STM JOURNALS

Journal of
Artificial Intelligence Research & Advances

ISSN: 2395-6720
Volume 8, Issue 2, 2021
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Research JoAIRA

Maadhyam: Sign Language Translator (ASL)

Naman Gupta[1,*], Radhika Sharma[2], Priyanshi Jindal[3], Priyam Chauhan[4], Sapna Juneja[5]

Abstract

There are several million deaf and dumb people around the world and their only means of communication is sign language. They are usually not able to receive education and even some of them are unemployed due to lack of communication. This often leads to mental health issues among them. Deaf and dumb people rely on sign language interpreters for communications because other people do not prefer learning their language. However, our research aims to fulfil this gap, so that other people can understand sign and gestures without any human interpreter. It is a web application based on neural networks on which user can provide any random gesture and the project will automatically identify what gesture it is. We have currently trained our model for ASL (American Sign Language) letters only i.e. from A to Z. We have used CNN (Convolutional Neural Network) for training and recognizing gestures.

Keywords: Neural Network, Convolutional Neural Network (CNN), sign language, gestures, OpenCV, image data augmentation

INTRODUCTION

The communication through sign language (SL) is made by particulars of hand and facial colloquialisms to communicate their perspectives and contemplations of discourse and hearing handicapped people with the typical (discourse and hearing) individuals. The vast majority of the ordinary people may not unmistakably comprehend the gesture-based communication. In this manner, there is a huge correspondence hole between the hard of hearing networks with the overall population. There is a certainty of innovation support for discourse disability individuals as human interpreters are profoundly difficult to support discourse hindered people in their everyday exercises constantly. With the progression in science and innovation, we can consider planning a methodology that can decipher motion signs into humanoid or machine-understandable content. This reduces the communication gap between typical and hindered individuals.

Ordinary individuals can convey their contemplations and thoughts to others through discourse. One significant method for a specialized technique for the consultation disabled local area is the utilization of gesture-based communication, as 500,000 and 2,000,000 individuals utilize Sign Language as their significant day by day specialized instrument. These numbers may go amiss from other various sources; however, it is shockingly famous as referenced in Trudy Suggs book that American Sign Language is the third most-utilized language in the United States. It appears to be that 3.68% of the complete populace is discovered to be in need of a hearing aid and 0.3% of the absolute populace is practically deaf, out of an all-out populace of around 268,000,000 (2005) in the US. In Canada and the USA, American Sign Language (ASL) is for the most part liked as the vehicle of correspondence for the people in need of a hearing aid and the hard of hearing the same.

***Author for Correspondence**
Naman Gupta
E-mail: naman.g2208@gmail.com

[1-4]Student, Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, India
[5]Professor & Head, Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, India

Received Date: August 03, 2021
Accepted Date: September 20, 2021
Published Date: October 07, 2021

Citation: Naman Gupta, Radhika Sharma, Priyanshi Jindal, Priyam Chauhan, Sapna Juneja. Maadhyam: Sign Language Translator (ASL). Journal of Artificial Intelligence Research & Advances. 2021; 8(2): 42–50p.

A few techniques have been proposed in the past to decipher the signs utilizing the signals and highlights of the underwriter. Essentially, Ko and Yang fostered a finger mouse that empowers a client to indicate orders with the fingers. Other epic methodologies incorporate the shaded glove-based strategy, skin shading division, video arrangement appearance displaying and Hidden Markov Model (HMM) frameworks. The other class manages the utilization of shaded gloves to perceive the signs.

Neural networks, distance-based, HMMs, skin tone based, and other measurable strategies have effectively tackled the gesture based communication acknowledgment when considered for word/sentence acknowledgment. It just requires video-based information assortment and consequently prompts a superior regular interface for the client. Gesture based communication when contrasted with the communicated in language has distinctive sentence structure. In communicated in language, the discourse is gathering of sentences where words in the sentence are straight (for example single word followed by another), though in communication via gestures, a synchronous design exists with an equal worldly and spatial arrangement.

There are in excess of 120 particular sign dialects are utilized by the discourse hindered local area of different countries all through the universe like American Sign Language (ASL), Indian Sign Language, Australian Sign Language, Italian Sign Language, Sri Lankan Sign Language, and some more. Well beyond 70 million individuals in the universe and around 10 million individuals in India are utilizing gesture-based communication as their superb vehicle of correspondence.

ASL is the most broadly utilized SL on the planet and the fourth most usable etymological in North America. In the United States, however ASL is additionally utilized in Canada, Mexico, West Africa, and Asia. In excess of 20 different countries including Jamaica, Panama, Thai, Malaysia etc., in which English is the significant correspondence language use ASL for their hearing-debilitated local area correspondence. Almost 2,000,000 deaf individuals of the USA and Canada are utilizing ASL as their essential premise of correspondence [1]. ASL is an expansive just as confounded language that uses signs made by activities of finger and hand total through stances of the body and appearances of the face. As ASL is viewed as an exact and certified language, it has ample varieties, as different dialects do, like French and Spanish. ASL is an exceptional type of communication and ideal for a colossal part of the discourse disability populace. Its establishment, existing conditions, prospect trusts, and worldwide effect are very astonishing and educational.

ASL gives a bunch of 26 motion signs named as an American Manual Alphabet that can be pushed off to explain a considerable lot of the English words accessible. The 19 different hand states of ASL are pushed off to make 26 American Manual Alphabets. An indistinguishable hand shape with assorted directions is utilized for 'K' and 'P' letters signs. ASL additionally offers a bunch of 10 numeric motions to sign the numbers '0' to '9'. ASL does not involve worked in ASL counterparts finishes paperwork for exact things and specialized terms. Alongside ASL Alphabets and Numbers, there are a great many hand and facial motion signs which are accessible to sign the different English words too.

MATERIALS AND METHODS

We used Python and some other basic libraries like cv2 which focuses on video capture and image processing, NumPy, pickle, sqlite3 (for interacting with database), os, tensorflow, keras, sklearn, matplotlib, pyttsx3, pydot and graphviz. These were the basic requirements if one needs to perform it from scratch.

We started with creating a dataset and for that we need to initialize our hand's histogram. "We use the orientation histogram as a feature vector for classification and interpolation. For moving or dynamic gestures, the histogram of the spatiotemporal gradients of image intensity form the analogous feature vector and maybe useful for dynamic gesture recognition" [2].

For setting our hand's histogram, we need to collect basic features of our hand so that it can clone the exact shape of our hand. We have to take care of proper lightening while setting histogram, as the whole working of the project will depend on this only and getting a good hand histogram is really important. As we can see in Figure 1, at right we have actual hand 50 green squares which are helping in collecting features of our hand. We need to make sure that our hand covers all the squares, as well as we are getting proper boundary of our hand in the left side. This is how we detect the basic features of hand using OpenCV, and created our hand's histogram. We have to save it for further training our model [3].

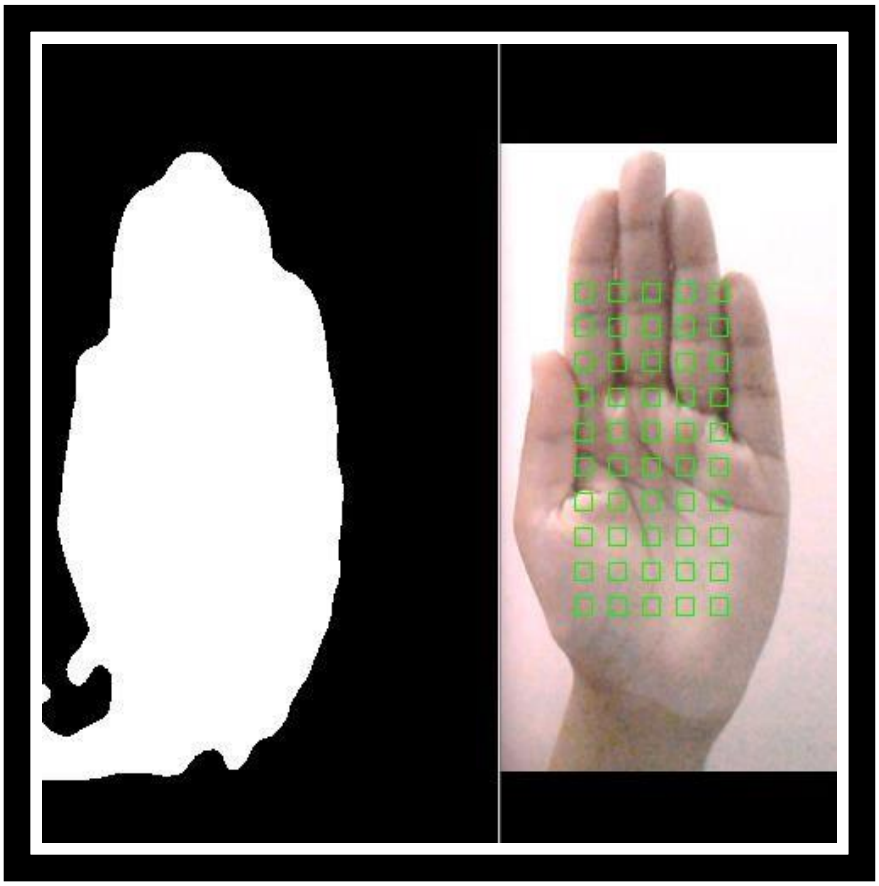

Figure 1. Setting hand histogram.

Our next step is to create database. In this process, we created total 27 gestures (26 English Alphabets and one "Best of Luck" gesture) using OpenCV. We can refer Figure 2 for basic gestures in American Sign Language. For each gesture we captured total 1200 images, each of 50×50 pixels. We used sqlite for interacting with database [4].

All of the images were stored in grayscale in database. We used OpenCV for the purpose of converting our images to grayscale. Figure 2 shows the way data is stored for training purpose.

Then we performed image data augmentation. Image data augmentation is used to expand the size of training dataset by creating modified versions of images in dataset. Training deep learning neural network can result in better accuracy and more skilled models. Augmentation techniques can create variations of the images that can improve the ability of the models to generalize what they have learnt to new images [5]. Hence, we flipped each and every image under data augmentation, and at the end we had total 2400 images of each gesture. We can see the original and flipped version of gestures in Figure 3.

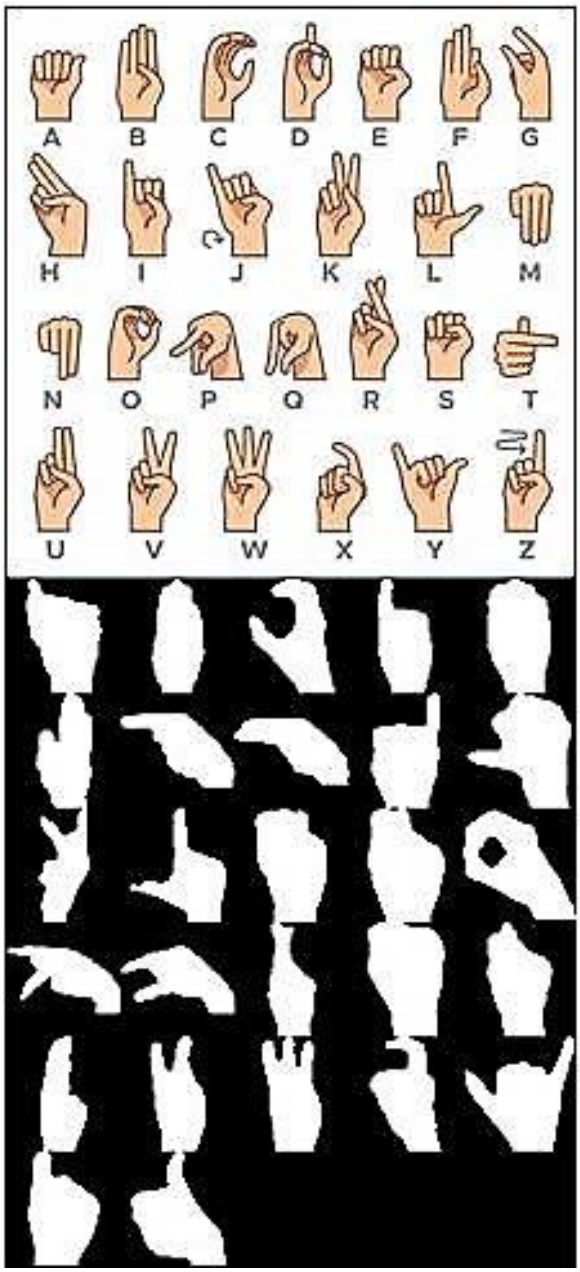

Figure 2. American sign language gestures and their corresponding gestures stored in database.

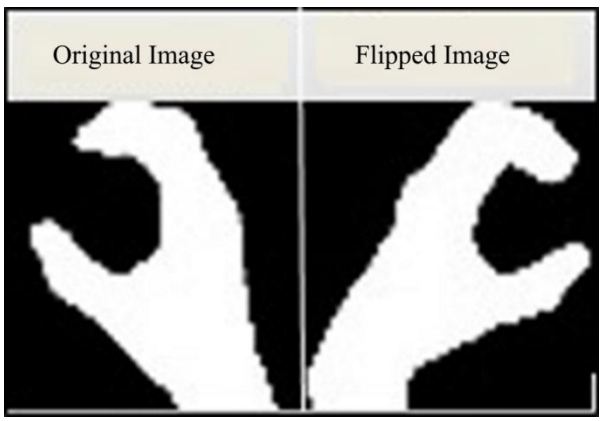

Figure 3. Original vs. flipped image.

Next step is to use our dataset to train our model using CNN (Convolutional Neural Network). We split our dataset into training and testing images and send training data to CNN layers as shown in the Figure 4. CNNs are used for classifying and recognizing images because of its high accuracy. There are following three types of layers in CNN:

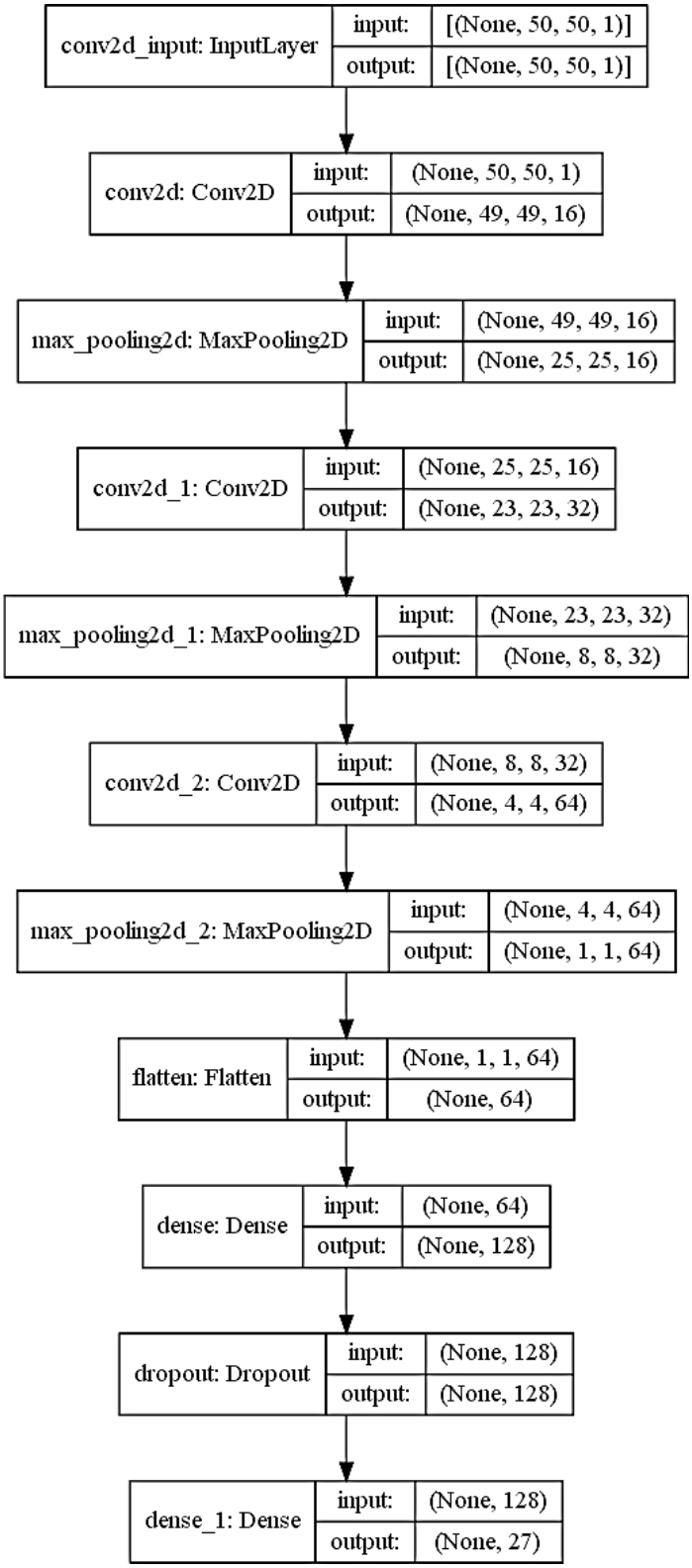

Figure 4. Flowchart of CNN layers.

Convolutional Layers

They are comprised of filters (basically neurons of the layer) and feature maps. It is the output of one filter applied to previous layer.

Pooling Layers

They down-sample the previous layer's feature map. It follows a sequence of one or more convolutional layers and are intended to consolidate the features learnt and expressed in the feature map of previous layer. It may also be considered as a technique to reduce overfitting.

Fully Connected Layers

They are used at the end of the network. They are used for making predictions by the network. These layers might have non-linear activation function or a softmax activation in order to output probabilities of class predictions.

We used all these layers in our model. Figure 4 shows the flow of layers in our Convolutional Neural Network model. We used 'Flatten' after a series of convolutional and pooling layer. Flatten is a function which converts the pooled feature map to a single column that is passed to the fully connected layers. As we are going towards the end, we are using 'dense layer'. Dense is a fully connected layer; it means that all the neurons in a layer are connected to all those neurons in the next layer. There are several reasons behind using dense layer. One of them is as follows: A convolutional layer relies on consistent features with a small repetitive field, whereas dense layer provides learning features from all the combinations of the features of the previous layer. In the second last phase we are using 'dropout'. Dropout is used to prevent a model from overfitting [6].

After training we pickled our model for future use. Pickling is the process in which a python object hierarchy is converted into a byte stream. Python pickle module is used for serializing and deserializing python object structure. Hence, we pickle so that it can be preserved on the disk [7–15].

RESULT AND DISCUSSION

The main purpose of our research was to link a particular section of society with the normal world. We made it possible by converting sign language gestures to text and then that text to sound.

We have used CNN (Convolutional Neural Network) to train and recognize gestures. Here we are using our already pickled model for testing and recognizing purpose. We were able to successfully identify all the English alphabets along with a 'Best of Luck' gesture with a very good accuracy. This model has been trained on 54000 samples and tested on 5400 samples. We received an accuracy of 99.93% with 0.07% CNN error, which can be depicted through Confusion Matrix as shown in Figure 5.

Confusion Matrix is a matrix that is used to evaluate the performance of a classification model. It is an N×N matrix where N is the number of target classes. From Figure 5 we can see that we were able to correctly classify most of the gestures. This model is capable of predicting output in very less time. The average prediction time is 0.000630 sec, and the time taken to predict 5400 test images came out to be 3 sec, which is comparatively lesser than the average time taken by other methods. Figure 6 shows a classification report which is used to measure quality of predictions from a classification algorithm. It helps to predict how many predictions are true and how many are false using four techniques that are precision, recall, f1 score and support. Precision is considered as the ability of a classifier not to label an instance positive that is actually negative. Recall, tells us the ability of classifier, to find all positive instances. F1-scores are lower than accuracy measures. They combine precision and recall together to evaluate the accuracy of the system. Support gives us the number of real occurrences of the class in the specified dataset.

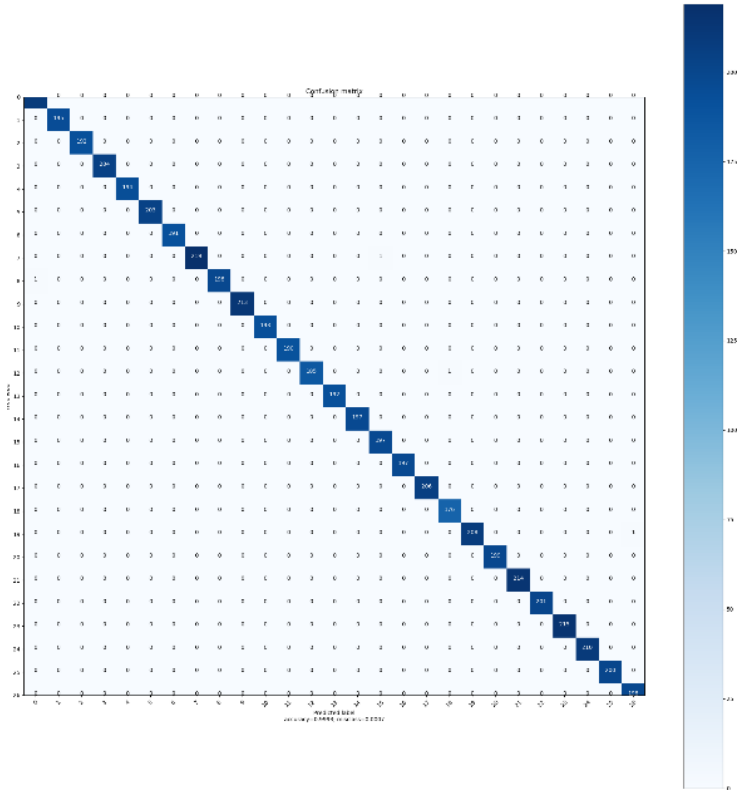

Figure 5. Confusion Matrix.

Classification Report

	precision	recall	f1-score	support
0	1.00	1.00	1.00	209
1	1.00	1.00	1.00	195
2	1.00	1.00	1.00	192
3	1.00	1.00	1.00	204
4	1.00	1.00	1.00	194
5	1.00	1.00	1.00	203
6	1.00	1.00	1.00	191
7	1.00	1.00	1.00	220
8	1.00	0.99	1.00	199
9	1.00	1.00	1.00	213
10	1.00	1.00	1.00	193
11	1.00	1.00	1.00	190
12	1.00	0.99	1.00	186
13	1.00	1.00	1.00	192
14	1.00	1.00	1.00	197
15	0.99	1.00	1.00	197
16	1.00	1.00	1.00	197
17	1.00	1.00	1.00	206
18	0.99	1.00	1.00	176
19	1.00	1.00	1.00	209
20	1.00	1.00	1.00	199
21	1.00	1.00	1.00	214
22	1.00	1.00	1.00	201
23	1.00	1.00	1.00	215
24	1.00	1.00	1.00	210
25	1.00	1.00	1.00	200
26	0.99	1.00	1.00	198

Figure 6. Classification Report.

Hence, from the classification report we can observe that we were able to achieve a very good precision, recall, f1-score, and support. It is very easy to use as well because all of this will happen through a web application and hence making it less complicated and user-friendly.

CONCLUSION

It is very difficult for the deaf and dumb people to communicate with normal people. We hope that this research will help the deaf and dumb people to connect and communicate with the normal people. This research illustrates an approach, to achieve the transliteration of all the 26 alphabets in English with a good luck hand gesture in addition. For hand gestures, we use the histogram of local orientations for recognition. The model is based on the simple hand gesture technique, the user makes an alphabet and the model will process using CNN model and then give the output in the form of text or speech. The accuracy of the model is very good i.e. 99.93%. The model is trained on 54000 samples of English alphabets of American Sign Language, including a best of luck gesture also. It is validated on 5400 samples. We can further work on this approach and train our model on more gestures like numbers or some common sentences. We can also add some additional features which can provide the result in real time. We can also add an Avatar based approach which will convert any text or speech into their corresponding gestures so that deaf and dumb people can understand and communicate in better way.

REFERENCES

1. Priyanka Mekala, Ying Gao, Jeffrey Fan, Davari A. Real-time Sign Language Recognition based on Neural Network Architecture. Conference: System Theory (SSST), IEEE 43rd Southeastern Symposium. 2011.
2. Freeman William T, Michael Roth. Orientation Histograms for Hand Gesture Recognition. Mitsubishi Electric Research Laboratories, TR94-03. 1994 Dec.
3. Dhanashree Bhandarkar, Amit Kale, Prajakta Mukhekar, Rohini Nere. Sign Language Converter and Recognition. International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE). 2020 May; 9(5): 170–172.
4. Pitambar Thapa, Pooja Daundkar, Mandar Jadhav, Rajesh Khape. Application for Analysing Gestureand convert gesture into text and voice. Int J Innov Res Comput Commun Eng. 2016; 4(10): 1899–1902.
5. Alina Kuznetsova, Laura Leal-Taixe, Bodo Rosenhahn. Real time sign language recognition using a consumer Depth camera. IEEE International Conference on Computer Vision Workshops. 2013.
6. Liang R, Ouhyoung M. Real-time continuous gesture recognition system for sign language. Proc 3rd IEEE International Conference on automatic Face and Gesture Recognition. 1998 May; 558–567.
7. Lubin J, Jones K, Kornhauser A. Using Backpropagation networks to assess several image representation schemes for object detection. International Joint Conference on Neural Networks (IJCNN). 1989.
8. Holden E-J, Lee G, Owens R. Australian Sign Language Recognition. Mach Vis Appl. 2005; 16(5): 312–320.
9. Ibraheem Noor A, Khan Rafiqul Z. Vision Based Gesture Recognition Using Neural Networks Approaches: A Review. International Journal of human Computer Interaction (IJHCI). 2012; 3(1): 1–14.
10. Huang, Deng-Yuan, Wu-Chih Hu, Sung-Hsiang Chang. Vision based hand gesture recognition using PCA+ Gabor filters and SVM. 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009 (IIH-MSP'09), IEEE. 2009.
11. Karishma Dixit, Anand Singh Jalal. Automatic Indian Sign Language Recognition System. 3rd IEEE International Advance Computing Conference (IACC). 2013.
12. Trigo TR, Pellegrino MRS. An Analysis of Features for Hand Gesture Classification. In 17th International Conference on Systems, Signals and Image Processing. 2010; 412–415.
13. Panwar M, Mehra PS. Hand Gesture Recognition for Human Computer Interaction. In International Conference on Image Information Processing. 2011; 1–7.

14. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a Convolutional Neural Network. 2017 International Conference on Engineering and Technology (ICET). 2017; 1–6. doi: 10.1109/ICEngTechnol.2017.8308186/
15. Connor Shorten, Khoshgoftaar Taghi M. A survey on Image Data Augmentation for deep learning. J Big Data. 2019; 6: Article Number: 60.

