E-Learning Platform for Hearing Impaired Students

Niroshan Krishnamoorthy
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it18144772@my.sliit.lk

Accash Raveendran
Department of Data Science
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it18069600@my.sliit.lk

Pirathikaran Vadiveswaran
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it18068610@my.sliit.lk

Sangeeth Raj Arulraj
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it18152074@my.sliit.lk

Kalpani Manathunga
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
kalpani.m@sliit.lk

Samanthi Siriwardana
Department of Information Technology
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
samanthi.s@sliit.lk

***Abstract*— With the Spread of global pandemic Covid-19, the learning was transformed to online from traditional learning. The use of e-Learning platforms was increased. But this idea has issues with certain communities of people around the world. The hearing-impaired people have many issues with eLearning platforms because of their deficiency in hearing sound. Therefore, through this paper we are introducing a platform through which hearing impaired people can effectively involve in learning. The proposed system uses sign language in addressing the students. We also introduce ways on which hearing impaired students can communicate with the tutors. The system has a low light enhancement module to enhance the videos uploaded by the tutor, module to convert the uploaded videos to American Sign Language and it also converts the questions asked via sign language to text.**

Keywords— Low Light Detection; Low Light Enhancement; Cumulative Histogram; ASL; Speech Recognition: Tokenization; Stemming, RCNN

I. INTRODUCTION

In the year 2020 the world encountered a global pandemic problem with the spread of COVID-19 virus. This pandemic situation transformed many of the industries to online basis with the use of Internet. This new transformation of Industries to online was quickly adapted by the people around the globe. One such sector which transformed to online was the education sector where students started learning through online platforms. Even though this transformation was effective in continuing the learning, some group of people encountered lots of difficulties compared to traditional learning. One such group of people is the Hearing-Impaired people. In [1] the author states that the study made by WHO had suggested that approximately 466 million of total population around the world has some sort of hearing deficiency in 2018. This is a total of 6.1% of the world's population. Out of this, 432 million are adults and 34 million are children.

Initially we conducted a survey with the hearing-impaired community regarding an implementation of an e-learning environment. This survey was conducted at a deaf school in Sri Lanka and we received 75 responses. According to Fig. 1. 60% of the response preferred to have lectures in sign language on the e-learning platform. This showed us the importance of having an eLearning system using sign language.

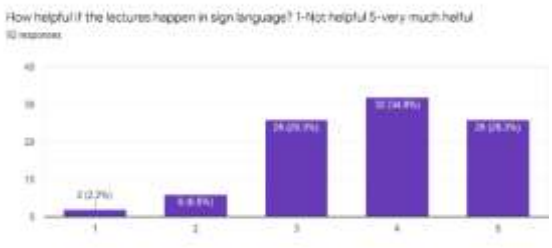

Fig. 1. Survey result

Based on the survey that is conducted, we have clearly identified where the hearing-impaired community is facing their learning problems and how we can utilize the technological advancements to provide a better solution to the problems that are faced by the hearing-impaired students.

II. RELATED WORK

Some Authors have worked on providing solutions for hearing impaired community with various technologies in the past, some of those related works are discussed here.

According to [2] the authors have discussed about using the human computer interaction approach to propose a new eLearning interface with interactional features for the use of students with varying visual and hearing needs. The proposed system is useful for visual/hearing impaired students. The assistive adaptive features used in this technology for hearing impaired students are providing sign language when the user places the cursor over the text content in the page, pre-recorded sign language videos, sign language features to read command in toolbar, explanation of graphs using sign language. In another work, authors have also proposed solutions using modern technologies which can be used in day today tasks. In [3] the authors have proposed an interpreter system that can be used as an Android application. This application can convert the sign language into Normal speaking language. The proposed model was successful for conveying messages from deaf people to others.

We also planned to add low light enhancement for the uploaded videos in our system because it was identified that low light videos can cause some problems for hearing impaired students. This is elaborated more in section III A. To identify low light videos we gathered information from some relevant past works conducted by various authors. Out of them the Authors of [4] have proposed an algorithm for the enhancement of low-light videos. The algorithm is first inverting an input low light video and then applying the optimized image de-haze algorithm on the inverted video.

Simulations results of this algorithm shows good enhancement results when compared to other frame wise enhancement algorithms. In the work [5], the authors have proposed a real time video enhancement technique for videos with complex conditions like insufficient lighting. This method provided a better approach to enhance the video in low-lighting conditions without any loss of color. This algorithm is providing effective enhancement using simple computational procedures. The results were analyzed on the videos that were taken on the bad light conditions. But this approach doesn't provide any evidence on the effectiveness of this algorithm in normal light.

When creating a sign language interpreter for our system, we identified the importance of captioning techniques in the preprocessing part of the application. Therefore, we considered some important past research on the captioning techniques used on eLearning systems. According to [6], Authors discuss about two speech recognition techniques. They are using real time captioning using IBM ViaScribe and Post lecture transcription using IBM hosted transcription service. The main processes of these two techniques are the process of recording instructor's speech, captioning the methods using a human captioner or providing transcript using the human transcript services and finding errors of the transcripts or captions. These captioning techniques can now be utilized using modern cloud solution.

After captioning, we had to include some text-to sign language conversion techniques. For text to sign language conversion, a few research works performed in the past includes of direct translation and rule-based translation. In the Direct translation approach the English words are directly converted into sign language sequence, regardless of the meaning of the sentence. This is highlighted as a major defect in the research. Further, as in the rule-based approach [7], the user given text input is processed under syntactic transformation and it is converted into a median text. Rule based approach can be considered as a successful approach when translating text to sigh language.

When we are creating an eLearning system for hearing impaired students, they should be able to use the system desirably. This brought us the challenge of identifying hand gestures using our system. We looked on to some of the systems that were developed for this challenge. The proposed system [8] focuses on recognizing ASL alphabets and double-handed gestures for deaf and dumb people. In their system, they have used 4 main components which are real-time hand tracking, hand segmentation, feature extraction and gesture recognition. The camshaft method and Hue-saturation Intensity (HSV) color model was used for hand tracking, gestures detection and segmentation. In the work [9], a proposal for deaf and dumb people to communicate with ordinary people using a framework that recognizes hand gestures was introduced. Their approach was first to take an image applying skin segmentation using Hue-Saturation-value (HSV) histogram and finding edges and then applying Harris Algorithms for feature extraction. The next steps on their work were feature matching and recognition where they calculate the minimum value of the matrix. Also, they have used skin segmentation techniques to detect the hand gestures here. The proposed system [10] focuses on making the dual way communication between hearing-impaired and normal people. But they identified some issues on dynamic hand gestures that give the same meaning. Pre-processing, segmentation, feature extraction, and classification are the techniques they have used. For Preprocessing they have used the Gaussian Blur to reduce the noise in the images. Segmentation was applied to find hand gestures regions from the image area. For feature extraction, they have used Eigenvalues and Eigenvectors which can be used to classify hand gestures.

With All the above-mentioned studies, some of the key features and functionalities for a solution on hearing impaired students' inability to learn through eLearning system was proposed in the next section.

III. METHODOLOGY

We have created an eLearning system for the hearing-impaired students using four main functionalities. These individual functionalities act as the solutions for some of the problems faced by the hearing-impaired students in eLearning systems. Each Module in the Fig. 2. represents the individual functionalities.

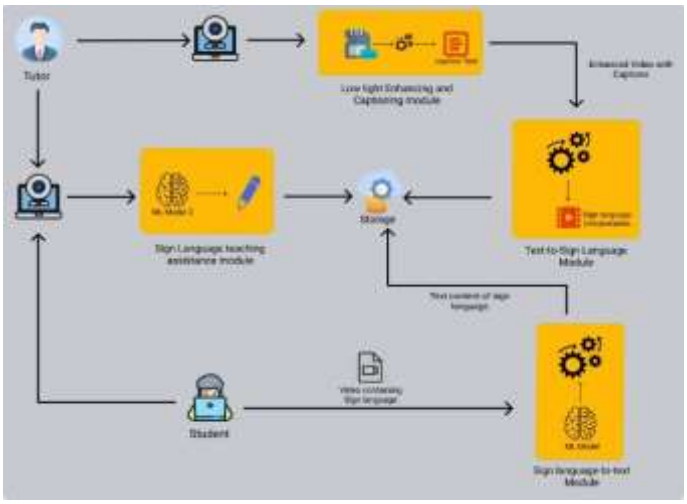

Figure. 2. System Overview Diagram

A. Low-light enhancement captioning module.

This module is responsible to identify low light parts of a video and enhance them. After enhancing the low light parts, the speech in the video is converted into text and added as the subtitle text for the video. This subtitle text is used in the text to speech module to produce sign language interpretation.

a) Low-light identification and enhancement

This module was decided to be implemented with respect to the responses from the survey.

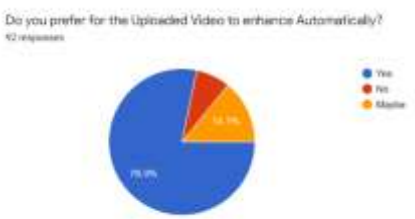

Fig. 3. Video enhancement response

According to the Fig.3 majority of the people preferred enhancing the uploaded video because the low light videos hinder some valuable information when learning through

videos. Therefore, we used a technique to identify the low light videos automatically using some basic computation procedures and we enhanced them using low light enhancement techniques.

To identify the low light parts of the video, we first converted the video from color to gray scale because it is easy to do computation on the gray scale rather than using a color scale like RGB. Then we constructed cumulative histograms for each frame of the video. A cumulative histogram gives the cumulative number of pixels at each intensity level on each video frame or image. According to [11] Authors have identified that intensity distribution of images captured under low light conditions shows that more than 50% of total number of pixels fall under low intensity values and for day light condition more than 50% of total number pixels fall under high intensity values. Based on this theory, we selected a threshold value from the intensity values which can separate dark intensity values and bright intensity values using a gray intensity scale.

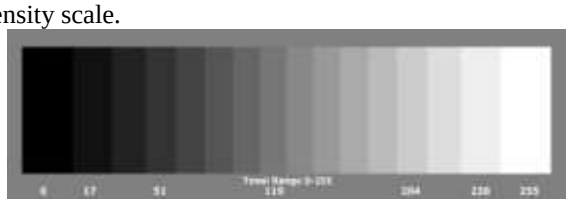

Fig. 4. Gray intensity scale

Based on Fig. 4. we selected an intensity value which is >100 and <120 as our threshold value. This value is then applied to the cumulative histogram that is constructed for each frame. If the cumulative pixels percentage for this intensity value is >=75% of the overall cumulative pixel percentage, then this identified as a low light frame else it is identified as a bright light frame.

Once the low light frames are identified, median filter is used to denoise the low light frames and the following gamma correction equation is used to improve the contrast of the low light frames.

$$O = C * I \wedge (1 / G) \qquad (1)$$

In equation (1) I is the input image and G is the gamma value. C is a constant and it is 1. The output O is then scaled back to the intensity range [0,255]. In our solution a gamma value(G) of 2.0 is used to enhance the low light frames. Once this is completed the pre-processing of the uploaded video is completed. The processed video is then saved on the disk for the next processing.

b) Captioning

For this purpose, we are using the Google Speech to text model, because it is proved that they have better performance compared to other Automated Speech Recognition systems [12]. First, we extracted the audio from the enhanced video. This audio is next sent into the Google's Speech to Text model. We got the transcription of the audio as the output of this model. This transcript file is then sent into an Algorithm which can separate its words into six words sentences and its respective timestamps. Once the subtitles are created it is saved in a .srt file and used as the subtitles for the uploaded video.

B. Text-to-Sign language Module

a) Data collection

For this module the data was collected from the available online sources and the source dataset was retrieved from the captioning module mentioned above. The American Sign Language videos were downloaded manually and put into respective folders with the names of the English words. Unrelated or duplicated videos were deleted by manually checking them.

b) Solution design

The process of this module has four sub modules.

1. Tokenization
2. Stop words removal
3. Stemming
4. Output video conversion

The tokenization module is responsible of breaking the sentence into words by selectively choosing only the meaningful words that needed to be converted into sign language. The stop words removal module is made to eliminate all the common words that usually appear in most documents or contexts, like the articles (a, an, the), prepositions (of, from, on), models (could, may, can), conjunctions (and, for, but), etc. The stemming module is created to reduce a word to its root. In other words, the words that are given as an adverb or an adjective, are converted into its root word.

E.g. Words such as completely, completed, completion were reduced to its root word that is '*complete'*.

Followed by these modules, the remaining words from the captions retrieved from the captioning module are taken and matched with the video dataset. Finally, in the output video conversion module, a video sequence is produced for the given English sentence, resembling the ASL.

a) Development language

The implementation of this text-to-sign language module was done using Python.

b)Tool / Libraries

Some tools and libraries like NumPy, movies.py were used here.

C. Sign Language-to-Text module

This functionality is to convert sign language into text and make a meaningful sentence. Deaf and dumb student can ask questions by uploading their questions as a video file. Then the system saves the video and does the relevant steps. Video pre-processing is the first thing done by the system. Pre-processing has three steps in the video. The first step is converting the video into a frame by frame image, the second step is to adjust the contrast and the final step is to resize the image. After video pre-processing image background removal, the next step is to convert the image into binary form, the following step is the feature extraction, and the later step is to do the gesture recognition and finally taking output text, fine-tune and save it into the database.

a) Dataset collection

Our research mainly focus on the American Sign Language. Therefore we used the American Sign Language data set from Kaggle, Microsoft-ASL and also form our own data set Fig. 5.

Fig. 5. Dataset Sample

b) Video pre-processing

There are three subprocesses in this process. The steps are to convert video into frame by frame, adjust contrast, and resize images. First of all, the system takes video from the database and start video processing. In the first step of video pre-processing, video input is converted frame by frame and stored as sequences of images. The second step is to take images one by one to analyze and adjust the contrast according to the requirement. The last step is to resize the image. This step maintains a unique size of image and resolution for all images. It deduces the analysis time for the calculations.

1) Removal background object

After the video pre-processing, the image is taken for the next step. In this step, image's background and objects are removed. It makes the image more definition to identify the hand region.

Fig. 6. Removal background object

2) Feature Extraction

The feature extraction is used to pick up certain features from the unique hand image for each identity. We used HOG transformation to extract feature from the frame. The Histogram of Oriented Gradients (HOG) is a feature descriptor for object recognition in computer vision and image processing. This method counts the number of times a gradient orientation appears in a certain area of a picture.

3) Classification

Classification is a method after feature extraction that recognizes hand gestures with various hand gesture images. We used here Stochastic Gradient Descent (SGD) classifier. The SGD optimization technique is used to determine the values of parameters/coefficients of functions that minimize a cost function. After the classification, outputs are saved into the specific database location.

D. Sign Language teaching assistance module

In this module we are educating sign language to hearing-impaired students and also this can be a great opportunity for the general user to learn and improve their knowledge in sign language. This is an easy mode of learning sign language remotely from their premises. The LMS helps learning from basic and test their knowledge in what they have studied so far in the module, for the user will need a personal computer, webcam, and internet connection. The user will be given basic lessons such as alphabet letters in American sign language and once the task begins he/she is asked to repeat as per the tutorial displayed in the system that shows the sign language of alphabet letters. The user should do that in front of the webcam as in Fig. 7 which shows how the system detects the sign of the user.

Fig. 7. How the system detecting sign

In the preprocessing segment using OpenPose user's motion is detected and send to the ML model where it is analyzed [13]. If the motion of the user is correct, then it will be saved and the user is taken to the next task but if he/she is performing it incorrectly, then the student is asked to repeat the task again.

This module of the system has two sub parts which are,

1. Annotate Hand Gesture
2. Image Classifier.

In the Annotate Hand Gesture, we use Faster Region Based Convolutional Neural Networks(RCNN) configuration. It is used to implement this feature on the TensorFlow ML model trainer. The teaching process is based on low-resolution imagery which benefits with faster training of ML model, storage efficiency and low latency network connections (low internet speed). Once the images are detected they are separated has training dataset segment and testing dataset segment using PASCAL VOC labeling tool and XML (extensible markup language) record is obtained for each image. To the train and test image datasets, images are manipulated into CSV (Comma Separated Value) files. Using Faster RCNN and TensorFlow 2 Detection Model Zoo we train the ML model with the dataset and using the Convolutional Neural Networks(CNN). Based on that we can process different images and we can categorize them accordingly. The *'Keras'* library is used to implement CNN ML model. Dataset of American sign language with 26 classes which is the alphabet letters from A-Z, each class trained with minimum 500-750 images. The Fig. 8. shows the trained dataset. After successfully training the ML model, we tested the model with webcam to detect images using trained ML model [14].

Fig. 8. Trained Dataset

The image goes through different stage in CNN classifier; Convolutional Layer, Nonlinearity and Pooling Layer. Nonlinearity is a term used in statistics to describe a situation in which an independent variable and a dependent variable do not have a straight-line or direct relationship. Changes in the output are not proportional to changes in any of the inputs in a nonlinear relationship. Using pooling layers, the feature maps' dimension are reduced. The number of time parameters used to learn and the amount of computation in the network are both reduced as a result. The pooling layer sums up the features presented in an area of the feature map formed by a convolution layer. Once the filtration is satisfied the system notifies that the user passed the task [15].

IV. Discussion & Results

A. *Low-light identification, enhancement captioning module.*

To test this functionality, low images, bright light images and low light videos were used. Following are the results of what we received after processing.

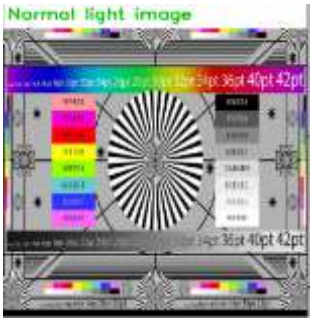

Fig. 9.

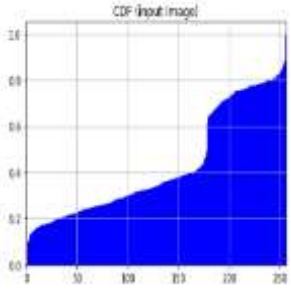

Fig. 10

Fig. 11.

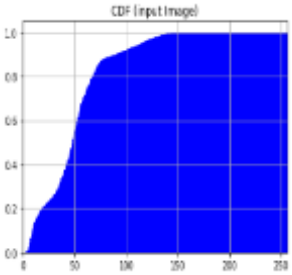

Fig. 12.

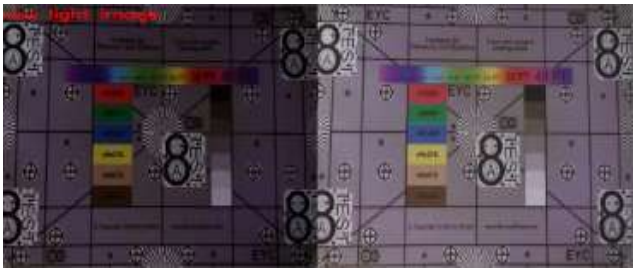

Fig. 13.

For this test, we selected our threshold value as 105 Fig.10. shows the cummulative histogram of Fig.9 and it is identified as a normal light image because the cummulative number of pixels at the intensity level 105 is <75% of the oveall number of pixels of that image. Fig.12 shows the cummulative histogram of Fig.11. and it is identified as a low light image because the cummulative number of pixels at the intensity level 105 is >=75% of the overall number of pixels of that image.

Fig.13 shows the enhanced output of a low light image. Left of fig shows the low light image and the right of that image shows the gamma corrected version of that. Finally this test was conducted with the low light videos and the low light videos were enhanced.

B. *Text-to-Sign language module.*

In order to test the accuracy of this functionality, we gave some test cases and the system translated it into ASL grammar format. The results were tested according to the standard. Table. I. shows the conversion of natural language grammar to ASL grammar.

Verb Pattern	Rule	Input Sentence	Parsed Sentence	Output Sentence
verb + object	VP NP	go school	(VP (VB Go) (NP (NN school)))	school go
subject + verb	NP V	birds fly	(NP (NNS birds)) (VP (VBP fly))	birds fly
subject + verb + subject complement	NP V NP	his brother became a soldier	(NP (PRP$ his) (NN brother)) (VP (VBD became) (NP (DT a) (NN soldier)))	his brother a soldier became
subject + verb + indirect object + direct object	NP V NP NP	i lent her my pen	(NP (FW i)) (VP (VBD lent) (NP (PRP her)) (NP (PRP$ my) (NN pen)))	i her my pen lent

Table. I: Standard grammar to ASL grammar

C. *Sign-to-Text Module.*

We tested this functionality with a machine learning model and Fig. 14. shows the Cohen kappa score accuracy of 0.803 and it predicted 80% of images according to the relevant label.

	precision	recall	f1-score	support
resizedrink	1.000000	1.000000	1.000000	5.000000
resizehelp	1.000000	1.000000	1.000000	15.000000
resizehome	0.500000	0.200000	0.285714	5.000000
resizehow	0.750000	1.000000	0.857143	6.000000
resizeno	0.636364	0.700000	0.666667	10.000000
resizewhat	0.857143	0.923077	0.888889	13.000000
resizewhen	0.928571	0.866667	0.896552	15.000000
resizewhere	0.800000	0.800000	0.800000	10.000000
resizewhich	0.800000	0.923077	0.857143	13.000000
resizeyes	0.500000	0.333333	0.400000	6.000000
accuracy	0.826531	0.826531	0.826531	0.826531
macro avg	0.777208	0.774615	0.765211	98.000000
weighted avg	0.814644	0.826531	0.814130	98.000000

```
metrics.cohen_kappa_score(y_test,y_pred_test)
```

```
0.8034218289085546
```

Fig. 14. Model Accuracy

Cohen's kappa is a popular measure for determining how well two raters agree. It may also be used to judge a classification model's effectiveness. According to Cohen, values of $\leqslant 0$

indicate no agreement, 0.01 – 0.20 indicate none to the sparse agreement, 0.21 – 0.40 indicate fair agreement, 0.41 – 0.60 indicate moderate agreement, 0.61 – 0.80 indicate significant agreement and 0.81 – 1.00 indicate virtually perfect agreement.

D. Sign Language teaching assistance module.

We tested this functionality in different testing circumstance such as in complex background, bright light environment and dark light environment, and we got expected out from the system. Table. II. shows the result of test the functionality.

Test no.	Test Scenario	No. of test runs	Mean Detection Accuracy (%)
1	Dark background	15	85
2	Bright environment	15	81
3	Complex environment	15	89

Table. II. Hand detection Test Result

When user interacts with the system, the webcam detects the user's hand notation and with the help of faster RCNN, we can get better real-time output and with trained ML model, the system will analyze user hand notation and compare whether it is similar with the system request. This ML gives above 80% average accuracy level, so that we can consider this as an optimal component. If the user follows the correct instruction and repeat as per the system request, he/she can pass the stage. We initialized the system with alphabetic letters because the user begins their course as a new learner. The user will be continuing the course with alphabets in chronological order.

V. CONCLUSION

In this paper, we discussed the functionalities used in our e-learning system which was developed for hearing impaired students. We showed that these methods were successful in achieving the overall goal of the system. From identifying low light parts of the uploaded video using cumulative histogram to enhancing using gamma correction, everything showed positive results. Currently the system will only output images of signs in the translating section. Using these techniques, we can use animations to output sign gestures in the future. For teaching sign language section, we can add quizzes and other teaching techniques in the future. We also added a section where hearing impaired students can clear their doubts through the system by raising questions using sign language.

ACKNOWLEDGMENT

This research was supported by the Sri Lanka Institute of Information technology CDAP module.

REFERENCES

[1] A.C. Davis and H. Hoffman, "Hearing loss: rising prevailance and impact," NCBI, 1 October 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6796666/.

[2] W. Farhan and J. Razmak, "A comparative study of an assistive e-learning interface among students with and without visual and hearing impairments," in *Disability and Rehabilitation Assistive Technology*, 2020.

[3] G. Nath and V. Anu, "Embedded sign language interpreter system for deaf and dumb people," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems*, 2017.

[4] Jiangto, Wen and W. Li, "An Efficient and Integrated Algorithm for Video Enhancement in Challenging Lighting Conditions," *Computer Vision and Pattern Recognition*, 2011.

[5] G. Mittal, S. Locharam and S. sasi, "An Efficient Video Enhancement Method Using LA*B* Analysis," in *IEEE International Conference on Video and Signal Based Surveillance*, 2006.

[6] R. Ranchel, T.-D. Teresa, Y. Guo, K. Bein, H. Martin, J. P. Robinson and B. S. Duerstock, "Using speech recognition for real-time captioning and lecture transcription in the classroom," *IEEE Transactions on Learning Technologies*, vol. 6, no. 4, pp. 299-311, 2013.

[7] Z. Tmar, A. Othman and M. Jemni, "A rule-based approach for building an artificial English-ASL corpus," in *2013 International Conference on Electrical Engineering and Software Applications*, Hammamet, 2013.

[8] A. S. Ghotkar, K. Rucha , K. Sanjana , A. Surbhi and H. Mithila , "Hand gesture recognition for Indian Sign Language," in *International Conference on Computer Communication and Informatics*, Coimbatore, 2012.

[9] A. Sood and M. Anju , "AAWAAZ: A communication system for deaf and dumb," in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, 2016.

[10] S. Rathi and G. Ujwalla , "Development of full duplex intelligent communication system for deaf and dumb people," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Noida, 2017.

[11] M. Kim, D. Park, D. K. Han and H. Ko, "A novel framework for extremely low-light video enhancement," in *IEEE International Conference on Consumer Electronics (ICCE)*, 2014.

[12] I. Maglogiannis, "A Benchmarking of IBM, Google and Wit Automatic Speech Recognition Systems," in *Artificial Intelligence Applications and Innovations.*, 2020.

[13] X. Hu, L. Tan, J. Zhou, S. Ali, Z. Yong, J. Liao and L. Liu, "Recognizing Chinese Sign Language Based on Deep Neural Network," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Toronto, Canada, 2020.

[14] M.G.Grif, "Approach to the Sign Language Gesture Recognition Framework Based on HamNoSys Analysis," in *XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*, Novosibirsk, Russia, 2018.

[15] D. M. Kumar, K. Bavanraj, S. Thavananthan, G. M. A. S. Bastiansz, S. M. B. Harshanath and J. Alosious, "EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence," in *2nd International Conference on Advancements in Computing (ICAC)*, Malabe, Sri Lanka, 2020.