below table 5.

Table 5: Comparison of Accuracies from Category 1 and Category 2

Category	Level	Number of Input Features	Accuracy	F1 Score
1	Animals	6	91.67%	0.9111
	Colours	9	94.44%	0.9404
	Numbers	10	100%	1.0
2	Light Colours	11	95.45%	0.96190
	Dark Colours	11	95.45%	0.96190
	Family Members	3	100	1.0

8.1.2 Sign Language Capturing module

8.1.2.1 Speech emotion recognition

To identify the emotion using speech four models were tested for accuracy to select the most suitable classification model. These model takes an audio file as input and analyzes the audio file to identify the emotion. The following table shows the different models used and the results of each model.

Table 6: SER Results

Model	Accuracy	F1 score	Precision	Recall
MLPC	79.03%	0.79	0.80	0.80
Random Forest	68.55%	0.69	0.70	0.70
Decision Tree	70.83%	0.68	0.69	0.69
Logistic Regression	61.8%	0.61	0.61	0.61

The accuracy of the build model was tested using a test set of data and the achieved accuracy was 79.55% with an F1 score of 0.79. Multi-Layer Perceptron Classifier (MLPC) was used for the classification since MLPC had the highest accuracy of the used models.

8.1.2.2 Text Emotion Analysis

To identify the emotion using text four models were tested for accuracy to select the most suitable classification model. These model takes text as input and preprocess the text and analyzes the to identify the emotion. The following table shows the different models used and the results of each model.

Table 7: Text Emotion Analysis Results

Model	Accuracy	F1 score
VADER Sentiment	84%	0.84
Random Forest	76.84%	0.73
Multinomial Naïve Bayes	66.33%	0.66
Logistic Regression	70.5%	0.69

The used technology for sentiment analysis was able to predict the emotion with an accuracy of 84% with a F1 score of 0.84. This was higher than the other predicted models hence were selected as the model for text emotion prediction.

8.1.2.3 Facial expression detection

A deep face library was used for the detection of emotion in facial expressions as this has an accuracy of 97% for the detection of a face and an accuracy of 80% for emotion detection. The combination of all was tested with a set of video files manually and has an accuracy of 78.9%.

8.1.3 SSL to ASL translation

In the categories that were chosen to train, the Alphabet, and the numbers [1-10] contain static gestures while the Months and Days include dynamic gestures altogether. Based on the category a suitable model was chosen and was able to obtain a good accuracy for each component. The following diagram displays the accuracies and the model that was used for each category of signs

Table 8 :Accuracy comparison for each category of words

Category	Type	Model	Accuracy
Alphabet	Static	CNN	84%
Numbers	Static	CNN	70.82%
Months	Dynamic	LSTM	79.82%
Days	Dynamic	LSTM	100%

8.2 Discussion

8.2.1 Teaching SSL and evaluation

Keras in TensorFlow is used to build a sequential neural network. The model is trained using 3 sets of LSTM layers followed by 2 dense layers. Relu activation function is used for the training of LSTM layers and first two dense layers and Softmax activation function is used for the training of last dense layer as can be seen in the figure 18.

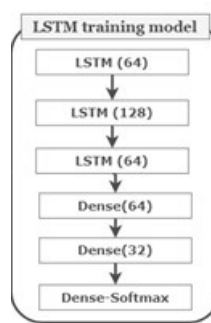


Figure 21: LSTM training model

All the models are trained with satisfactory accuracies as can be seen in table 5. All accuracies obtained are greater than the expected accuracy of 80%. Limiting the number of input variables for each model is the key for obtaining such high accuracies.

Animals (Category 1)

Model for ColorSign is trained with 300 epochs and with a categorical_crossentropy loss of 0.5222 and categorical accuracy of 0.8465.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 64)	82688
lstm_1 (LSTM)	(None, 30, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 6)	198
Total params: 237,350		
Trainable params: 237,350		
Non-trainable params: 0		

Family Members(Category 2)

Model for Family members is trained with 500 epochs and with a categorical_crossentropy loss of 0.3808 and categorical accuracy of 0.8626.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 64)	82688
lstm_1 (LSTM)	(None, 30, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 3)	99
Total params: 237,251		
Trainable params: 237,251		

Non-trainable params: 0

8.2.2 Sign Language Capturing module

At the initial stages of development, it was noticed that emotion couldn't be identified only using speech as it was planned. This was mainly since the person who is speaking may have a monotonous tone and hence the content may be positive, but the tone of the voice will get classified as negative due to this. Therefore, after further research it was decided to use a combination of speech, text, and facial expressions for emotion detection.

When comparing with other learning platforms that are developed for HI it was proved that there was no system exists with this functionality and focused on SSL. Therefore, this the accuracies achieved through functionality can be determined to be the best success rate.

8.2.3 SSL to ASL translation

- Alphabet

Trained using a CNN model using 2 sets of convolutional layers, 2 sets of Maxpooling layers, and 2 dense layers. The figure below shows a summary of the model including a count of trainable and non-trainable params for each set.

Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
conv2d (Conv2D)	(None, 62, 62, 32)	320
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 32)	0
flatten (Flatten)	(None, 6272)	0
dense (Dense)	(None, 128)	802944
dense_1 (Dense)	(None, 26)	3354
=====		
Total params: 815,866		
Trainable params: 815,866		
Non-trainable params: 0		

- Numbers [0-10]

Numbers belonging to the static sign category were also trained using the same CNN model as the Alphabet. It includes the same number of layers as mentioned above. However, the total parameters are slightly reduced compared to the previous model

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	320
max_pooling2d (MaxPooling2D)	(None, 31, 31, 32)	0
conv2d_1 (Conv2D)	(None, 29, 29, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 32)	0
flatten (Flatten)	(None, 6272)	0
dense (Dense)	(None, 128)	802944
dense_1 (Dense)	(None, 11)	1419
Total params: 813,931		
Trainable params: 813,931		
Non-trainable params: 0		

- Days

Trained using 3 sets of LSTM layers and 3 sets of Dense layers for action detection. The figure below is a summary of the trained model for the 'Days' category.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 64)	82688
lstm_1 (LSTM)	(None, 30, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 7)	231
Total params: 237,383		
Trainable params: 237,383		
Non-trainable params: 0		

- Months

Months of the year contains another set of dynamic gestures. It was also trained using 3 sets of LSTM layers and 3 sets of Dense layers. The parameter count for this model is slightly greater than the model trained for 'Days'

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 64)	82688
lstm_1 (LSTM)	(None, 30, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_2 (Dense)	(None, 12)	396

=====
Total params: 237,548

Trainable params: 237,548

Non-trainable params: 0

9.CONCLUSION

The research is objected to address the issues faced by HI in Sri Lanka with respect to learning SSL including limited learning systems for SSL .The research result is the implementation of Learning Platform "Hastha" providing an effective learning experience to learn SSL for the kids/adults with hearing impairments in Sri Lanka. The system is consisted of level based games and comprehensive evaluation at the end of each level guaranteeing effective learning experience for HI kids.Further, the system is comprised of a sign language capturing module and chat-bot feature to search gain additional knowledge in SSL for HI adults. Sign Language capturing module is capable of translating a YouTube video input to SSL, addressing the learning limitations faced by HI in Sri Lanka due to less learning resources.It is also capable of emotion analysis of the video feed.The system further facilitates SSL to ASL translation filling the gap in learning ASL for the HI in Sri Lanka.It is expected to gain more exposure to Online learning systems and better learning experience by the HI in Sri Lanka with the implementation of "Hastha".There are numerous possible improvements that can be made to the Hastha in future. One possibility is expanding the system to other sign languages since Hastha is currently developed only for Sri Lankan Sign Language.Further,the sign language capturing module can be expanded to convert content of any video to sign language.Moreover, the same method can be applied in different platforms such as in Social Media, in addition to Google search enabling users to navigate through people minds via social media content.

REFERENCES

- [1] "High frequency of hearing disorders in Sri Lanka," 29 March 2013. [Online]. Available: <https://www.hear-it.org/high-frequency-of-hearing-disorders-in-sri-lanka#:~:text=Nine%20percent%20of%20Sri%20Lankans,some%20sort%20of%20hearing%20disorder>

- [2] S. He, "Research of a Sign Language Translation System Based on Deep Learning," 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 392-396, doi: 10.1109/AIAM48774.2019.00083.

- [3] A. Kumar, K. Thankachan and M. M. Dominic, "Sign language recognition," 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), 2016, pp. 422-428, doi: 10.1109/RAIT.2016.7507939.

- [4] Herath, R.J., Ishanka, P. (2022). An Approach to Sri Lankan Sign Language Recognition Using Deep Learning with MediaPipe. In: Motahhir, S., Bossoufi, B. (eds) Digital Technologies and Applications. ICDTA 2022. Lecture Notes in Networks and Systems, vol 454. Springer, Cham. https://doi.org/10.1007/978-3-031-01942-5_45

- [5] C. Savage, "The importance of mother tongue in education," 30 August 2019. [Online]. Available: <https://ie-today.co.uk/comment/the-importance-of-mother-tongue-in-education/>.

- [6] R. Brooks, "A Guide to the Different Types of Sign Language Around the World", The Language Blog, 2018. [Online]. Available: <https://www.k-international.com/blog/different-types-of-sign-language-around-the-world/>

- [7] J. R. Ferreira Brega, I. A. Rodello, D. R. Colombo Dias, V. F. Martins and M. de Paiva Guimarães, "A virtual reality environment to support chat rooms for hearing impaired and to teach Brazilian Sign Language (LIBRAS)," 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), 2014, pp. 433-440, Available: 10.1109/AICCSA.2014.7073231.

- [8] B. Saunders, N. C. Camgoz and R. Bowden, "Everybody Sign Now: Translating Spoken Language to Photo Realistic Sign Language Video", arXiv preprint, 2020. [Online] Available: <https://arxiv.org/abs/2011.09846>

- [9] D. Kelly, J. McDonald and C. Markham, "A system for teaching sign language using live gesture feedback," 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1-2, Available: 10.1109/AFGR.2008.4813350 [Accessed 20 January 2022].

- [10] K. Fernando and H. Wickramaratne, 2018, August. "Sri Lankan Sign Language Tutor," presented at 1st International Conference on Business Innovation 2018.

- [11] W. S. S. W. Martianda Anggraeni, "Indonesian Sign Language (SIBI) Vocabulary Learning Media Design Based on Augmented Reality for Hearing-Impaired Children," [Online]. Available: <https://jurnaleccis.ub.ac.id/index.php/eccis/article/view/620>.

- [12] M. E. Anggraeni, I. Maulania and W. Sarinastiti, "Interactive Learning Media for Hearing-Impaired Children using Indonesian Sign Language (SIBI) — Simple Sentence Arrangement," 2020 International Electronics Symposium (IES), 2020, pp. 662-668, doi: 10.1109/IES50839.2020.9231955.

- [13] D. Dewasurendra, A. Kumar, I. Perera, D. Jayasena and S. Thelijjagoda, "Emergency Communication Application for Speech and Hearing-Impaired Citizens," 2020 From Innovation to Impact (FITI), 2020, pp. 1-6, doi: 10.1109/FITI52050.2020.9424899.

- [14] A. Almutairi and S. Al-Megren, "Augmented Reality for the Literacy Development of Deaf Children: A Preliminary Investigation," in *Proceeding of the 19th International ACM SIGACCESS Conference on Computer Accessibility*, Online, 2017.

- [15] J. R. Liddell, "American sign language: The phonological base," [Online]. Available: <https://muse.jhu.edu/article/507116/summary>.

- [16] D. Kelly, J. McDonald and C. Markham, "A system for teaching sign language using live gesture feedback," 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1-2, doi: 10.1109/AFGR.2008.4813350.

- [17] D. Manoj Kumar, K. Bavanraj, S. Thavananthan, G. Bastiansz, S. Harshanath and J. Alosious, "EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence", *2020 2nd International Conference on Advancements in Computing (ICAC)*, 2020, pp. 506-511. Available: 10.1109/icac51239.2020.9357154 [Accessed 20 January 2022].
- [18] Y. Perera, N. Jayalath, S. Tissera, O. Bandara and S. Thelijjagoda, "Intelligent mobile assistant for hearing impairers to interact with the society in Sinhala language," 2017.
- [19] I. Dissanayake, P. Wickramanayake, M. Mudunkotuwa and P. Fernando, "Utalk: Sri Lankan Sign Language Converter Mobile App using Image Processing and Machine Learning", *2020 2nd International Conference on Advancements in Computing (ICAC)*, 2020, pp. 31-36. Available: 10.1109/icac51239.2020.9357300 [Accessed 20 January 2022].
- [20] H. Amnur, Y. Syanurdi, R. Idmayanti and A. Erianda, "Developing Online Learning Applications for People with Hearing Disabilities", *JOIV: International Journal on Informatics Visualization*, vol. 5, no. 1, 2021. Available: 10.30630/joiv.5.1.457.
- [21] N. Krishnamoorthy, A. Raveendran, P. Vadiveswaran, S. Arulraj, K. Manathunga and S. Siriwardana, "E-Learning Platform for Hearing Impaired Students", *2021 3rd International Conference on Advancements in Computing (ICAC)*, 2021, pp. 122-127. Available: 10.1109/icac54203.2021.9671113 [Accessed 20 January 2022].

- [22] DailyFT, "Sinhala Sign Language the main communication mode for the Deaf in Sri Lanka", p. Single, 2019. [Online]. Available: <https://www.ft.lk/Opinion-and-Issues/Sinhala-Sign-Language-the-main-communication-mode-for-the-Deaf-in-Sri-Lanka/14-671078>
- [23] Prof. J. Booth, "How do deaf or hard of hearing children learn to read?", BOLD, 2019. [Online]. Available: <https://bold.expert/how-do-deaf-or-hard-ofhearing-children-learn-to-read/>
- [24] B. Racoma, "Why Do Sign Language Interpreters Make Faces?", eTranslation Services Blog, 2021. [Online]. Available: <https://etranslationservices.com/blog/translations/why-do-sign-language-interpreters-make-faces/>

LIST OF APPENDICES

Appendix A : Survey Questions

What do you think is the best way to display translation of videos? (විඩියෝ පරිවර්තන පෙන්වීමට හොඳම ක්‍රමය කුමක්දැයි ඔබ සිතන්නේ කුමක්ද?)

- ☐ Text as subtitles (උපසිරැසි ලෙස පෙළ)
- ☐ Sign language through an animated avatar along with the video (විඩියෝව සමඟින් සංඥා භාෂාව සජීවීකරණ භරතා)
- ☐ Sign language through an animated avatar only (සංඥා භාෂාව සජීවීකරණ avatar භරතා පමණක්)

How likely are you to use a system that is teaching Sri Lankan sign language through an automated platform with interactive features and animated avatars ? 1 being highly unlikely , 10 being highly likely (අන්තර්ක්‍රියාකාරී විශේෂාංග සහ සජීවීකරණ avatar සහිත ස්වයංක්‍රීය වේදිකාවක් භරතා ශ්‍රී ලාංකේය සංඥා භාෂාව උගන්වන පද්ධතියක් භාවිතා කිරීමට ඔබ කෙතරම් දුරට ඉඩ තිබේද? 1 බොහෝ විට නොහැක්කකි, 10 බොහෝ දුරට ඉඩ ඇත)

- | | | | | | | | | | |
|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Is it useful if the hearing impaired people are able to get real time feedback on the mistakes in gestures when interpreting words in sign language? (ශ්‍රවණාබාධිත පුද්ගලයන්ට සංඥා භාෂාවෙන් වචන අර්ථකථනය කිරීමේදී අභිනයන්වල ඇති වැරදි පිළිබඳව තත්ත්‍ය කාලීන ප්‍රතිපෝෂණ ලබා ගත හැකි නම් එය ප්‍රයෝජනවත්ද?) *

- ☐ Yes (ඔව්)
- ☐ No (නැත)

How would you expect the system to provide feedback on the answer they have given? (ඔවුන් ලබා දී ඇති පිළිතුර පිළිබඳව පද්ධතිය ප්‍රතිපෝෂණ ලබා දෙනු ඇතැයි ඔබ අපේක්ෂා කරන්නේ කෙසේද?)

- ☐ State only the correct answer without any feedback on the answer. (පිළිතුර පිළිබඳ කිසිදු ප්‍රතිපෝෂණයකින් තොරව නිවැරදි පිළිතුර පමණක් සඳහන් කරන්න.)
- ☐ State only if the answer is correct or wrong (පිළිතුර නිවැරදි හෝ වැරදි නම් පමණක් සඳහන් කරන්න)
- ☐ State what percentage of answer is correct or wrong along with the mistake they have done and the correct answer (කර ඇති වැරද්ද සහ නිවැරදි පිළිතුර සමඟ නිවැරදි හෝ වැරදි පිළිතුරේ ප්‍රතිශතය කොපමණද යන්න සඳහන් කරන්න)

Do you think it would help the hearing impaired if content of youtube videos to be translated to Sri Lankan sign language to gain knowledge? (ශ්‍රවණාබාධිත අයට දැනුම ලබාගැනීම සඳහා යු ටියුබ් විඩියෝවල අන්තර්ගතය ශ්‍රී ලාංකේය සංඥා භාෂාවට පරිවර්තනය කළහොත් එය උපකාරයක් වේ යැයි ඔබ සිතනවාද?) *

- ☐ Yes (ඔව්)
- ☐ No (නැත)