

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/364731666>

# Sign language translator: Web application based deep learning

Conference Paper in AIP Conference Proceedings · October 2022

DOI: 10.1063/5.0093367

CITATIONS

0

READS

126

3 authors, including:



[Abdullah Baktash](#)

University of Telafer

2 PUBLICATIONS 5 CITATIONS

[SEE PROFILE](#)



[Ammar Yahya](#)

Middle Technical University

15 PUBLICATIONS 49 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Computational Intelligence [View project](#)

# Sign language translator: Web application based deep learning

Cite as: AIP Conference Proceedings **2398**, 050020 (2022); <https://doi.org/10.1063/5.0093367>  
Published Online: 25 October 2022

Abdullah Qassim Baktash, Saleem Latteef Mohammed and Ammar Yahya Daeef



View Online

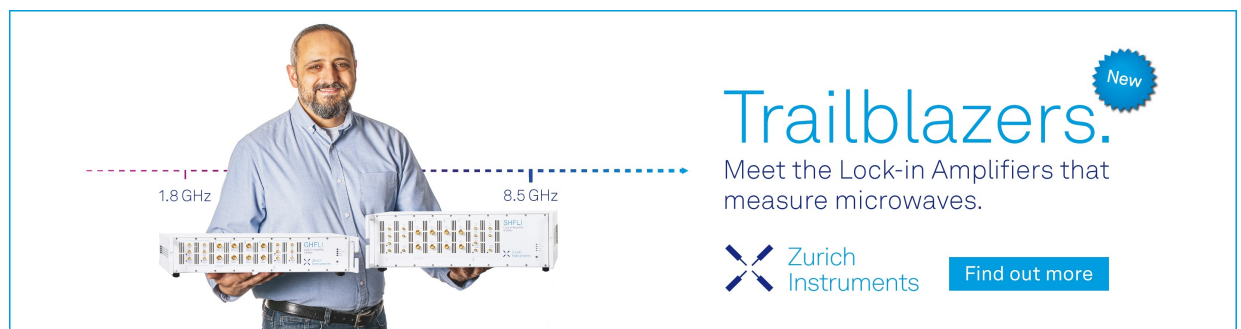


Export Citation

## ARTICLES YOU MAY BE INTERESTED IN

[Environmental pollution "causes - types - effects"](#)

AIP Conference Proceedings **2398**, 040023 (2022); <https://doi.org/10.1063/5.0093364>



**Trailblazers.** New

Meet the Lock-in Amplifiers that measure microwaves.

Zurich Instruments [Find out more](#)

# Sign Language Translator: Web Application based Deep Learning

Abdullah Qassim Baktash <sup>a)</sup>, Saleem Latteef Mohammed <sup>b)</sup>, and Ammar Yahya Daeef <sup>c)</sup>

*Department of Medical Instrumentation Techniques Engineering, Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq.*

<sup>a)</sup> Corresponding author: abdallah\_baktash85@uotelafer.edu.iq

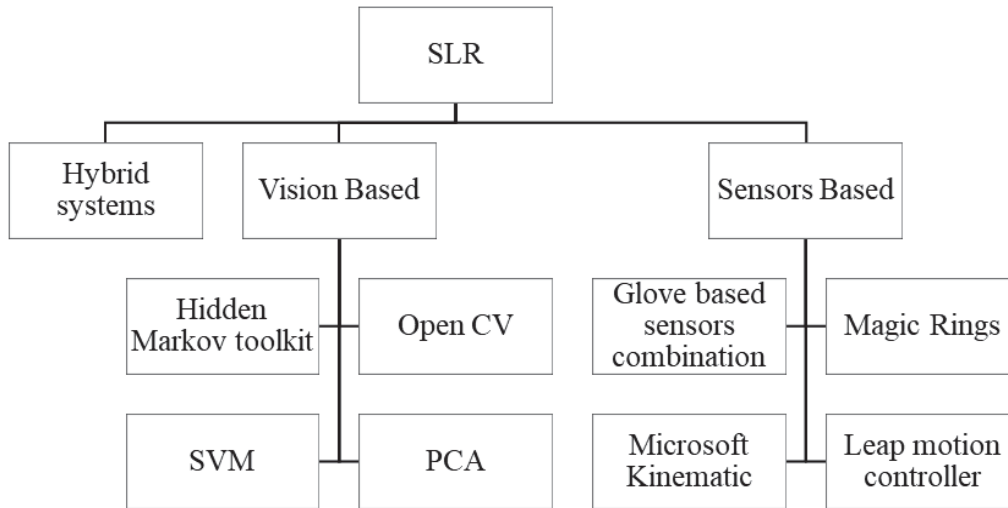
<sup>b)</sup> saleem\_lateef\_mohammed@mtu.edu.iq

<sup>c)</sup> ammaryahyadaef@mtu.edu.iq

**Abstract.** As a result of different physiological and accidentals causes for the inability of a man to speak. It becomes necessary to develop an efficient user-friendly technique to translate visual sign language into speech. In this paper, a visual-based translator is proposed as a hand gesture classification model. Region of interest (ROI) and hand segmentation is performed using a mask Region-based Convolutional Neural Network (R-CNN). The classification model is trained using a large number of gestures dataset using Convolutional Neural Network (CNN) deep learning and hosted in a web server. The system has realized a high accuracy of 99.79% and a loss of 0.0096. The trained model is loaded from the server to an internet browser using a special JavaScript library. The hand gesture is captured using a smart device camera and applied to the model to provide a real-time prediction.

## INTRODUCTION

Sign language translation continues to be an important way of communicating with the speech-impaired into the rest of normal people. Communication is the major challenge faced by deaf and mute people. This problem becomes serious when it is associated with the inability of reading and writing. Hence, the development of a sign language translation into text technique becomes an important prerequisite. Numerous kinds of visual communication available such as sign language, leap reading, air writing, fingerspelling, and different gestures with facial expressions. However, visual sign language is the most widely used way to convey the messages of physically impaired people [1]. The performed gestures can be detected either by using electronic sensors, visual methods, or by a combination of both. Fig.1 depicts the taxonomy of gesture translator systems. A sensor-based system employs different types of cost-effective and small-sized electronic sensors and microcontrollers to acquire enough information about the hand gesture. Whereas, the vision-based techniques include capturing the hand gesture's image and analyzing using image processing, and artificial intelligent algorithms. The hybrid recognition approach comprises a camera and one or more electronic sensors. Aside from information provided from the camera, additional data is provided by the sensor about the hand posture [2].



**FIGURE 1.** SLR Taxonomy

A short overview of previously used approaches for understanding sign language is mentioned below:

## VISION-BASED SYSTEM

Sadik et al. [3] introduced a binary masking skin segmentation and Support vector machine (SVM) classifier for Bangla sign language recognition. A binary image mask is generated by converting image color into YCbCr color and Fuzzy C- means clustering which provided a skin segmentation. The proposed system was trained for 10 classes of alphabet containing 40 images of hand gestures using a multi-class SVM classifier. The accuracy obtained of 99.8%. However, the system is developed to classify only static hand poses and the dataset has a few images to train the model. Rathi and Gawande [4] developed an intelligent system for deaf and dumb communication assistants. The proposed system designed 6 different words of Indian sign language from video input. The video frames are converted into HSV color space for skin color filtering. Furthermore, Hue saturation was used for skin detection. A binary image mask is applied after transforming the image into a gray-scale level applied. The region was extracted using the biggest binary linked object in the image. The classification was done using Euclidean distance values in the Eigenvectors recognition techniques. However, the system is designed to translate only six words. Bantupalli and Xie [5] introduces a deep learning classification method for gesture translation. the hand video recorded by a camera and video frames preprocessed and generated a 2400 image sample. The researchers utilized the Inception model developed by Google for feature extraction. The system provided two output classification methods: a SoftMax prediction layer or a pooling layer. The features were applied to Long-Short Term Memory which obtained higher accuracies. However, the proposed system trained only for 1800 image samples in the same image background. Sripairojthikoon and Harnsomburana [6] relied upon 3D Convolutional Neural Networks (CNN) for Thai Sign Language (TSL) translator which compared two hands posture, shape, and motion sequences to represent the alphabet of the language. Obtained accuracy was maximum when 15 depth frames were used and input data belong to skeleton video which was 97.7%. however, the Microsoft Kinetic device was not portable and it costs the user. Kalam et al. [7] developed a digit recognition system in sign language based on a feed-forward convolutional neural network (CNN) with the help of residual learning. The proposed system was built on a supervised learning technique called mini-batch stochastic gradient descent. The researchers were aimed to develop 10 layers deep CNN that able to classify digits (0 – 9). The system has shown an accuracy of 97.28% on the proposed method and 86.71% on the deep CNN. The system was limited to digit number detection. Furthermore, a high-specification computer requires to perform the classification. Karmel et al. [8] proposed the Internet of Things (IoT) assistive and an application programming interface Google (API) cloud system for the blind, deaf, and dumb persons. The major

part was Raspberry Pi supported with Google API, the device also included a microphone, camera module C310HD, speaker, and LCD.

## **SENSOR-BASED SYSTEM**

Al-Nuaimy [9] introduced an inexpensive electronic glove to help speechless people to communicate with normal people. The authors proposed hand press was sensed by a force flex sensor in the glove translated into visual and audio form. The system comprised three flex Force sensors, Arduino Uno microcontroller, display, and speaker. However, was tested only for six words. Lee et al. [10] developed a smart hand wearable sign gesture interpreter to identify American Sign Language (ASL) characters was familiarized. The system comprised three units: hand wearable sensors unit, processing unit, and mobile application unit. Experiment results have pointed the accuracy of the system was 65.7% without using pressure sensors and enhanced to 98.2% when pressure sensor used on the middle finger. Vijayakumar et al. [11] introduced a gesture to speech (G2S) system to portray a large number of messages with lesser sensors used. The prototype consisted of four flex sensors placed on the glove worn by hand, accelerometer sensor, analog to digital converter (ADC), raspberry pi 3 controllers, display, and speaker. The experimental results have shown the accuracy changed from 79% to 93% with an average accuracy of 86% for 14 gestures. The system was a prototype, not a confirmed system. Kumar et al. [12] projected a data glove prototype that transforms the figure's motion into audible sound and visual form. The authors relied upon three flex sensors that generate eight outputs. The proposed system was used 5 sensors to generate  $(2^5) = 32$  corresponding gestures. Also, the system accuracy could be increased by using an accelerometer. Kala et al. [13] designed and implemented a hand gesture translator embedded device to reduce the gap between a mute individual with normal people. The developed model comprised three parts: sensing part, processing the input and output communication. The device was comprised flex sensor attached for each finger of the glove, a three-axis accelerometer, Arduino ATMEGA 2560, Bluetooth module HC-05, LCD, Android smartphone. However, the operated from main supply and device is not portable. Developed an Indian Sign language into a voice translator. The gesture in this system is captured using five flex sensors and an inertia measurement unit (IMU).

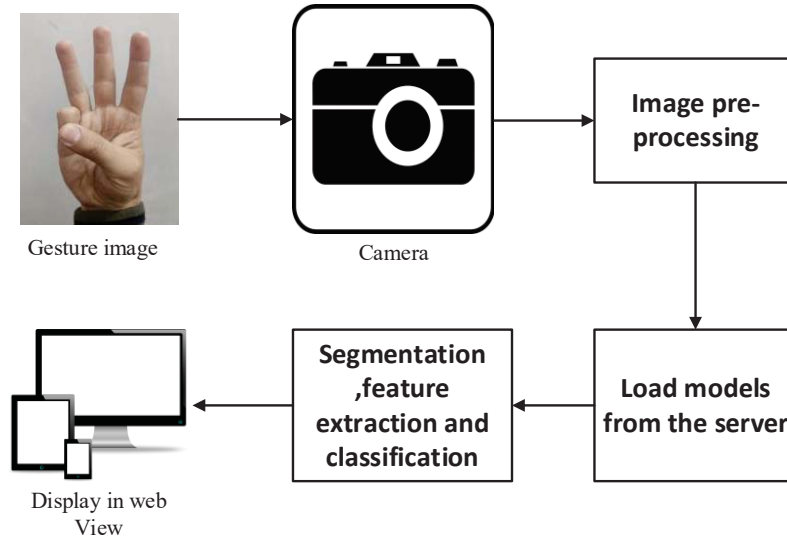
## **HYBRID SYSTEM**

El Badawy et al [14] proposed a sensor-based and camera-based combination for Arabic sign language translation. The authors have used an LMC device for hand gesture detection in addition to using two digital cameras to capture body movement and facial expression. The recognition accuracy achieved from the hybrid system was 95% obtained from 20 dynamic sign datasets for facial expression and body movement. However, the camera must hold closer to the user's face to cancel other body part's movement and hand motion. Besides, a good background lighting condition must be provided. Furthermore, the proposed system is costly and not portable.

In this paper, a web application is established to translate Arabic Sign Language gestures using a CNN learning algorithm. The developed system is stored in the internet server and can be accessed from everywhere and from a web browser of a smart device.

## **PROPOSED METHODOLOGY**

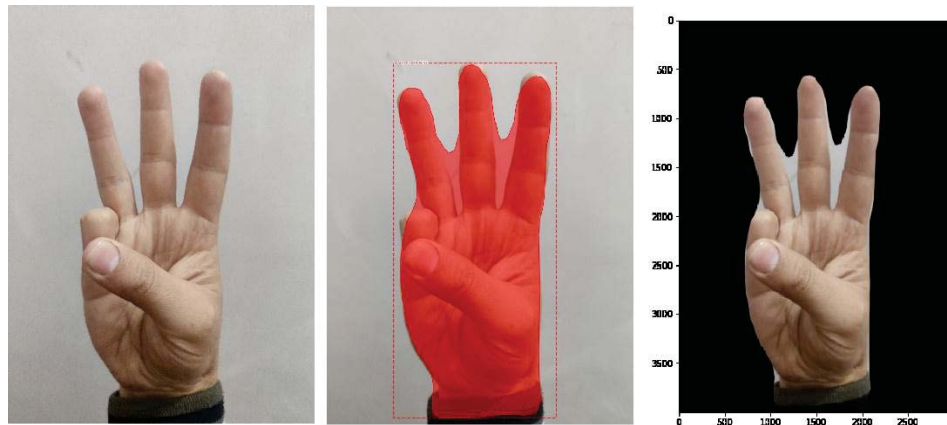
In this approach, the hand gesture is recognized and classified using CNN deep learning computer vision techniques. The block diagram in Fig.2 shows the process flow of the system. The hand gesture is recorded by a smart device digital camera. The image frames are preprocessed by normalizing, resizing, and a grayscale color mode conversion of the frames to match the input layer of the neural network. The region of interest is detected and segmented using mask region-based Convolutional Neural Network (R-CNN). The images are applied to the classification model which is loaded from a hosted web server using a special JavaScript library developed to run and train Artificial Intelligence (AI) models from a web browser.



**FIGURE 2.** Block diagram of the proposed system.

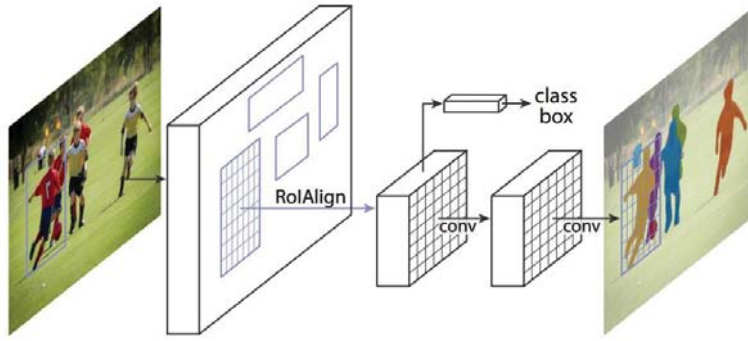
## REGION OF INTEREST AND SEGMENTATION

A mask R-CNN architecture is developed by Facebook AI Research (FAIR) to find out the region of interest and segmentation purpose. A mask R-CNN is extended from Faster RCNN which is utilized for object detection and labeling by coordinating a bound box for each object in the image. Besides, from object detection with the bound box, a mask R-CNN provides a binary mask in the output stage for a detected object as shown in Fig.3.



**FIGURE 3.** Hand segmentation process

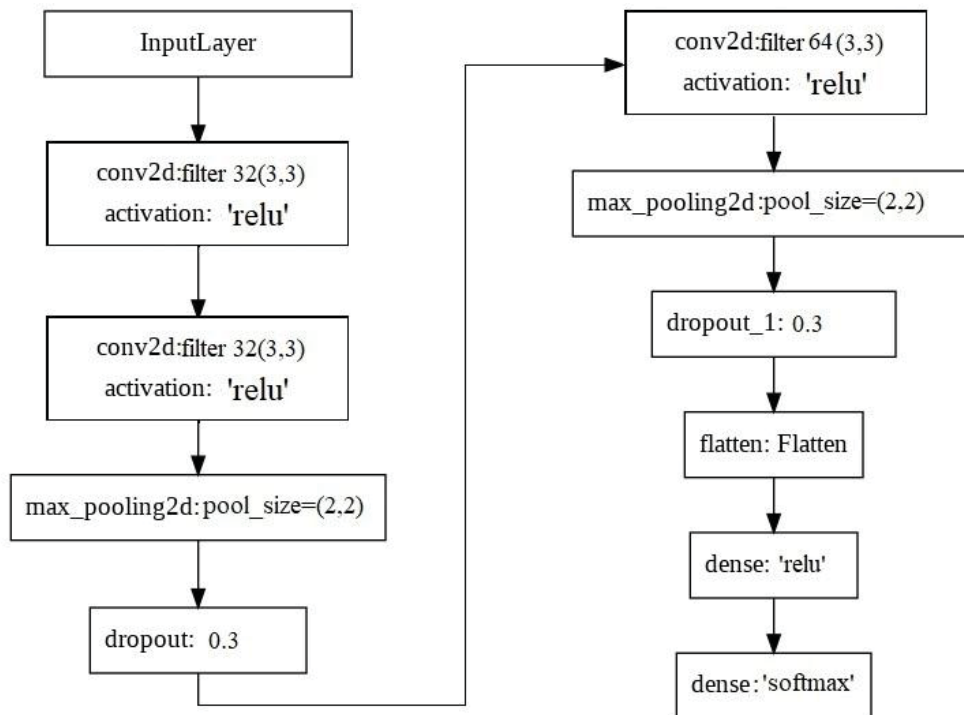
Figure 4 shows the architecture of the mask R-CNN. A convolutional net is used for feature map extraction from the image and the extracted data is passed to Region Proposal Network (RPN) stage which suggests a bounding box for each object in the image according to the feature map extracted. In the next stage, the feature is extracted from the candidate box region using the RoI Pooling layer, and all regions are processed to be in the same shape and perform classification through a fully connected output layer. Finally, a binary mask is generated for each region [15].



**FIGURE 4.** Mask R-CNN segmentation architecture. Image source [15]

## IMAGE CLASSIFICATION

The most important task in any computer vision project is a classification process. One of the most popular techniques in deep learning classification is Convolutional Neural Network (CNN) [16]. It is mainly used to solve complex image-based pattern recognition tasks, with a simple framework. Fig. 5 shows the structure of the model trained in this technique. It comprises three convolutional layers with (32, 32, 64) filters with 3x3 kernel sizes respectively to perform two-dimensional convolution of the image data. The main function of the convolutional layers is to extract the feature map of image objects. Two layers of max pooling 2d which reduces the unwanted portion of the image by generating sub-region from the image and keeping only a maximum valued region, two dropout layers to reduce the overfitting issue [17], flatten layer containing two dense layers with Relu and SoftMax Activation function for 29 output classes [18]. The model is trained through Google Colaboratory (Colab) platform [19] which provides a powerful online machine learning environment to write and execute python code with free GPU access to accelerate training time.



**FIGURE 5.** CNN model structure

The dataset for American Sign Language (ASL) [20] is used to train the model containing a 3000 sample of 200x200 pixels image for each hand gesture with a total of 87000 images for English alphabet in addition to ‘space’, ‘delete’, and ‘nothing’ commands. The dataset contains different background illuminations and skin colors.

## WEB APPLICATION DEPLOYMENT

The establishment of a sign-language translator web application requires a trained model to be hosted on an internet server. A hardware-accelerated opensource JavaScript library called “TensorFlowjs” [21] developed by Google to train and load the machine learning models by a front-end web page designed to interface with the user and display the translation results on the smart devices web browsers.

## RESULTS AND DISCUSSIONS

Table 1 shows typical precision for each class in the trained model which stands for the number of true positive divided by the total of false-positive and true positive as in equation 1 [22].

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Where: TP = True positive, FP = False positive.

The model is trained for 10 epochs and the losses obtained are very small. The loss is 0.0288 for the training set and 0.0096 for the validation set. Whereas, the accuracy of 99.21 and 99.79 for training and validation process respectively as shown in Fig.7 a, b

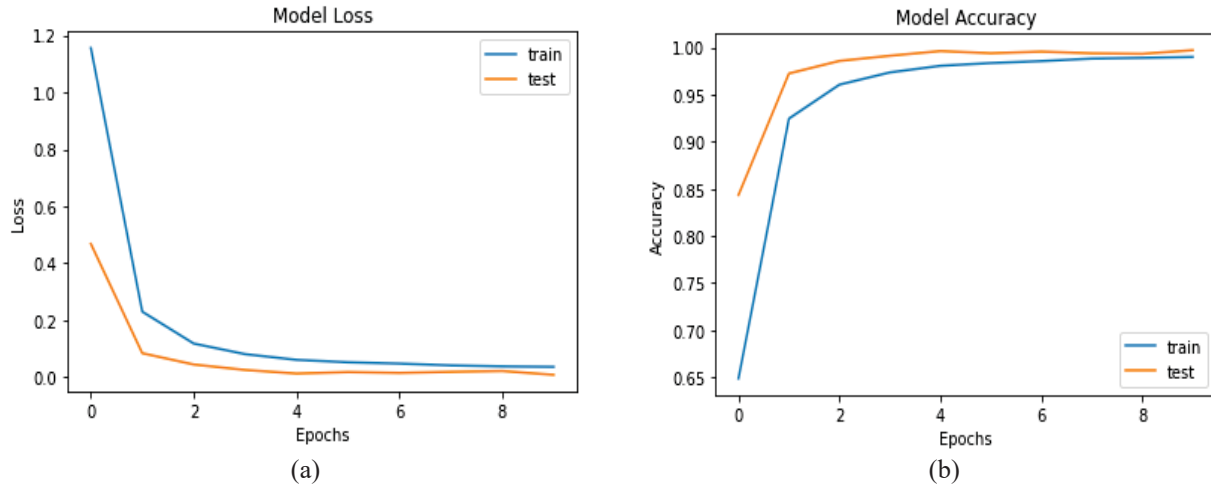


FIGURE 7. Model losses and accuracy

TABLE 1. Trained letters precisions.

Letter	Precision	Letter	Precision
A	0.9965	P	1.0000
B	0.9925	Q	1.0000
C	1.0000	R	1.0000
D	1.0000	S	1.0000
E	1.0000	T	0.9967
F	0.9966	U	0.9978
G	0.9989	V	0.9977
H	0.9967	W	0.9920



I	1.0000	X	0.9875
J	1.0000	Y	1.0000
K	1.0000	Z	0.9978
L	1.0000	del	1.0000
M	1.0000	nothing	1.0000
N	0.9893	space	0.9976
O	1.0000		

The real-time translation process occurred on the client-side on a personal computer after loading the model from the internet server. Fig.6 shows examples of the translation process on the personal computer internet browser and the reliable translation is obtained.

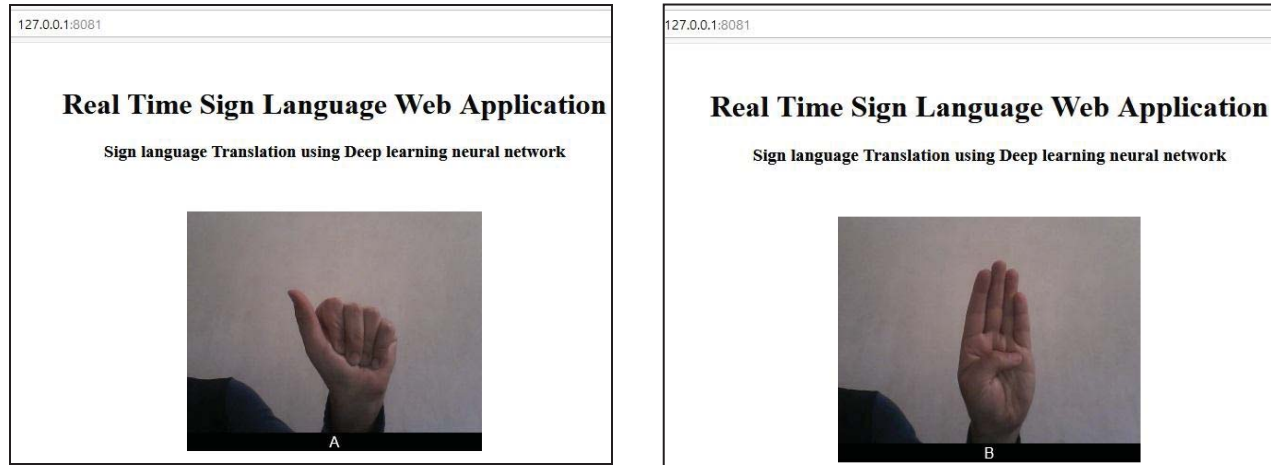


FIGURE 6. The real-time gesture translation process

## CONCLUSION AND FUTURE WORKS

Due to the great advancement in the field of Internet communication and the availability of the TensorFlow JavaScript library enables us to load and run a machine learning artificial neurons model on the web browser. So, the development of cross-platform sign language translator web application becomes possible to assist speechless people to communicate with normal people. The model is trained for 87000 image samples of the ASL dataset using the google Colab platform and obtained a validation accuracy of 99.79%. Future work will be possible in translating a complete sign language with two hands not only for hand gestures for letters only.

## REFERENCES

1. S.S. Wazalwar and U. Shrawankar, Proc. 2019 6th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2019 418 (2019).
2. K. Kudrinko, E. Flavin, X. Zhu, and Q. Li, IEEE Rev. Biomed. Eng. **3333**, 1 (2020).
3. F. Sadik, M.R. Subah, A.G. Dastider, S.A. Moon, S.S. Ahbab, and S.A. Fattah, 2019 5th IEEE Int. WIE Conf. Electr. Comput. Eng. WIECON-ECE 2019 - Proc. 1 (2019).
4. S. Rathi and U. Gawande, Proc. 7th Int. Conf. Conflu. 2017 Cloud Comput. Data Sci. Eng. 733 (2017).
5. K. Bantupalli and Y. Xie, Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018 4896 (2019).
6. N. Sripairojthikoon and J. Harnsomburana, ACM Int. Conf. Proceeding Ser. 186 (2019).
7. M.A. Kalam, M.N.I. Mondal, and B. Ahmed, 2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019 1 (2019).
8. A. Karmel, A. Sharma, M. Pandya, and D. Garg, *Procedia Comput. Sci.* **165**, 259 (2019).
9. F.N.H. Al-Nuaimy, Proc. 2017 Int. Conf. Eng. Technol. ICET 2017 **2018-Janua**, 1 (2018).
10. B.G. Lee and S.M. Lee, *IEEE Sens. J.* **18**, 1224 (2018).

11. K.P. Vijayakumar, A. Nair, and N. Tomar, [Int. J. Innov. Technol. Explor. Eng.](#) **9**, 1241 (2020).
12. V. Kumar, S.K. Raghuwanshi, and A. Kumar, Springer Proc. Math. Stat. **308**, 347 (2020).
13. H.S. Kala, S. Sushith Rai, S. Pal, K. Uzma Sulthana, and S. Chakma, Proc. - 2018 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control. ICDI3C 2018 97 (2018).
14. M. El Badawy, A. Samir Elons, H. Sheded, and M.F. Tolba, Adv. Intell. Syst. Comput. **323**, 721 (2015).
15. K. He, G. Gkioxari, P. Dollár, and R. Girshick, [IEEE Trans. Pattern Anal. Mach. Intell.](#) **42**, 386 (2020).
16. K. O'Shea and R. Nash, 1 (2015).
17. G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R.R. Salakhutdinov, 1 (2012).
18. A.F.M. Agarap, ArXiv 2 (2018).
19. T. Carneiro, R.V.M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V.H.C. De Albuquerque, and P.P.R. Filho, [IEEE Access](#) **6**, 61677 (2018).
20. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I.J. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Józefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, ArXiv abs/1603.04467, (2016).
21. N. Kumar, Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2017 2017-Janua, 244 (2017).