ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

INDIAN SIGN LANGUAGE RECOGNITION USING HAND-POSE KEY POINTS AND TRANSFER LEARNING

Shilpa Ingoley[1]* and Jagdish Bakal[2]
[1]PhD Scholar and [2]Principal, Pillai HOC College of Engineering and Technology, Rasayani, India
[1]shilpaingoley1@gmail.com, [2]jwbakal@mes.ac.in
[1]ORCID ID: 0009-0006-2592-6282 and [2]ORCID ID: 0009-0003-4952-127X

ABSTRACT
We want world class facilities like good infrastructure, transportation, housing, healthcare etc. in Indian cities. These ultra-modern cities should be well equipped and friendly for everyone including Divyangjan-differently abled people. One of the challenges is to facilitate Divyangjan, especially the people who are deaf and mute(unable to speak). People with speaking abilities convey their thoughts, ideas and share their experiences by getting vocal while interacting with the people around them. Vocal language is one of the primary media of communication used by the human. However, the people who cannot speak and hear, use sign language to convey their views and emotions. But majority of the people have trouble in understanding sign language. This creates a barrier in the communication process. To fill this gap, the projected system converts the alphabets signs' of ISL(Indian Sign Language) into text. In ISL, some alphabets have single way of representation while some have two or more ways of representing gestures. For example, alphabet 'A' has single gesture while alphabet 'E' can be sign using three different ways. The primary objective of a communication system is to convey the message that a person wishes to express. To implements this, the proposed system makes uses of the concept of keypoints of hand gesture and transfer learning techniques. The suggested solution provides validation accuracy of 99.97% and is also appropriate for situations encountered in daily life. Additionally, we advise installing kiosks in smart cities with interpreting software set up at various significant locations to have cordial conversation with them.

Keywords: Indian Sign Language(ISL), Machine Leaning, Computer Vision, Deep Leaning(DL), Hand Gesture Recognition(HGR), Convolution Neural Network(CNN).

INTRODUCTION
India is developing rapidly in cities and villages. In order to create "smart cities" where everyone, including those with disabilities, feel inclusive and for everyone state of art facilities should be available. To promote these ideas in future smart cities this article addresses a solution of communication among differently abled people, especially speech and hearing-impaired people. Communication is a fundamental necessity for humans. However, the people who are hard of hearing and are unable to speak, use visual form of communication that is a sign language. When these people wish to communicate with rest of the society who do not understand sign language, creates barriers in the communication process. Without the knowledge of sign language(SL), it's really challenging to communicate with such people. Different gestures of SL are made using hands, fingers, arms and head along with facial expressions. Practically each country has their own SL, few examples are- American SL, Chinese or Mandarin SL, German SL, Swiss SL, Turkish SL, Argentinian SL, Arabic SL,Thai SL etc.

The proposed effort is based on Indian sign language(ISL). It identifies alphabets used in ISL. Recognition of alphabets would be really helpful in sign languages in mentioning names of people, specify various places, to identify brands of products, movie titles etc. Like in vocal languages, we do have some variations region wise. Similarly, ISL is not an exception. In ISL, some alphabets have single gesture while some have multiple ways of representation. For example, alphabet 'A' has single gesture, 'B' has two ways of representation and while alphabet 'E' can be sign using three or more ways. The chief goal of communication system is to identify and convey the meaning of what a person wishes to communicate. To implement this, the proposed system works on static images and uses the transfer learning techniques and the concept of key-points of hand pose.

LITERATURE REVIEW

SL interpretation is one of the common applications of gesture recognition(GR). Considerable work is done in the field of SL recognition. Signs to be recognized can be either static or dynamic. Static signs require us to make a pose whereas for dynamic signs movement of hands and/or fingers is needed to interpret the sign. Some signs are performed by single hand while some signs use both the hands. For identification of signs, many authors have proposed solution which can be largely classified into following types:

Glove-Based: This system uses wired or wireless glove with sensors[1-3] or colored based gloves without sensors[4]. However, wearing gloves may not be comfortable. Facial expressions/gestures cannot be identified with this method.

Vision-Based: This system makes use of camera of computer/laptop[5-9] or mobile[10]. In this type, comparatively complex computation must be performed on the images/videos.

[1] In vision based system recognition of signs can be of two types static or dynamic. Mostly, static gesture recognition makes uses of images[11-13] whereas dynamic gesture uses videos[14 -18] as a input. Authors[17] collected video dataset of 8 emergency signs, performed by 26 subjects at varied conditions. Used of deep-learning technique such as CNNand LSTM were used by[15] for word-level SLR. Twenty dynamic ISL gesture performed in complex-backgrounds by ten subjects were identified by [18].

Significant volume of effort has been carried on American SL[1][5][19][20]. Fernandes et al., 2020 [1] uses glove-based as well as vision-based approach. As glove-based system's accuracy was not high, afterwards the accuracy was improved using CNN.

On ISL, many authors [14][16][21-22][24-25] have done enormous contribution. ISL alphabet uses both single and mostly double handed signs. There is misconception that ISL is Hindi language, however it considers an English language. Reference [2] proposed a hardware glove based system, which produces artificial speech with the help of a Bluetooth. Their system can be called as "Talking Hands". In [22] recognize a subset of ISL, which is static single hand gestures. This paper also briefly mentions ISL and its dialects, varieties and current exertions in the course of its standardization. In their paper they mentioned about the various regional dialects of ISL within India and below are some of them:

Mumbai-Delhi SL , Calcutta SL, Bangalore-Madras SL

In [14] created the database which contains in all 130,000 videos and use segmentation-based technique for implementation. Katoch et al., 2022 [21], developed a system on static HGR using the concept of keypoints. In [26], emphasis is placed on training ML/DL models using both left as well as right hand to ensure accurate sign interpretation whether signs are conducted by a left or right-handed person.

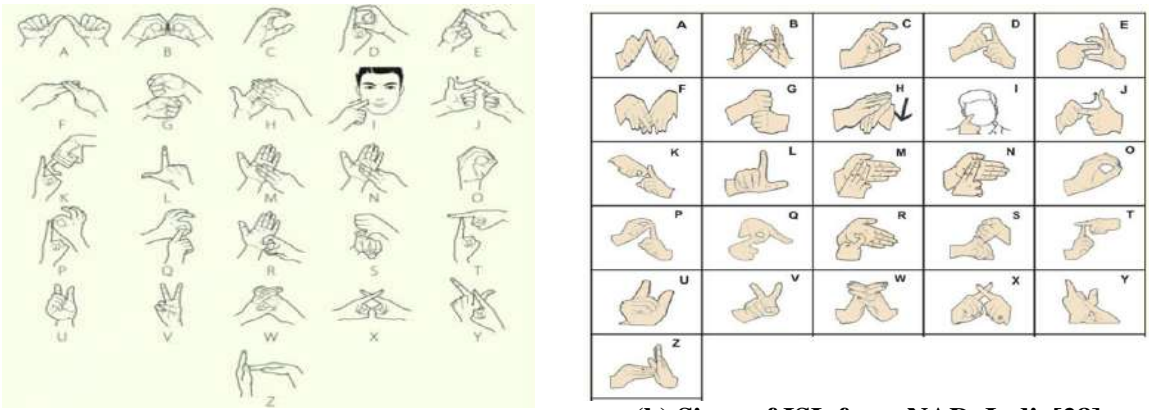
Proposed ISLARS and Its Modules

Projected ISLARS system has several modules which are explored in the following section:

ISL Alphabet Sign Gestures. The objective of the ISLARS is to recognize the alphabets(A-Z) of ISL. The goal is to interpret the correct meaning and to convey the information. If multiple ways of representations are used for same alphabetic gestures, then it should correctly identify that alphabet. Otherwise, system will misinterpret or wrongly classify the alphabet. To support this statement, lets observe the following ISL alphabetic signs used by the different authors:

'B', 'C', 'E', 'I', 'J', 'O', 'W' and 'Z' [2][21][23][24]

For the comparison purposes we have taken only two ISL sign picture images which can be seen in Fig.1. Fig. 1(a) is the poster of the manual *Alphabet* in ISL taken from *ISL Research and Training centre* (*ISLRTC*)[27]. Fig. 1(b) shows signs of ISL from *National Association of the Deaf(NAD)*[28]. From these two pictures, we can observe that in the first row of images, the formation/appearance of letters B, C and E are different.



(a) Signs of ISL from ISLRTC[27] **(b) Signs of ISL from NAD, India[28]**

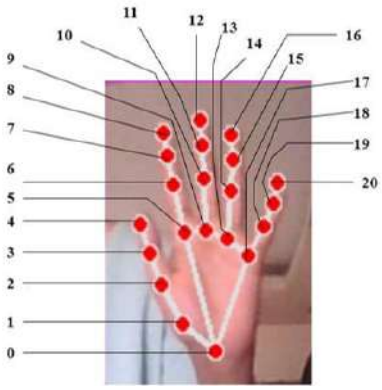
Fig. 1 Some ISL signs with difference representations

From gesture recognition point of view, the signs made for 'B' in Fig. 1(a) and Fig. 1(b) are different. Similarly, one can observe that the formation of shapes is different for signs 'C' as well as 'E'. For Sign Recognition System, these things matter. Otherwise, it will incorrectly classify gesture. From the above discussions, we have considered multiple signs for classification of the same alphabet as indicated in Table 1.

Table 1: ISL Alphabets and number of signs considered

ISL Alphabets	Numbers of signs considered	ISL Alphabets	Numbers of signs considered	ISL Alphabets	Numbers of signs considered	ISL Alphabets	Numbers of signs considered
A	1	H	1	O	1	V	1
B	2	I	2	P	1	W	2
C	2	J	2	Q	1	X	1
D	1	K	1	R	1	Y	1
E	3	L	1	S	1	Z	2
F	1	M	1	T	1		
G	1	N	1	U	1		

Hand Pose Keypoints Detection. The proposed work uses key point detection of hand(s) and can be term as hand 'landmarks' detection. This type of detection of the landmarks on the hands can be applied to images and videos. Fig. 2 shows the key point for hand pose detection. For a single hand, there are 21 points in total, which are labelled from 0 to 20. Same set of points can be detected from other hand. For implementation, we have used media-pipe library's hand-pose model, which is an open source[29].



0:WRIST	11:MIDDLE_FINGER_DIP
1:THUMB_CMC	12:MIDDLE_FINGER_TIP
2:THUMB_MCP	13:RING_FINGER_MCP
3:THUMB_IP	14:RING_FINGER_PIP
4:THUMB_TIP	15:RING_FINGER_DIP
5:INDEX_FINGER_MCP	16:RING_FINGER_TIP
6:INDEX_FINGER_PIP	17:PINKY_FINGER_MCP
7:INDEX_FINGER_DIP	18:PINKY_FINGER_PIP
8:INDEX_FINGER_TIP	19:PINKY_FINGER_DIP
9:MIDDLE_FINGER_MCP	20:PINKY_FINGER_TIP
10:MIDDLE_FINGER_PIP	

Fig. 2 Key Point of hands with numbers and their names

Building Model - Steps in creating ISLARS. Main steps and flow to build an ISL Alphabets Recognition System are mentioned in Fig. 3.

a) Image Acquisition: Images are acquired using camera of laptop. No special camera like 3D/depth is used for capturing different sign images. Images are captured along with hand-key-points.

b) Image Pre-processing and Re-shaping: After identifying hand(s), bounding-box is formed around the hand(s). Later, the image is cropped. All the images are of different size because of the different hand gestures as well as use of single and double hands in formation of ISL signs. Due to this, the dimensions of all the images vary. Hence, they are converted to same size of 'M*M' using the concept of padding. Image normalization is done through rescaling.

c) Creating Dataset: To create dataset of ISL alphabets as mentioned Table 1, we have captured images for 34 signs of alphabets. Whole dataset is created with the assistance from the three signers. Multiple signers performed hand gesture in various background and illumination. Images are also captured at varied distance from the camera.

d) Trained the Model using Transfer Learning Technique: We have used transfer learning technique for implementation of our ISLARS system. Vgg16 pre-trained model has been used. CNN is best for images and Vgg16 makes uses of CNN. We have done some customization before training. Vgg16 is 16-layers deep. It has thirteen-convolutional layers, five-max-pooling layers, and three-fully-connected layers. We have feezed the first thirteen-layers and trained last 3 dense layers.

e) Validating and Testing the Model for Accuracy: Dataset was kept balanced by taking same number of images for all the classes/signs. For training and validating, the dataset is separated into ratio of 70:30. Out of these, 70% are used for training, while the remaining 30% are considered for validation.

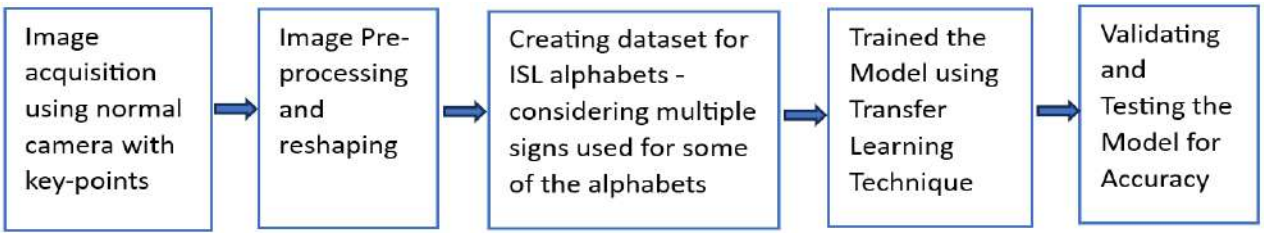


Fig. 3 Flow of the ISLARS

Experimental Results and Analysis

Implementation Details. As mentioned in Table 1, number of classes would be 34. We have labelled alphabets in the following manner. If an alphabet uses single sign gesture then we have labelled it as it is. If an alphabet is represented with two dissimilar ISL signs, then we have used labels like B and B'. Suppose, alphabet E is represented with 3 different signs then labelling is done likewise namely E, E' and E''. This is done to understand that we are taking different sign gestures for the same alphabet. For each of the above alphabet we have taken 450 signs(images). Out of which we have kept 315(70%) images for training and 135(30%) of the images for validation.

- Total number of images would be 34*450 = 15300
- Total 34*315=10710 images belonging to 34 classes used for Training
- Total 34*135=4590 images belonging to 34 classes used for Validation

The resolution of every image in the dataset is change to 224*224*3 before giving it for training. Where '3' represents RGB-channels. For the first two dense layers after layer 13, we have kept activation function as 'Relu' and for the last dense-layer activation function used is 'Softmax,' being a multiclass classification problem. To

trained the model batch size is kept as 64. Losses are measured using 'categorical_crossentropy'. Optimizer used is 'adam' since it convergences fast.

Performance Matrix. To evaluate the performance of suggested work, we have considered confusion matrix, as indicated in Fig. 4.(a) and classification report is display in Fig. 4(b). To analyzed and assess the performance of suggested work, below measures are taken into consideration:

$$Accuracy(A), \quad \text{Recall(R)}, \quad \text{Precision(P) and } F(F1\_Score)$$

Accuracy is the most indispensable performance criteria. Formulas are shown in Eq. 1 to Eq. 4, where TN is True_ Negative, TP→True_ Positive, FN→False_Negative and FP→False_ Positive.

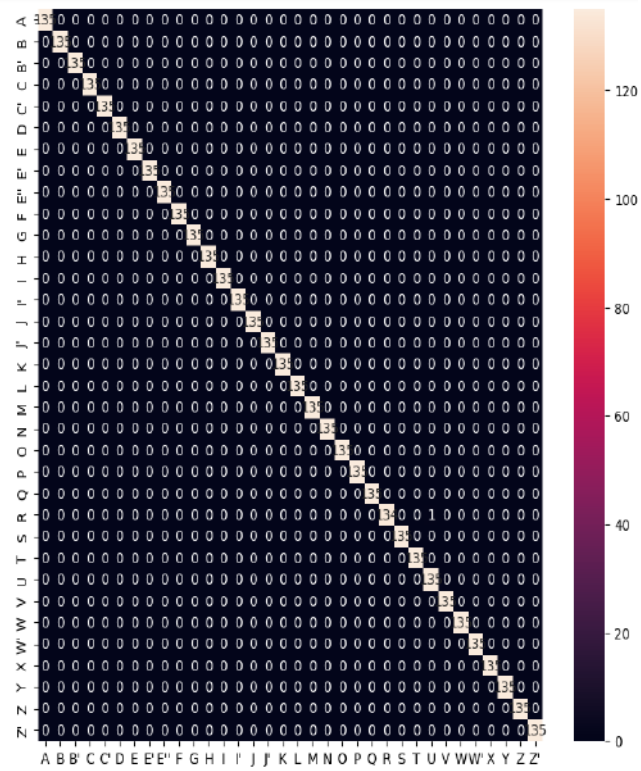
$$A(Accuracy)=\frac{\text{correctly identified signs}}{\text{total number of signs}}=\frac{(TN+TP)}{(TN+TP+FN+FP)} \quad (1)$$

$$P(Precision)=\frac{TP}{(FP+TP)} \quad (2)$$

$$R\ (Recall\ )=\frac{TP}{(FN+TP)} \quad (3)$$

$$F(F1\text{-}Score\ )=2*\frac{(P*R)}{(R+P)} \quad (4)$$

Using vgg16 model, it is observed that for our ISLARS system an overall 100% accuracy is achieved on the training dataset set whereas validation accuracy is 99.97%. The class-wise achieved accuracy is indicated in Fig. 4(b).



(a)Confusion matrix for true label and predicted labels

	precision	recall	f1-score	support
A	1.00	1.00	1.00	135
B	1.00	1.00	1.00	135
B'	1.00	1.00	1.00	135
C	1.00	1.00	1.00	135
C'	1.00	1.00	1.00	135
D	1.00	1.00	1.00	135
E	1.00	1.00	1.00	135
E'	1.00	1.00	1.00	135
E''	1.00	1.00	1.00	135
F	1.00	1.00	1.00	135
G	1.00	1.00	1.00	135
H	1.00	1.00	1.00	135
I	1.00	1.00	1.00	135
I'	1.00	1.00	1.00	135
J	1.00	1.00	1.00	135
J'	1.00	1.00	1.00	135
K	1.00	1.00	1.00	135
L	1.00	1.00	1.00	135
M	1.00	1.00	1.00	135
N	1.00	1.00	1.00	135
O	1.00	1.00	1.00	135
P	1.00	1.00	1.00	135
Q	1.00	1.00	1.00	135
R	1.00	0.99	1.00	135
S	1.00	1.00	1.00	135
T	1.00	1.00	1.00	135
U	0.99	1.00	1.00	135
V	1.00	1.00	1.00	135
W	1.00	1.00	1.00	135
W'	1.00	1.00	1.00	135
X	1.00	1.00	1.00	135
Y	1.00	1.00	1.00	135
Z	1.00	1.00	1.00	135
Z'	1.00	1.00	1.00	135
accuracy			1.00	4590
macro avg	1.00	1.00	1.00	4590
weighted avg	1.00	1.00	1.00	4590

(b)precision, recall and f1-score on validation dataset

Fig. 4: Performance matrix on validation dataset

Fig. 5. show results of testing carried out in varied background. In spite of varied background and colour used, the result clearly shows the ISLARS has correctly predicted the given ISL sign alphabets.

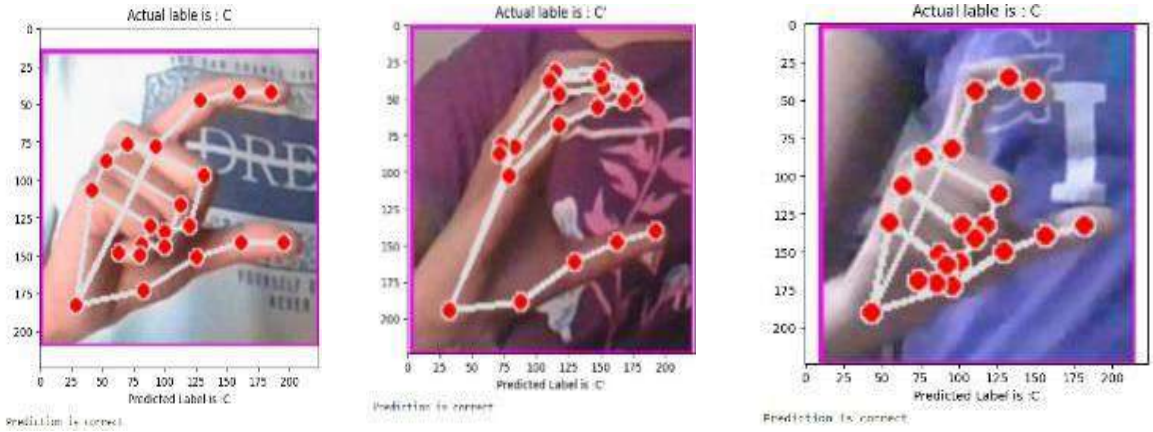


Fig. 5: Testing of model on three different images of C (C,C',C) with varied background

SUMMARY

One of the most important technological advancements that is required in our future cities is to facilitate differently-abled people. To make speech and hearing-impaired people an inseparable part of our society, we have suggested ISLARS system. The affected people use visual language known as sign language. When these people wish to communicate with normal people, signs made by them needed to be converted into text or speech. Aim of this research is to recognize the alphabets of ISL. Finding correct sign is important as finger spelling is useful in many situations like mentioning names, places etc. For some of the alphabets in ISL uses multiple signs. To interpret it correctly the proposed ISLARS system gives consideration for the same. Thus, our system assists to remove the communication-barrier between impaired people and healthy individuals. This effort makes use of the concept of hand pose key points. It builds the model using transfer learning technique, vgg16 model. Training accuracy is 100% and validation accuracy of the system is 99.97%. It's a user-independent system also capable of handling diverse background conditions.

ACKNOWLEDGEMENTS

We are thankful to Ms. Neeta Mukherji, a certified ISL-trainer of Jhaveri Thana wala School for the Deaf, Thane(w), for her valuable time and for clarifying our doubts about ISL.

REFERENCES

[1] Fernandes, L., Dalvi, P., Junnarkar, A., Bansode, M. (2020). Convolutional neural network based bidirectional sign language translation system. *ICSSIT 2020*, 769–775.

[2] Heera, S. Y., Murthy, M. K., Sravanti, V. S., Salvi, Talking hands - An Indian sign language to speech translating gloves. *IEEE, ICIMIA 2017*, 746–751.

[3] Sharma, S., Gupta, R., Kumar, Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition. Journal of Ambient Intelligence and Humanized Computing, (2022) *13*(7), 3527–3537.

[4] Kanvinde, A., Revadekar, A., Tamse, M., Kalbande, D. R., Bakereywala, N., Bidirectional Sign Language Translation. *Proceedings - ICCICT 2021*.

[5] Amrutha, K., Prabu, P., ML based sign language recognition system. ICITIIT 2021.

[6] Zheng, J., Chen, Y., Wu, C., Shi, X., Kamal, Enhancing Neural Sign Language Translation by highlighting the facial expression information. (2021) 462–472.

[7] Sharma, A., Panda, S.,Verma, Sign Language to Speech Translation. ICCCNT 2020.

[8] Harini R, Janani R, Keerthana S, Madhubala S, Venkatasubramanian S, "Sign Language Translation" ICACCS 978-1-7281-5197-7/20/ 2020 IEEE

[9] Adithya, V., Rajesh, A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition. (2020) 2353–2361.

[10] Ku, Y. J., Chen, M. J., King, A virtual sign language translator on smartphones. (2019) 445–449.

[11] Shenoy, K., Dastane, T., Rao, V., & Vyavaharkar, *Real-time Indian Sign Language (ISL) Recognition*, IEEE – 43488, 2018

[12] Bora, J,Dehingia,S,Boruah, A, Chetia, A A, Gogoi, Real-time Assamese Sign Language Recognition using MediaPipe & Deep Learning, (2023) *218*,1384–1393.

[13] Sreemathy, R., Turuk, M. P., Chaudhary, S., Lavate, K., Ushire, A., Khurana, Continuous word level sign language recognition using an expert system based on machine learning. *International Journal of Cognitive Computing in Engineering* (2023) 170–178.

[14] Purva C. Badhe, Vaishali Kulkarni "Indian Sign Language Translator Using Gesture Recognition Algorithm", CGVIS 2015 IEEE

[15] Du, Y., Xie, P., Wang, M., Hu, X., Zhao, Z., Liu, Full transformer network with masking future for word-level sign language recognition. (2022) *500*, 115–123.

[16] Sridhar, A., Ganesan, R. G., Kumar, P., & Khapra, INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition. *Proceedings of the 28th ACM International Conference on Multimedia*, (2020) 1366–1375.

[17] Adithya, V., Rajesh, Hand gestures for emergency situations: A video dataset based on words from Indian sign language. 2020.

[18] Singh, D. K., 3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling. *Procedia CIRP*, (2021) *189*, 76–83.

[19] Ardiansyah, A., Hitoyoshi, B., Halim, M., Hanafiah, N., Wibisurya, Systematic Literature Review: American Sign Language Translator. *Procedia Computer Science*, (2021) *179*, 541–549.

[20] He, Research of a Sign Language Translation System Based on Deep Learning. *International Conference on, AIAM 2019*, 392–396.

[21] Katoch, S., Singh, V., Tiwary, Indian Sign Language recognition system using SURF with SVM and CNN. 2022.

[22] Futane, P. R., Dharaskar, "Hasta Mudra": An interpretation of Indian sign hand gestures. *ICECT 2011 - 2*, 377–380.

[23] Rokade, Y. I., Jadav, Indian Sign Language Recognition System. *International Journal of Engineering and Technology*, 2017, 189–196.

[24] Grover, Y., Aggarwal, R., Sharma, D., Gupta, Sign language translation systems for hearing/speech impaired people: A review, *ICIPTM 2021*, 10–14.

[25] Sharma, A., Panda, S., Verma, Sign Language to Speech Translation. *ICCCNT 2020*.

[26] Ingoley, S. N., Bakal, Use of Key Points and Transfer Learning Techniques in Recognition of Handedness Indian Sign Language. IJRITCC *2023*, 535–545.

[27] Information on https://islrtc.nic.in/poster-manual-alphabet-isl

[28] Information on https://ceobihar.nic.in/pdf/Fingerspelling_Alphabet.pdf

[29] Information on https://developers.google.com/mediapipe/solutions/vision/hand_landmarker