International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 62-68, January-2024

https://doi.org/10.47001/IRJIET/2024.801008

Machine Learning Based Customized Solution for Deafness and Mute People

[1]Suranjini Silva, [2]Thilini Jayalath, [3]Madhushani E.A.Y.C., [4]Amarasooriya H.D., [5]Elpitiya S.N., [6]Frank Perera

[1,2,3,4,5,6]Department of Information Technology, Sri Lanka Institute of Information Technology, New Kandy Road, Malabe, Sri Lanka

***Abstract* - Individuals with hearing or speech impairments often possess a higher degree of familiarity with sign language due to its widespread usage. Consequently, a noticeable communication divide exists between those with such impairments and the general populace. A pivotal solution in bridging this gap involves employing human sign language interpreters. Regrettably, the scarcity of sign language interpreters worldwide in comparison to the number of individuals with hearing or speech impairments has resulted in certain individuals being unable to bear the cost of a human interpreter each time they engage in conversation. To address this issue, it is imperative to automate communication in a manner that liberates the deaf community from dependence on human translators. This article centers on the development of an application capable of real-time conversion between American Sign Language and text, along with supplementary functionalities aimed at dismantling the communication barriers faced by individuals with hearing or speech challenges when interacting with the broader public. the primary objective of this application is to accurately perceive and interpret the user's sign language. The initial stage involves training the system to recognize and interpret signs utilizing algorithms for object recognition and motion tracking. For this purpose, a convolutional neural network model was trained employing a meticulously curated American Sign Language dataset. Subsequently, the recognized signs are translated into English, forming grammatically sound words. Incorporated within the system are a language translator and a virtual sign keyboard, augmenting its capabilities. Moreover, a text-to-text transformer, structured on an encoder-decoder framework, is harnessed to identify grammatical inaccuracies and generate coherent phrases. For a comprehensive understanding of the process, detailed elaboration will follow in subsequent sections.**

Keywords: American Sign Language, Convolutional neural network, Grammatically correct sentences, Object Detection, Real-time, Speech and hearing impairments, Voice-to-sign language.

I. INTRODUCTION

Today, a sizable proportion of the population suffers from hearing loss or deafness. According to the World Health Organization (WHO), these illnesses impact around 400 million people. Given this large number of people, the need for a real-time sign language translator is critical. Various sign language interpreters are currently engaged to meet the communication needs of both the deaf and mute community and the public. Although verbal communication is the most common, not everyone can engage in it. Exclusion from our societal structure, on the other hand, is not a feasible answer. These deficits can be caused by a variety of circumstances.

There are between 138 and 300 unique sign languages in use worldwide. Sign language communicates mostly through hand movements and facial expressions. Each sign language has its own grammar rules and sentence patterns. Notably, the word order of phrases differs between English and American Sign Language. Furthermore, sign language only includes basic English word signals, with no prepositional markers. These distinctions are critical in the development of an efficient sign language translator that promotes communication between sign language users and people who are not familiar with sign language.

Artificial intelligence (AI) and machine learning (ML) models are currently being used to translate Sinhala sign language into textual and aural outputs. The methods used to facilitate seamless communication between the impaired community and the broader publics are described in this article. Case studies completed by Ezhumalai P, Raj Kumar M, Rahul A S, Vimalanathan V, Yuvaraj, and Mr. Mahesh Kumar, among others, were used in the study. For gesture recognition, techniques such as Latent Dirichlet Allocation (LDA) in conjunction with Naive Bayes (NB) are used. To extract characteristics from the dataset, MATLAB is used, and convolutional neural networks (CNN) are used for early analysis of sign language datasets possible to infer that the model attained an average accuracy of 97% during testing, demonstrating excellent precision, particularly when examining similar indicators in real-time circumstances. Although this program was initially built as a web application,

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 62-68, January-2024

https://doi.org/10.47001/IRJIET/2024.801008

there are plans to expand it into a mobile platform in order to overcome barriers and reduce processing time.

II. LITERATURE SURVEY

As is commonly acknowledged, a large number of scholars from all around the world have been conducting considerable research into sign language translation. Omkar Vedak, Prasad Zavre, Abhijeet Todkar, and Manoj Patil of the Department of Computer Engineering at Datta Meghe College of Engineering, Mumbai University, India, conducted research titled "Sign Language Interpreting Research Using Image Processing and Machine Learning." Their research aims to develop a system for translating Indian Sign Language into English.

The suggested system used a webcam to take photos of sign lettering, which were then subjected to sophisticated feature extraction via processing. They developed a specific algorithm to facilitate the translation of sign language into English text, with facilities for eventual conversion into speech. However, it should be noted that this method was limited to only 26 pre-registered characters, resulting in a limited range of hand gestures. The system's accuracy was roughly 88%. The targeted scope of the application included both mobile and web applications.

Mr. Mahesh Kumar investigated the use of Linear Discrimination Analysis (LDA) in conjunction with the Naive Bayes (NB) approach for gesture identification. This project involves feature extraction from datasets using MATLAB as well as fundamental study on sign language datasets using the CNN algorithm. Notably, this method produced satisfactory training and testing results. The suggested approach includes analyzing images using selfie sign language and testing them using a technique known as stochastic pooling. The CNN model-trained dataset outperformed the other approaches under consideration, with an output accuracy of 92.88%.

As evidenced by research undertaken in Sri Lanka, advances in Artificial Intelligence (AI) and Machine Learning (ML) were used to translate Sinhala Sign Language into both textual and aural outputs. This program aimed to make it possible to translate spoken speech into Sinhala Sign language. The creation of the translator was divided into four stages: learning to recognize hand motions using pre-trained algorithms, classification and translation of recognized hand signals, usage of image classifiers, and finally presenting the discovered character as written or aural output. This translator, known as "Easy Talk," aims to improve communication between the disabled community and the public.

At RMD Engineering College's Department of Computer Science and Engineering, Ezhumalai P, Raj Kumar M, Rahul A S, Vimalanathan V, and Yuvaraj A started a study to create a speech-to-sign language translator for the hearing impaired. Their strategy included the use of natural language processing (NLP) techniques to build a system that could translate spoken or written English into sign language. This entailed employing speech recognition technology to transcribe the audio input into text, which was then converted into sign language. They used Python, OpenCV for image and video processing, Pyaudio for speech recognition, and Tkinter for the graphical user interface with the main objective of increasing access to Indian Sign Language through speech translation. Although previous projects had some technological similarities to our system, it's important to note that their user interfaces lacked the interactive elements required by modern technology.

III. METHODOLOGY

There are five crucial elements that must be present for the suggested system to function well and be implemented successfully in order to enable smooth communication. The following is a list of these elements:

1) Detecting and identifying Sign language and converting it to text.
2) Convert Text into sign Language.
3) Convert Sign Language into voice.
4) Detect Voice from Ordinary Person & Convert the Message into Text form.

A) Detecting and Identifying Sign language and converting it to text

The first step is for the camera to be able to detect and understand sign language for this feature to work as intended. The following functions will be programmed into the system's assigned segment:

- Configuration of key points.
- Extraction of key point values.
- Training of key point values.
- Creation of preprocess data, labels, and features.
- Development and training of an LSTM neural network

Figure 1: MP Holistic Configuration-Hand

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 62-68, January-2024

https://doi.org/10.47001/IRJIET/2024.801008

Figure 2: MP Holistic Configuration – Body

To create a motion detection configuration, it is crucial to incorporate landmarks and important locations. MediaPipe Holistic, which enables real-time tracking of body postures, facial landmarks, and hand movements, is the ideal technology for achieving this. In this situation, it is thought redundant to record body postures because hand tracking and face landmark recognition are the main concerns. As a result, the first step entails configuring crucial locations using MediaPipe Holistic. Figures 1 and 2 show visual representations of images that have been configured using MediaPipe and serve as examples.

After MP Holistic's configuration process is complete, the next step is to extract and train the acquired data. The generation of pre-processed data or a dataset with hand-sign expressions becomes necessary after these processes, and the manual compilation of these datasets is crucial. Finally, the LSTM Neural Network setup and training comprise the last steps.

After completing the aforementioned tasks successfully, the system will be able to recognize motions using the pre-existing datasets stored in the database and detect motions by utilizing the key points that have been determined. The use of a cloud-based database is thought to be the most appropriate approach for this system in order to guarantee universal access to the data.

B) Convert Text into Sign language

Now, when a user types of text, the text will be recognized and detected. The next step is to put a technique in place for translating the extracted texts into symbols for America Sign Language.

1) ASL word dataset creation and ASL font creation

For words, sentences, alphabets, and numbers, ASL has signs. This stage involved creating the ASL sign data set that contained the ASL video for the pertinent term.

There are no signs for every word in American Sign Language. For example, let's take a person's name, "Hasiru". In that case, there is no sign for "Hasiru" in ASL. Therefore, we require a method of showing that name in ASL. To accommodate this, an ASL font for English alphabetic. characters were created. The ASL font was created using the images of hand signs in TrueType Font file (ttf) format. The created ASL font contains signs for 0 - 9, a-z / A-Z. Additionally, it contains punctuation marks to enhance the readability or the clarity of the translation of text images to ASL.

2) Final implementation of extracted text into ASL conversion

Using the previously trained and quantized float16 model, the extracted text from theinput image will be obtained in this stage. The text is then kept in either alphabetical or word format in an array list. The retrieved text's pattern recognition/matching parameters are then evaluated against the database's collection of ASL signals. After the correct values have been matched, the relevant result will be displayed. If a sign in the database matches the extracted text, the algorithm returns graphics that explain the motion. The ASL alphabet (ASL typeface) is also used to translate additional words that do not have a corresponding symbol in the database.

C) Convert Sign language into Voice

There is a get predict grammatically correct sentences using the words relevant to the given sign. The sign language word and the English sentences must first be combined. The phrases will be combined to form a single sentence.

After that a converter will be used to ensure that the statement is grammatically sound. It used Natural Language Processing (NLP) to convert the text to voice. The steps of the following tasks.

1) Input the text message.
2) Identify the incorrect word.
3) Using the grammar former and correct the sentences. (Correct concatenation sentences)
4) Using NLP and convert text to voice.

To address their problems, separate data sets have been created.

1) Only the English root words have sign in ASL.
2) In ASL, there are no signs for prepositions.
3) The word order in an ASL sentence is object, subject and verb whereas in English is a subject, verb and object.

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 62-68, January-2024

https://doi.org/10.47001/IRJIET/2024.801008

With the grammatically faulty statement as input and a correct sentence as output, each dataset has more than 1000 data a point.

For better accuracy, the dataset was trained incrementally by gradually adding more data and altering the training arguments. There was a definite reduction in loss and improvement in accuracy with the addition of more data and model fine tuning.

The modified sentence is shown in real time text format. Users can choose the narration option to have the machine narrate a grammatical sentence which will make communication feel more natural.

D) Detect Voice from Ordinary person and convert the message into text form

The main components that will be discussed in this component would be as follows:

a) Voice Detection
b) Lip Detection
c) Message Identification
d) Sentence Generation

In the modern world, there are many technical barriers that are designed in the modern world, however, most of them encounter different barriers during the implementation process and problems occur during the process of receiving and delivering. Generally, a voice conversation algorithm is utilized to assist to modify a source speaker's speech to the voice/sound that is produced by the speaker is presented. The application of speech technology especially in the fields of codeless/wireless communications, speech recognition, and digital hearing aids are a few examples of such systems that often involve a noise reduction technique operating in combination with a more specific and precise voice activity detector (VAD).

a) Voice Detection

When we consider voice detection, this system will be developed by using Machine Learning, and Deep Learning. In addition to training the system to recognize natural languages using data sets, the system is trained to recognize the characteristics of the voice. Automatic Speech Recognition (ASR) techniques will be used to identify the natural languages. Once we review much literature with reference to speech recognition systems, genuinely seek the first attention towards the discovery of Alexander Graham Bell, to identify the process of converting sound waves into electrical impulses thus it is noted that the first speech recognition system was developed by Davis et al. for recognizing telephone quality digits spoken at normal speech rate. ASR is basically positioned on building up an electronic circuit for identifying and recognizing ten digits of telephone quality.

There are mainly two phases that are involved in Automatic speech recognition systems: The training phase and the recognition phase. A difficult training procedure should be followed in order to map basic speech such as phone, syllables to acoustic observation. Likely we can ASR the initial stage of the training phase, the known speech is recorded and later pre-processed, which then enters the first stage of i.e., Feature extraction. Later will be followed by HMM creation HMM training and HMM storage. Following the recognition phase, it begins with the analysis of acoustic analysis of the unknown speech signal. The captured signals are thereafter converted to a series of acoustic feature vectors, by using a suitable algorithm, the input observations are then processed, and the speech is compared against the HMM's networks and the word that is pronounced will be displayed. However, during the implementation of ASR systems, it can only recognize what is learned during the training process. Out of the several modules that are identified in the current context, we will be using the feature extraction module which is a module used more commonly.

b) Lip Detection

Watching the speaker's lip movements to audition can help to improve speech understanding especially based on lip shape temporal evolution than on absolute mouth shape. Our report is to propose an algorithm that will extract lip movements over an image sequence. This algorithm does not require any makeup and works under natural lighting conditions. The lip detection algorithm describes the mouth based on the shape of the active model.

For continuous experiments, the methodology can be used in many aspects. For instance, Decision Making, (Fretias, Jimenez & Meando, 2004). VSR is used in many industries such as teaching techniques, health sectors etc.. However, there are a few areas that should be focused on developing the systems and those problematic situations are mainly to be addressed. Namely, Identifying the speakers' characteristics in the analysis phase and secondly how to incorporate the speaker-specific knowledge while synthesis during the transformation phase. Once we are to develop, we need to closely monitor these aspects. These characteristics need to be represented in a suitable manner while integrating them either in a voice-to-text mode or vice versa systems. Moreover, the experiment was done by Olive and Nakatani [J. P. Olive and L. H. Nakatani,''Rule−synthesis of speech by word concatenation, this method can be further studies and applied for our application too. For instance,' J. Acoust. Soc. Am. 55,

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 62-68, January-2024

https://doi.org/10.47001/IRJIET/2024.801008

660−666 (1974)] describes has shown that the numeric value of the telephone is rule−synthesized by merging numbers together, which were spoken in isolation. The suitable adjustment of the fundamental frequency is more important than adjustments of amplitude, and duration. The word coarticulation relationship is also important, for merging numbers' sounds as if they were spoken naturally.

c) Message Identification & Sentence Generator

In order to process the message identification and the sentence generator, data plays a key role. There are many researchers that have used similar approaches for message identification. However, we need to train the model correctly to accept a wide, diverse set of training techniques to prevent any language model from over fitting. While there are different ways, one approach is considering the symbolic expressions constructed by the rules in a context-free grammar while another approach is taken in how to use parse tree illustrations of symbolic formulas. It is suggested that the latter approach can be used for our module. System also considers the data that the user provided initially including age, gender, and BMI calculated by the system.

IV. RESULTS AND DISCUSSION

This section describes the test results performed on each model. The main objective of this research is to build and maintain a better communication bond between the public and people with hearing or speech impairment. High accuracy of the results must be met to ensure the accuracy of the translation while reducing the gaps that may arise in the communication process. A simple user survey was organized to identify the level of knowledge among the public about sign language and to gather their suggestions and opinions about sign language interpreters. The results of the survey show several key aspects.

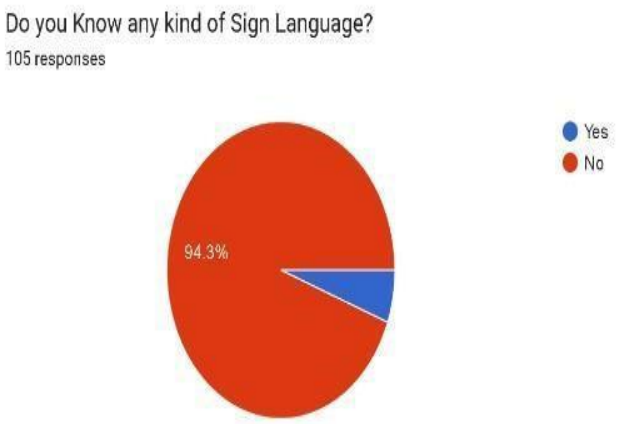

Figure 3

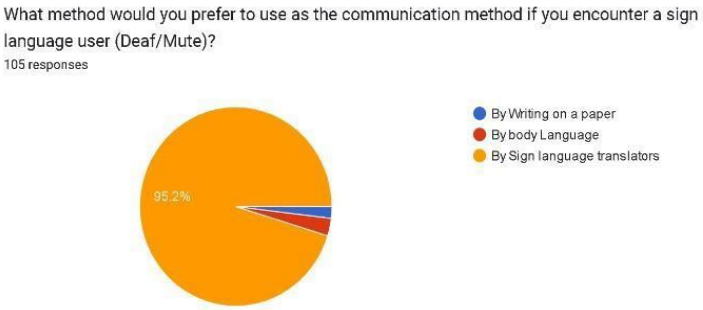

Figure 4

In the above paragraph, the focus is on a system with four primary features. The initial feature involves converting sign language (ASL) into text and subsequently translating it into voice. Another feature enables the conversion of voice messages from ordinary individuals into text format. The system also includes a feature that converts voice into sign language. These main features encompass sub-features such as hand gesture detection, image classification, and text and audio generation. To train hand gesture detection, a faster RCNN model was employed on TensorFlow. The model underwent testing in various scenarios, where input from a web camera was fed into the model. The model then generated a bounding box, determined the class, and provided accuracy. The accuracy percentages for recognizing hand signs, even against different backgrounds, are illustrated in Figure 5.

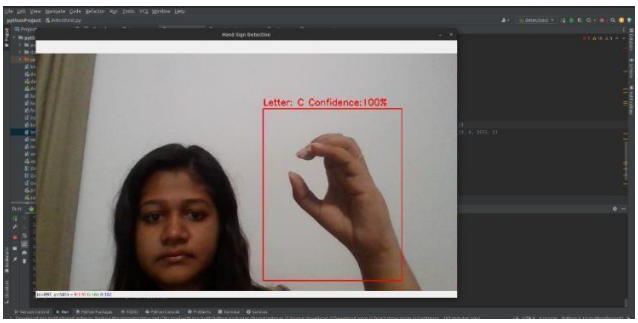

Figure 5

The table presented below displays the accuracy rates for the detection of hand images in the Data Acquisition component across various test scenarios.

Table 1

Test Scenario	Number of test runs	Mean Detection Accuracy (%)
The detection of hand gestures in complex backgrounds is being evaluated for accuracy.	20	93%
Hand gesture detection – Clear environment	20	100%
Hand gesture detection – dark environment	20	85%

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 62-68, January-2024

https://doi.org/10.47001/IRJIET/2024.801008

The Image classification component of this application involves training a CNN model with 26 classes, each consisting of 500 images. The dataset was divided into 85% for training and 15% for testing. To ensure better training, 50 epochs were utilized, and testing was performed using 38 samples from each class.

Table 2

Class	Accuracy (%)	Class	Accuracy (%)
A	100%	N	97%
B	100%	O	100%
C	100%	P	97%
D	97%	Q	95%
E	100%	R	100%
F	97%	S	97%
G	95%	T	100%
H	100%	U	97%
I	100%	V	100%
J	95%	W	97%
K	97%	X	95%
L	100%	Y	97%
M	97%	Z	97%

During the testing phase, the model demonstrated an average accuracy of 97%. While this level of accuracy is commendable, challenges arise when dealing with almost similar signs in real-time scenarios. This is one of the major drawbacks regarding this sub-feature.

V. CONCLUSION

In the world, there are fewer sign language translation tools than people who would be ready to utilize them for daily communication. This research study explains how online applications can be utilized for communication while translating American Sign Language to text/audio formats.

Faster RCNN-based model is first trained to identify the hand signs, and the Machine Learning-based API is then utilized to convert those detected hand images into verbal English. Future sign language-related advancements could make use of this API. It's not necessary for developers to create a categorization model from scratch. They only need a proper dataset and this API to use. The Text and Voice Generator helps to recognize word segments from collections of alphabets, then corrects any spelling errors before speaking the word segments out. The application translated sign language as well as NLP-related languages using an NLP based API. Additionally, the program has a reverse engine that allows regular people to convert spoken languages into American Sign Language. Using semantic analysis, this reverse translator transforms the text that the user sends into GIF images of the appropriate sign languages.

The system is now being considered as a web application but will soon also be available as a mobile application with quicker answers and less processing time.

ACKNOWLEDGEMENT

Our sincere gratitude goes to the Sri Lanka Institute of Information Technology, located in Malabe, Sri Lanka, for their invaluable support and provision of essential facilities and resources to carry out this research.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I.S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[8] Davis, K., Biddulph, R., and Balashek, S., “Automatic Recognition of Spoken Digit,” J. Acoust. Soc. Am. 24: Nov 1952, p. 637.

[9] Hemdal, J.F. and Hughes, G.W., A feature-based computer recognition program for the modeling of vowel perception, in Models for the Perception of Speech and Visual Form., MIT Press, Cambridge, MA. The Journal of the Acoustical Society of America 57 (2), 476-482, 1975.

[10] Kain and M. W. Macon, “Spectral voice conversion for text-to-speech synthesis,” in ICASSP, vol. 1, 1998, pp. 285–288.

© 2024-2017 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 62-68, January-2024
https://doi.org/10.47001/IRJIET/2024.801008

[11] Watcher, M. D., Matton, M., Demuynck, K., Wambacq, P., Cools, R., "Template Based Continuous Speech Recognition", IEEE Transaction.

[12] J. S. Sonkusare, N. B. Chopade, R. Sor, and S. L. Tade, "A review on hand gesture recognition system," 2015, doi: 10.1109/ICCUBEA.2015.158.

[13] U. Patel and A. G. Ambekar, "Moment Based Sign Language Recognition for Indian Languages," 2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2017, pp. 1–6, 2018, doi: 10.1109/ICCUBEA.2017.8463901.

[14] takelessons, 2019. ASL sentence structure: Word order in American sign language LASL lessons. YouTube. Available at: https://www.youtube.com/watch?v=nzrbvMeoBnE [Accessed. February 1, 2022].

[15] Z. A. Memon, M. U. Ahmed, S. T. Hussain, Z. A. Baig and U. Aziz, "Real Time Translator for Sign Languages," 2017 International Conference on Frontiers of Information Technology (FIT), 2017, pp. 144-148, doi: 10.1109/FIT.2017.00033.

[16] T. Vichyaloetsiri and P. Wuttidittachotti, "Web Service framework to translate text into sign language," 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), 2017, pp. 180-184, doi: 10.1109/CITS.2017.8035336.

[17] International Advanced Research Journal in Science, Engineering and Technology, Vol. 8, Issue 4, April 2021, DOI: 10.17148/IARJSET.2021.8412.

Citation of this Article:

Suranjini Silva, Thilini Jayalath, Madhushani E.A.Y.C., Amarasooriya H.D., Elpitiya S.N., Frank Perera, "Machine Learning Based Customized Solution for Deafness and Mute People" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 1, pp 62-68, January 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.801008

****

© 2024-2017 IRJIET All Rights Reserved