Indian Sign Language to Speech Conversion Using Convolutional Neural Network

Shashidhar R
Department of Electronics and Communication Engineering
Sri Jayachamarajendra College of Engineering, JSS Science and Technology University
Mysuru, INDIA
shashidhar.r@sjce.ac.in

Surendra R Hegde
Department of Electronics and Communication Engineering
Sri Jayachamarajendra College of Engineering, JSS Science and Technology University
Mysuru, INDIA
surendrahegde3@gmail.com

Chinmaya K
Department of Electronics and Communication Engineering
Sri Jayachamarajendra College of Engineering, JSS Science and Technology University
Mysuru, INDIA
chinmayak77@gmail.com

Ankit Priyesh
Department of Electronics and Communication Engineering
Sri Jayachamarajendra College of Engineering, JSS Science and Technology University
Mysuru, INDIA
priyesh.ankit3@gmail.com

A S Manjunath
Department of Computer Applications
Sri Jayachamarajendra College of Engineering, JSS Science and Technology University
Mysuru, INDIA
as.manjunath@sjce.ac.in

Arunakumari B. N.
Department of Computer Science and Engineering
BMS Institute of Technology and Management
Bengaluru, INDIA
arunakumaribn@bmsit.in

2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon) | 978-1-6654-9790-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/MysuruCon55714.2022.9972574

***Abstract*— True incapacity is the inability to speak. A person who has a speech impediment is unable to communicate with others through speech and hearing. Individuals use sign language as a form of communication to overcome this disability. Even though signing has become commonplace in recent years, it can still be difficult for non-signers to communicate with signers. The flow of information and emotions in a person's life has become increasingly dependent on communication over time. The only way for that person with special needs to communicate with the rest of the world is through sign language, which uses entirely distinct hand motions. With the most recent developments in computer vision and deep learning techniques, there has been significant improvement in the disciplines of motion and gesture identification. For American Sign Language (ASL), sign language recognition has been a well-researched subject. Nevertheless, there aren't much published analysis works on Indian Sign Language (ISL). The intended method will recognise 4972 static hand signs for the twenty-four different English alphabets (A, B, C, D, E, F, G, H, I, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y). The main goal of our effort is to create a deep learning-based application that uses the "Google text to speech" API to translate sign language into text, facilitating communication between signers and non-signers. We made use of the Kaggle-available dataset. Proposed method using custom Convolutional Neural Network and got the accuracy of 99%.**

***Keywords*— Convolutional Neural Network; Google text to speech API; Indian signing.**

I. Introduction

Due to the tremendous obstacles in the fast-paced world, people who are dumb or deaf would find it difficult to interact with others for personal reasons or in any workplace since they cannot hear them. They can't hear automobiles, bikes, or other people coming back, making it risky for them to travel alone. They struggle to respond to many traditional people and swiftly adjust to the environment around them. It can be difficult for people to express their feelings. Language used by Indians and also known as the plains sign talk is a gestural form of social communication used by American Indians in the great plains. Spanish explorers recorded its use in the southern plains around the sixteenth century, however its origins are still unknown. The eleventh five-year plan (2007–2012) recognised that the needs of people with hearing problems had been largely disregarded and proposed the establishment of a symbol language research & training centre to advance linguistic communication and to train teachers and translators. Everywhere in India, the deaf community uses Indian Sign Language (ISL). ISL isn't used in deaf schools to teach deaf children, though. Academics are not encouraged to teach in methods that incorporate ISL by teacher coaching programmes.

Language communication does not include any teaching materials. Parents of deaf children are often unaware of language communication's power to break down obstacles to communication. For the typical person, sign language is difficult to understand. According to the UN, around 5 percent of the world's population, or 466 million people, are dumb or deaf. Like all other languages that are internationally recognised, sign language could be a language. It is a form of non-verbal communication that the deaf and the mute use to communicate with one another and with others. Without speaking, it enables people to express themselves and understand one another. There are several different sign languages, including Chinese Sign Language, American Sign Language, and Indian Sign Language. American English is currently the most commonly used language worldwide. By identifying the sign victimisation, we are putting out a simple system to translate Indian verbal communication into text and then into speech.

Dumb individuals use hand signs to speak, therefore other normal individuals face drawbacks in recognizing their language by signs created. Therefore, there's a requirement of the system that acknowledges the different signs and conveys the knowledge to the conventional individuals within the style of text likewise to speech. Most people who have problems with speech and hearing,

978-1-6654-9790-9/22/$31.00 ©2022 IEEE

face a great deal of challenges in their day-to-day life. Further, to speak with others they struggle a lot by creating hand sign or by victimization body gestures and even when doing of these the one who is on the receiving side finds it difficult to grasp. So, as an associate degree innovative plan to unravel this drawback we have come up with this paper to make everyone's life easier.

The main objectives which we want to achieve from this paper are as follows:

1. To develop an associate degree automatic language recognition system with the help of a convolution neural network and laptop vision techniques.
2. To use natural visual sequences, without requiring the signer to wear coloured or informational gloves, and to be able to recognise a variety of signals.

The organization of remaining sections are as follows: section II explains the detail study of literature survey of the previous work, Section III explain the methodology of the proposed method, Section IV explain the result of the proposed method, section V conclude the proposed method with future work.

II. RELATED WORK

At first, we summarized the literature that used CNN as the classifier [1, 2, 7]. Communication is achieved through sign languages, which are used by the deaf to communicate through hand gestures, body movements, facial expressions, and body movements [3]. The details survey of the sign language to recognition explained by researchers [4, 5]. The architecture was designed to overcome the challenges faced by automatic recognition machines. It uses deep convolutional encoder-decoder architecture for capturing spatial information and LSTM architecture for capturing temporal information [6]. They choose a vocabulary of 14 one-handed signs from the LSA64 Dataset and achieved an accuracy of about 96.02% and 77.85% for the top three predictions in signer-dependent and signer-independent settings respectively [7].

Most papers discussed CNN, KNN, and SVM classifiers for sign language recognition. Rekha et al proposed the system using KNN and SVM classifiers on the ASL dataset and achieved the accuracy of 89.1% and 91.7% respectively. Agarwal et al introduced a method based on SVM and achieved an accuracy of 95.3% on the ISL dataset. He also introduced a method based on the Bayesian KNN algorithm which got an accuracy of 89.9% on the same dataset. Lilha and Sivamurthy introduced a method based on SVM and the accuracy was found to be 98.1% using ISL dataset. A vision based isolated sign language recognition model is developed. The model uses a convex hull for feature extraction and KNN classification. It is based on isolated hand gesture detection and recognition [8, 9].

Dutta et al proposed machine learning algorithm to recognize the sign language using MATLAB [10]. Neel et al. proposed convolutional neural network & deep learning algorithm for Indian sign Language processing and accuracy of 98.81%[11]. Muthu et al. work on sign language recognition for real time recognition [12]. Archana et al. proposed dynamic hand gesture recognition algorithm for Indian sign language using discrete time wrapping [13].

Kumud et al. working on the continuous gesture recognition and formation of the sentences using machine learning algorithm [14].

Shashidhar et al. proposed the two dimensional convolutional neural network for English Alphabets using Indian sign language recognition [15] using data set details are given in the weblink [16]

III. PROPOSED METHOD

This section will explain the approach we used in our paper with step by step procedure and a detailed explanation of the optimized model created with its mathematical expressions and the reason behind their usage. Fig.1 shows the block diagram is shown for an easier understanding.

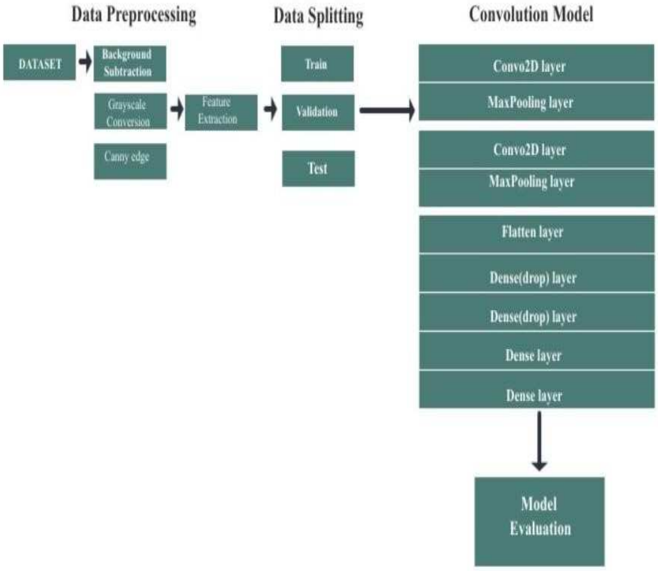


Fig 1. Block Diagram of Methodology

A. Dataset:

We have used the Indian sign language dataset which is an open-source dataset available on the Kaggle server. The dataset contains 35 classes of images which is from 1 – 9 and from A- Z. each class contains 1200 images. The dataset contains around 42000 images and we have divided that dataset in the ratio of 0.8: 0.2 to get the train images and test images.

B. Pre-Processing:

After importing the dataset, we need to do some modifications to that image before training. Since grayscale images contain only two values either 0 or 1 and colors are of not any importance, we have converted all the images to grayscale images. To remove the high-frequency component in the images we have used a gaussian filter. To get the desired image for training which contain only highlighted required edge of the sign, we have done some filtering which includes adaptive threshold filtering. To get the features 128 * 128 image is sufficient, we have resized the image to 128 * 128 pixels.

C. Model Architecture

We have designed the CNN sequential model for Indian sign language recognition. It contains 2 hidden layers, flatten layer, 3 dense layers with the ReLu activation function, and 1 dense layer with a softmax activation function to get the feature of every class in the dataset. A model summary is shown in the below table.

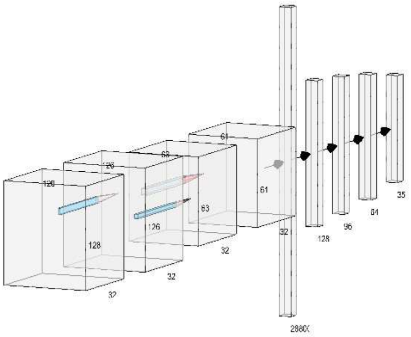


Fig 2. CNN Model

TABLE 1. DETAILED LAYER DESCRIPTION OF THE CNN MODEL

Layers	Type	Kernel size	Output shape	Parameters
Layer 1	Conv2D	3 x 3	126 x 126 x 32	320
Layer 2	Max-pool	2 x 2	63 x 63 x 32	0
Layer 3	Conv2D	3 x 3	61 x 61 x 32	9248
Layer 4	Max-pool	2 x 2	30 x 30 x 32	0
Layer 5	Flatten	-	28800	0
Layer 6	Dense 1	-	128	3686528
Layer 7	Dropout	-	128	0
Layer 8	Dense 2	-	96	12384
Layer 9	Dropout	-	96	0
Layer 10	Dense 3	-	64	6208
Layer 11	Dense 4	-	35	2275

A. Kernel Initialization

Gradient descent-based learning has one major drawback: the model could diverge or hit a local minimum point. In order to obtain convergence, intelligent weight initialization is necessary. These weights are represented in CNN as kernels (or filters), and one convolution layer is made up of several kernels. Utilizing xavier initialization is our suggested CNN model. This initializer generally balances the gradient scale across all kernels. The following range of initial weights is chosen at random: 6 In + 6 Out = x; [x, x]. The number of input and output units at the kernel weights are In and Out, respectively. Today, it is understood that the prior method is inferior to a small kernel with deeper layers. Therefore, we initialized our kernel size with 3 x 3 which is generally used in modern CNN classifiers.

B. Activation Function

It is used to decide if neural networks will produce a yes or no response. Depending on the function we employ, the result values are mapped to values between 0 and 1 or -1 and 1.

Nonlinear activation, such as rectified linear units (ReLU), leaky rectified linear units (LReLU), and exponential linear units, is frequently utilised in contemporary CNN models (ELU). We employed the ReLU activation function and the softmax activation function in our model.

a. ReLU

$$(x) = \max \quad (0, x) \qquad (1)$$

where x denotes input to the neuron and sparse categorical cross-entropy loss function is defined as,

$$\text{loss} = \sum_{i=1}^{m} yi \, . \log_{yi} \qquad (2)$$

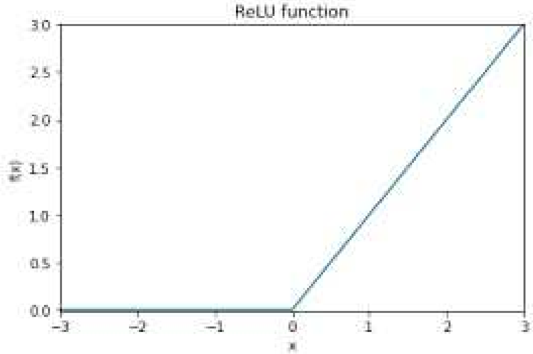


Fig 3. ReLU Activation function

C. Softmax

An activation function called Softmax scales numerical values into probabilities. A vector containing the probabilities of each potential result is the result of a softmax. All potential outcomes or classes are represented by a vector
of probabilities that add up to one .It is defined as,

$$S(y) = \frac{exp(yi)}{\Sigma exp(yj)} \qquad (3)$$

IV. Results Analysis

This section will show the results of the paper which includes accuracy graphs, losses graph, confusion matrix, given input ISL, and obtained output in text and speech. The training accuracy of the graph is found to be 0.99 after the 13th epoch before the 13th epoch it starts increasing from zero and the validation accuracy as observed in the graph is constant for all the epochs which are equal to 1.

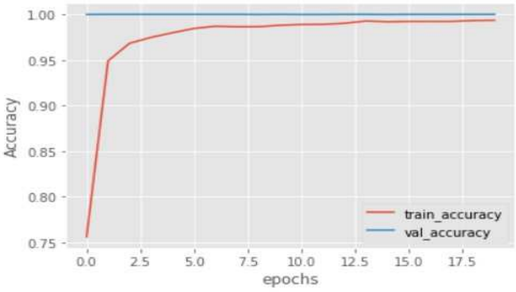


Fig 4. Accuracy graph

The training loss during the first epoch is 0.80 and it gradually decreases as seen in the graph and reaches a value of 0.0000000001 in the 17th epoch, the validation loss remains constant at zero.

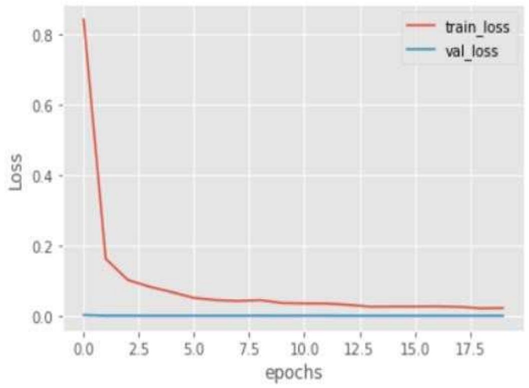


Fig 5. Losses graph

The model was saved and it consisted of 35 classes corresponding to numbers 0 to 9 and alphabets A to Z. The confusion matrix was plotted as shown in Fig. 6.

The below Indian Sign Language was given as the input which corresponds to GOODMORNING and the system accepted the input and showed output as shown in fig 7. The output of the Indian Sign Language to Text Conversion, the input is given as GOODMORNING using Sign Language.

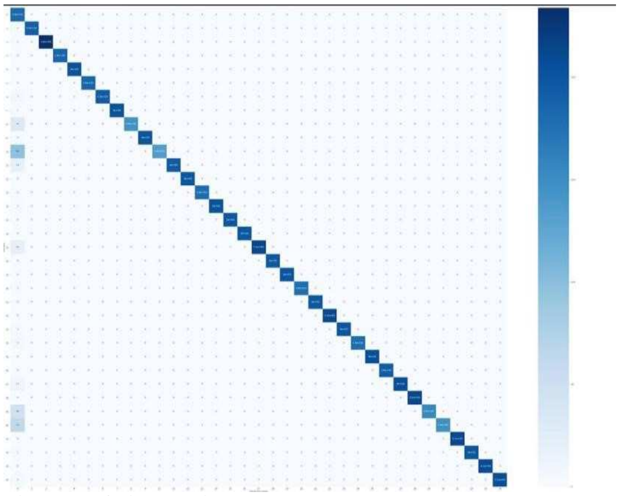


Fig. 6. Confusion Matrix

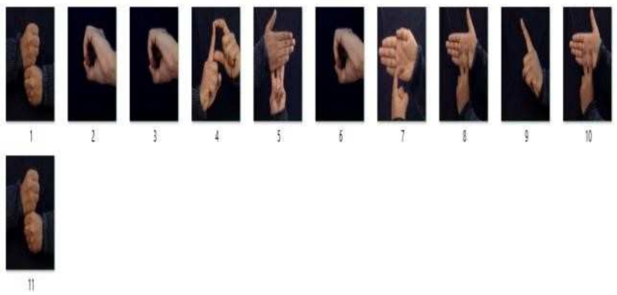
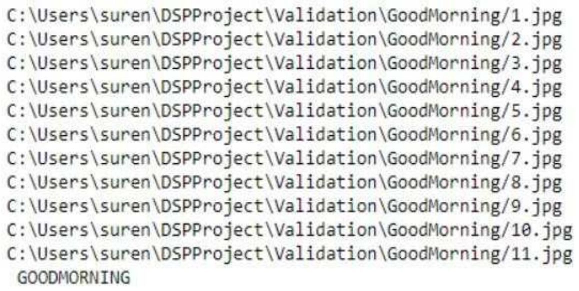


Fig. 7. The input of Indian Sign Language

The given input as shown in Fig 8 gives the audio output and is also shown in Fig 9 by using Google Text to Speech API. Table 2 gives the comparison of different algorithms used with their accuracy rate. Since these works have a different number of the test set, all parameters are to be considered and the efficient method can be chosen.



```
C:\Users\suren\DSPProject\Validation\GoodMorning/1.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/2.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/3.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/4.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/5.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/6.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/7.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/8.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/9.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/10.jpg
C:\Users\suren\DSPProject\Validation\GoodMorning/11.jpg
 GOODMORNING
```

Fig 8. Output for the given input of Indian Sign Language in Text

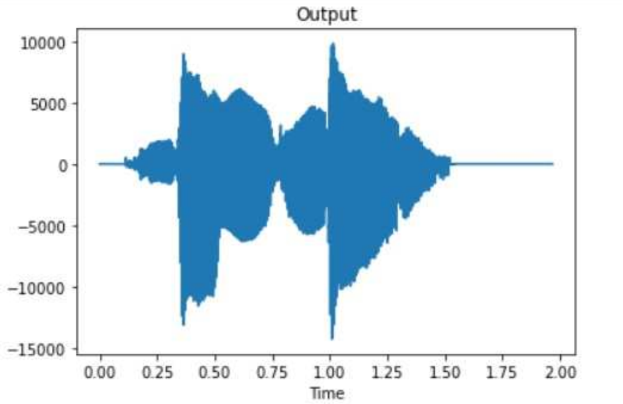


Fig 9. Output for the given input of Indian Sign Language in Speech

TABLE 2. COMPARISON TABLE

Work	Proposed Method	Accuracy Obtained (%)	Dataset
Sabeenian [8]	KNN	97%	ArSL
Yogeshwar et al.[1]	SVM	92.2%	ISL
Likhar[2]	LSTM	97.71%	ISL
Shashidhar et al[15]	2D CNN	95%	ISL
Proposed Method	CNN	99%	ISL

V. CONCLUSION

By adding computing to the communication path, our initiative intends to facilitate communication between deaf and dumb persons and other people by automatically capturing, recognising, translating, and displaying language as text. The procedures for converting languages vary by unit. Some of them employ a wired electronic glove, while others employ a visual strategy. Electronic gloves are highly expensive, and one person cannot wear another person's glove. Different techniques are used to recognise and match the collected gestures with gestures in the information in a vision-based approach. It is simple, cheap, and reliable to convert an RGB image to binary and match it with information using a comparing algorithmic software. This technique effectively converts language to text. Once converted to text, we usually renew the text into speech to increase the usefulness and dependability of our initiative for society. Future scope of the proposed work is Develop an android or iOS based application to use this project in day to day life of an individual. Increase the versatility of this project by adding more features to the sign language to detect more words and symbol.

References

[1]. Yogeshwar I. Rokade, Prashant M. Jadav "Indian Sign Language Recognition System" International Journal of Engineering and Technology (IJET), Vol. 9, No.3, 2019. Doi:http://dx.doi.org/10.21817/ijet/2017/v9i3/170903S030

[2]. P. Likhar, N. K. Bhagat and R. G N, "Deep Learning Methods for Indian Sign Language Recognition," 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), 2020, pp. 1-6, doi: 10.1109/ICCE-Berlin50680.2020.9352194.

[3]. Elakkiya, R. Machine learning based sign language recognition: a review and its research frontier. J Ambient Intell HumanComput 12, 7205–7224 (2021). https://doi.org/10.1007/s12652- 020-02396-y

[4]. Cheok, M.J., Omar, Z. & Jaward, M.H. A review of hand gesture and sign language recognition techniques. Int. J. Mach. Learn. & Cyber. 10, 131–153 (2019). https://doi.org/10.1007/s13042-017-0705-5

[5]. M. Safeel, T. Sukumar, Shashank. K. S, Arman M. D, Shashidhar R and Puneeth S. B, "Sign Language Recognition Techniques- A Review," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-9, doi: 10.1109/INOCON50539.2020.9298376.

[6]. Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr.Dayananda P, "Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network", International Journal of Engineering Research & Technology (IJERT), NCAIT, volume 8, Issue 15, 2020.

[7]. Jai Shah, "Deepsign: A Deep-Learning Architecture For Sign Language Recognition", The University Of Texas At Arlington December 2018

[8]. R.S. Sabeenian, S. Sai Bharathwaj, M. Mohamed Aadhil, Sign Language Recognition Using Deep Learning and Computer Vision, Jour of Adv Research in Dynamical & Control Systems, Vol. 12, 05-Special Issue, 2020

[9]. K. Amrutha and P. Prabu, "ML Based Sign Language Recognition System," 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), 2021, pp. 1-6, doi: 10.1109/ICITIIT51526.2021.9399594.

[10]. K. K. Dutta and S. A. S. Bellary, "Machine Learning Techniques for Indian Sign Language Recognition," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), 2017, pp. 333-336, doi: 10.1109/CTCEEC.2017.8454988.

[11]. N. K. Bhagat, Y. Vishnusai and G. N. Rathna, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning," 2019 Digital Image Computing: Techniques and Applications (DICTA), 2019, pp. 1-8, doi: 10.1109/DICTA47822.2019.8945850.

[12]. H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.

[13]. Archana Santosh Ghotkar, Gajanan Kashiram Kharate "Dynamic Hand Gesture Recognition and Novel Sentence Interpretation Algorithm for Indian Sign Language Using Microsoft Kinect Sensor," Journal of Pattern Recognition ResearchVol 10, No 1, pp 24-38,2015 doi:10.13176/11.626

[14]. Kumud Tripathi, Neha Baranwal and G. C. Nandi, "Continuous Indian Sign Language Gesture Recognition and Sentence Formation", Eleventh International MultiConference on Information Processing (IMCIP-2015), Procedia Computer Science, Vol 54, pp.523 – 531,2015. Doi: https://doi.org/10.1016/j.procs.2015.06.060

[15]. Shashidhar R, Arunakumari B. N., A S Manjunath, Roopa M, "Indian Sign Language Recognition Using 2-D Convolution Neural Network and Graphical User Interface", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.14, No.2, pp. 61-73, 2022.DOI: 10.5815/ijigsp.2022.02.06

[16]. Dataset available online: https://www.kaggle.com/prathumarikeri/indian-signlanguage