A Comprehensive Review on Indian Sign Language Recognition System using Vision based Approaches

Poornima B.V.
SJCE, JSSSTU
Mysuru

Srinath S.
SJCE, JSSSTU
Mysuru

ABSTRACT
A sign language recognition system is a method in which a computer automatically recognizes the sign language motions and converts them into machine or human readable text or speech. Many researchers have proposed different algorithms for identifying the static and dynamic Indian sign language (ISL) gestures. This review presents a qualitative and a comprehensive study of the different approaches like digital image processing, machine learning and deep learning methods used for recognition of gestures. Research publications from the past 10 years have been collected from electronic databases like scopus, google scholar and researchgate for the review and the details of the publicly available dataset repositories are highlighted. This review helps the researchers, academicians and the technology oriented people to understand the importance of different technologies used to recognize the gestures automatically which in turn benefits the speech and hearing impaired people. The challenges present in ISL recognition, the short comings of the existing systems and the future research directions in order to improve the recognition rate is explained in this paper.

General Terms
Image processing, Pattern recognition, Machine learning.

Keywords
Sign language recognition (SLR), Gesture recognition system, Indian sign language (ISL).

1. INTRODUCTION
Sign language [1] is a communication tool for speech or hearing impaired people which transmits the sign patterns visually to express the meaning. The most prevalent sensory impairment in modern people is speech hearing loss. According to world health organization estimates, there are over 63 million significant auditory impairment sufferers in India, putting the estimated prevalence at 6.3% of the population. It is necessary to develop an effective communication system for the impaired people. ISL [2] is predominantly used in South Asian region. India doesn't have a standard sign language like America and Europe. The majority of the ISL parts come from the British sign language. Localized sign languages have been used in India for a very long time. Researchers discovered that different towns' sign languages shared some signs in common as well as some differences. Figure 1 shows the ISL alphabets chart. Automatic translation of gestures [3] into text or speech is most important for interaction between people with disabilities and people who do not understand the sign language. The sign language interpretation is an extensive research area. A technique for identifying the signs in sign language performs automatic conversion of the gestures into human readable text or speech. The dataset is used to train the machine to detect, predict and analyze the target [4]. In order to improve the rate of gesture recognition, vision based technologies [5] can be used efficiently.

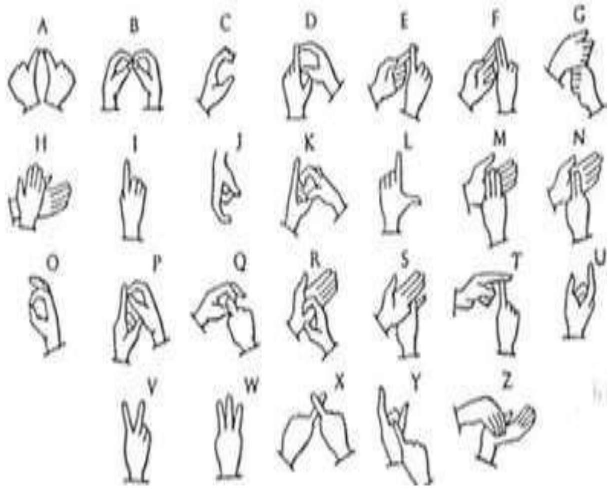

Fig 1: ISL Alphabets

1.1 Sign language recognition
It is a computational process to recognize and to classify the gestures into text or voice using various techniques. It is used in various sectors like human-computer interaction (HCI), General gesture recognition which could be required in specific places like hospitals, banks, railways stations, bus stops etc., Human robot interaction (HRI), sign language tutor for the hearing impaired society, special education for the differently abled people.

1.2 Steps in SLR process
SLR system works on few key factors like hand forms, hand motions, hand and head orientation, hand and head position, and facial expressions. The purpose of SLR system is to interpret the gestures correctly in the form of text or voice. This section explains about the different phases involved in the SLR process. The steps are as follows:

1. Image/Video capturing: It is an action of obtaining an image or video recordings from an external source for further processing. Images or frames will be obtained in this step.
2. Preprocessing: This phase involves processing the raw data and preparing it for the machine learning model. The purpose of preprocessing is to increase the image quality that eliminates the distortions and enhances some features which are important for further processing. It includes some transformations like resizing, orientation and color corrections.
3. Segmentation: It is used for determining the region

of interest (ROI). This procedure is used to recognize objects and boundaries (lines, curves, and so on) in images/frames.

4. Feature Extraction: It is a dimensionality reduction approach. It converts raw data into numerical features that may be processed while keeping the original dataset's content. The images or frames are then recognized and classified using the retrieved features.
5. Classification: It is a technique where the data is analyzed to identify or predict the class it belongs to.
6. Output: The gestures will be converted into either text or speech.

The various steps involved in the SLR process is shown in figure 2.

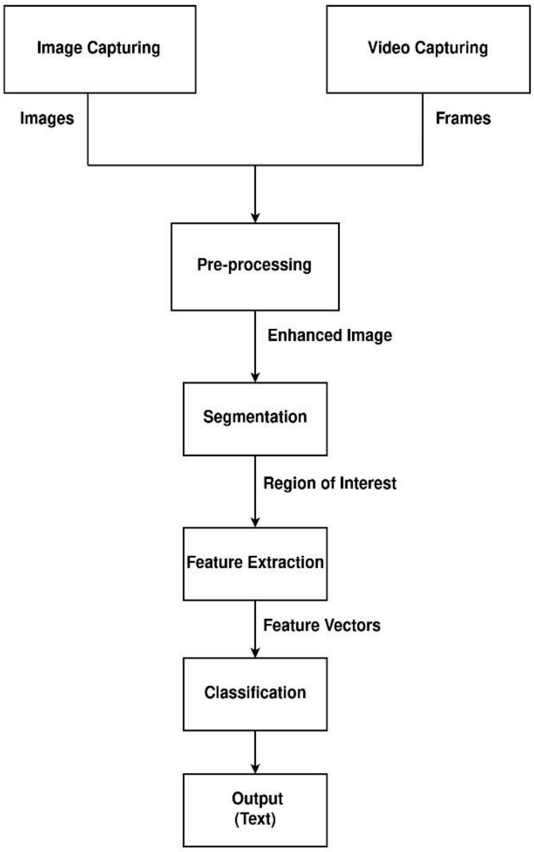

Fig 2: Steps in SLRprocess

1.3 Categories of ISL

There are two main categories of ISL which is represented in figure 3. It is categorized into one-handed and two-handed signs. In one-handed signs, a single hand is used to perform the gestures. Two-handed signs are represented by both the hands. Further, both one-handed and two-handed signs are divided into static and dynamic signs. Static signs are the still images and dynamic signs involves movement or video which can be either isolated or continuous. Isolated signs are formed by one or two words gestures and continuous signs are sentence level. Two-handed dynamic sign is further divided into type 0 and type 1. Both the hands are active in Type 0. One hand is dominant or active in Type 1. Gestures can be manual or non-manual [6]. Manual signs are made up of hand shape, orientation, location and movement. Non-manual signs are facial expressions, mouth gestures, body postures and eye-brow

movement which can be used in combination with hand and shoulder movements.

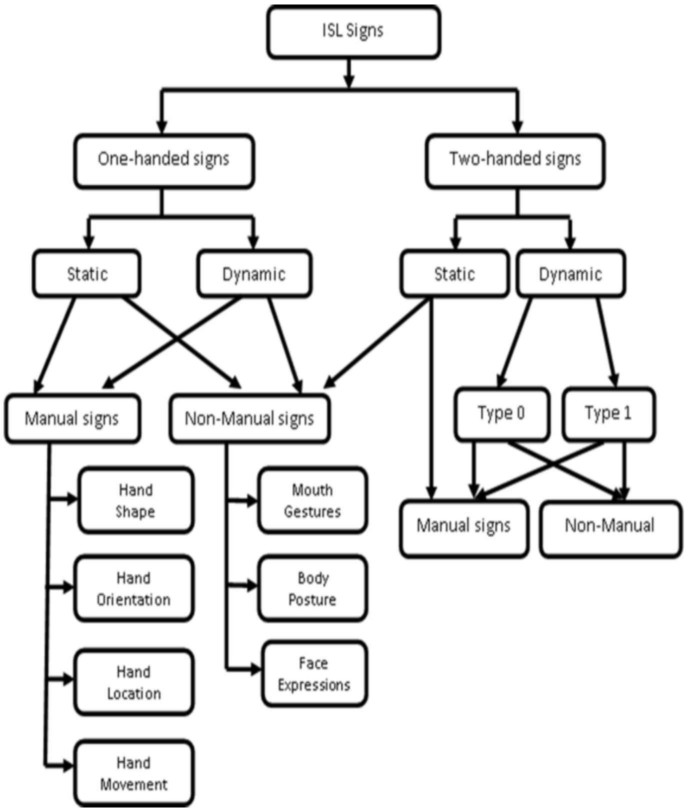

Fig 3: ISL categories

1.4 Features of ISL

- ISL uses both the hands to make most of the gestures and single hand for few particular signs.
- It includes non-manual signs, static and dynamic signs.
- One hand is more active and moves faster in few dynamic gestures.
- Complicated hand shapes, head and body movements are involved in ISL.

1.5 Outline

The goal of this study is to give a thoroughintroduction to image processing, machine learning and deep learning methods that have been used to solve SLR issues within the framework of ISL. It discusses a few techniques used to recognize the ISL. The paper is divided into the following sections:

- Research need and motivation is highlighted in section 2.
- Dataset availability details andthe techniques used in gesture recognition is explained in section 3&4 respectively.
- Section 5 contains the detailed comparative study of the related work.
- Challenges and future directions are explained in section 6 followed by conclusion and references.

2. RESEARCH NEED AND MOTIVATION

People with speech or hearing impairments frequently experience isolation and loneliness because they find it difficult to interact with the regular people. The impact on their social lives is significant. India has only around 300 skilled and qualified sign language interpreters [32].Having human translators available at all times is not practical. Due to these difficulties, a machine translation system that can convert ISL motions into a required pattern is needed, which helps the impaired people in communication. Intellectual processing, social interaction, and literary development are facilitated by the SLR system and also it has the potential to be used in educational sector.

3. DATASET DETAILS

Standard datasets are required for reliable algorithm testing and comparison. Table 1 provides the details of publicly available ISL gesture datasets.

- ISL-CSLTR (Indian sign language dataset for continuous sign language translation and recognition): It consists of sentences and word level gestures performed by 7 different Signers. Total volume of the dataset: 1334 Videos (19899 frames).
  Link to access the dataset: https://data.mendeley.com/datasets/kcmpdxky7p/1
- INCLUDE: It is a large scale dataset for ISL. It has 15 different word categories. Total volume of the dataset: 5484 videos. Link to access the dataset: https://zenodo.org/record/4010759#.YgIATupBxaQ
- ISL KAGGLE: The KAGGLE ISL dataset consists of alphanumeric signs (A-Z & 1-9) which is organized into 35 directories, where each directory is having 1200 files for each sign. Total volume of the dataset: 42000 images. Link to access the dataset: https://www.kaggle.com/prathumarikeri/indian-sign-language-isl
- MENDELEY DATA: It consists of the video dataset which includes words used in emergency situations. This dataset includes few words like (call, doctor, help, hot, lose, pain, thief, accident) which are used commonly to communicate during emergency situations. Total volume of the dataset: 416 words. Link to access the dataset: https://data.mendeley.com/datasets/2vfdm42337/1

Table 1. Publicly available dataset

Repository name and Published Date	**Background type**	**Gesture Type**	**Data Volume**
ISLCSLTR,22/01/2021	Uniform &Complex	Sentence,wordlevel	1334 videos (19899 frames)
INCLUDE 16/10/2020	Uniform & Complex	Word level	5484 videos
ISLKAGGLE	Uniform (Blackbackground)	Character level	42000 images
MENDELEY 27/08/2021	Uniform (Blackbackground)	Word level	416 Words

4. TECHNIQUES USED IN SLR PROCESS

This section gives the detailed information about the various methodologies which are generally used in the SLR process.

1. K-NN (K-Nearest Neighbor): In the paper [7], the authors have used the K-NN algorithm to recognize the hand poses on the clustered data. Detection and tracking is done by using the techniques like face detection, object stabilization and skin color segmentation. Further the image is subjected to grid based feature extraction technique which represents the hand pose in the form of a feature vector. Later K-NN is applied to classify the hand pose which has achieved 99% accuracy. K-NN is also known as the non-parametric method and it is used to solve the problems of classification and regression. It is straight forward to use because it has just two parameters: K value and the distance function (e.g. Euclidean or Manhattan). It does not perform well with the huge dataset and high dimensional values. It is sensitive to noisy data.

2. CNN (Convolutional neural network): In the paper [8], the authors have proposed a robust model using CNN for the feature extraction and classification of ISL gestures. The depth wise separable convolution has been used for SLR. The proposed model automatically detects the key features from an input frame. The model has been designed with appropriate number of convolutional layers and filters to reduce the computational cost. Experiment is done on three different set of self-made data using black as the background color. It has achieved the highest accuracy. CNN is specifically used to process the pixel data. It gives high accuracy in image recognition process. It automatically detects the relevant features. Computational cost is high. Training process takes more time without a GPU for complex tasks. Lot of training data is required.

3. RNN (Recurrent neural network): The authors in [9] have proposed a method to train the temporal features using RNN and spatial features using CNN. Individual frames were predicted using CNN to obtain sequence of predictions. Later the sequence obtained was given to RNN to train the temporal features. The proposed model with a pool layer has achieved a good accuracy of 96% with a self-made dataset which consists of isolated hand gestures. RNN is built to remember every detail of the process, which is useful when dealing with the time series difficulties. The size of the model does not increase with the growth of the input size. The computation process is slow due to its recurrent nature. Long sequences can be difficult to process. It is sensitive to issues such as gradient vanishing.

4. LSTM (Long short term memory): In [10], the authors have applied LSTM combined with GoogleNet which is a pre-trained model used for dynamic gestures. The network has been trained for 20 epochs with the sequence of feature vectors. The network layer is built with sequenceinput layer which is followed by a bidirectional LSTM layer which has 2000 hidden units. This approach withpre-trained model, LSTM, bidirectional LSTM layer, dropout layer and adam optimizer has given 96% accuracy. LSTM is a type of RNN which is capable of learning order dependence in sequence prediction problems. These are used in classification and prediction problems based on the time series data. It offers a wide range of parameters, including learning rates, input and output biases, and so on. As a result, no fine modifications are required. The vanishing gradient problem is almost removed. Longer time is required for the training and more memory is required.

5. MobileNetV2: The authors in [11] have proposed a transfer learning approach which uses a pre-trained model mobile-netv2 on the INSIGNVID which is the first ISL video dataset. 55 most frequently used sentences are used and total 1289 video samples are taken for the experiment which are further augmented to get large number of frames. In the proposed work CNN is initiated with the MobileNetv2 which is pre-trained on ImageNet dataset. The top layer configuration of the model is changed which gives the best classification result which also includes adding of LSTM and dense layer to get the prediction of english sentences. MobileNetV2 is a light weight model which uses the deep neural network and it is considered the best for mobile and embedded vision applications.

6. VPPN (View pair pooling network): The authors in [12] have proposed a deep learning method for multiview SLR. VPPN is used to generate view invariant features and it is called as multi-view feature learning network which learns the pooling process. An 8 stream CNN with motion attention model is proposed to extract features in multiple views. The learned pooling process has the ability to draw view specific features for maximizing recognition accuracy. The VPPN takes the individual view oriented features and adds them into the ensemble feature matrix. Authors have contributed a multiview sign language video dataset with 200 signs in 5 various camera view positions with multiple backgrounds recorded with 5 different signers. The obtained accuracy is 85%.

7. MLP (Multilayer perceptron): The authors in [13] have classified ISL using MLP neural network classifier with 86% accuracy. Dataset consists of one and two handed gestures. The proposed method performs different operations like cropping the original RGB images by using bounding box technique and then converting it to the grayscale images before applying the MLP. The network is tested to search the number of processing elements required in a hidden layer that gives the minimum squared error (MSE) on the training data. Maximum classification accuracy is obtained by varying the transfer functions in hidden layer like tanh, sigmoid, softmax and learning rules like step, momentum, conjugate gradient. According to the authors the letters K,S,T,Y have poor accuracy classification.

5. RELATED WORK

This section highlights the different approaches usedfor the SLR process which uses static, dynamic gestures or real time videos as input. A detailed survey about the different methodologies used on different gestures with different dataset with their accuracy is summarized.

5.1Digital image processing approaches

According to the studies the digital image processing techniques for SLR are considered as the basic algorithms or the traditional algorithms which works well for static gestures in plain background. Algorithms like maximum curvature point selection[14], pixel count algorithm[15], contour and convex hull, harris corner detector16] have used in SLR process. The survey tells that these approaches are not suitable for dynamic gestures and non-manual gestures. Limited work is done on non-uniform background and the accuracy ranges from 51% to 91% with the minimum dataset which ranges from5 imagesto 1300images.

5.2 Machine learning approaches

SLR using machine learning approaches have given better performance compared to the basic algorithms according to the survey. Various machine learning algorithms used by researchers are PCA[17], SVM [18], K-NN [19], K-means clustering. The following observations were done based on the literature review.

- Supervised and unsupervisedalgorithms have been used to recognize thegestures.
- Limiteddatasetsare used fortheexperiment.
- Not suitable for complex gestures and time series data which involves spatial and temporaldata.
- Accuracy ranges from 77% to 98%.

Dataset range:130 images to 42000 images. Majority of the existingsystems have used less number of data and veryfew systems have usedmore trainingdata. Most of the researchers have used the self-created data.

5.3 Deep Learning approaches

Deep learning algorithms are considered as the advanced models with the maximum usage by many researchers in recent years. Detailed comparison about different methodologies with the accuracy obtained is shown in table 2.

Table 2. Deep learningapproaches

Year	Static/ Dynamic	Dataset used	Methods	Remarks& Accuracy
2018 [20]	Dynamic	Own dataset of 18 com-monges-tures.	ANN with backpropa-ga-tion&minimum distan-ceclassifier.	A novel idea of using sign language in the smart phones is simulated and tested. Video capture using selfie stick has been intro-duced. Accuracy: 87%
2019 [21]	Static	A-Z (24,415 images).	CNN	CNN based model named signet is pro-posed to rec-ognize static gestures. Proposed method ex-cludes the letters J & H. Accuracy: 98%
2019 [22]	Both	Own dataset, A-Z & num-bers(45000 RGB & 45000 depth images), word gestures (1080 videos)	CNN for static, LSTM for word ges-tures.	A real time hand gesture recognition system is proposed based on the data captured by the Mi-crosoft Ki-nect RGB-D camera. Accuracy: 99.08%
2020 [23]	Static	A-Z (150000	VGG16 model, natu-	The hierar-chical model

		images)	ral language based output network	performs better than the VGG16 &Deep_CNN but faces problem in recognizing the alphabet J.Accuarcy:98 %
2020 [24]	Static	A-Z, 0-9 & word level signs(35 000 images)	CNN	The proposed approach has used 100 static signs and has been tested on 50 deep learning models with different optimizers. The model could be extended to recognize dynamic signs. Accuracy: 99%
2020 [10]	Dynamic	824 sample videos, ISL words used in emergency situation Source: Mendeley dataset	SVM, CNN,LSTM	Experimented in uniform background with normal lighting conditions. Accuracy: 92%
2020 [25]	Both	Own dataset. 36 static &10 dynamic gestures.	CNN, U-Net, ResNet.	Work done is on static signs and real time gesture recognition system in different lighting conditions is proposed here. Accuracy: 98%
2021 [9]	Dynamic	Self-made ISL dataset,38 Word level gestures.	RNN for temporal data, CNN for spatial data	This work can be extended to recognize continuous gestures. Accuracy: 54%
2021 [26]	Static	50391 images (A-Z,0-9) and seven words.	K-means clustering, BoV, SVM classifiers, CNN & RNN.	The proposed model cannot recognize complex signs. Accuracy: 82%
2021 [27]	Dynamic	Cas-Talk-ISL dataset, 50 ISL words	Hybrid CNN-RNN	Work done is on real-time videos in different backgrounds. Accuracy: 96%
2021 [28]	Dynamic	Benchmark multi-view dataset.	VPPN-multi-view feature learning network, CNN	The Proposed multi stream CNN creates a low dimensional feature vector at multiple stages in the CNN pipeline. View invariance problem is solved. Accuracy: 85%
2021 [29]	Dynamic	Medical/Everydayused Words.	1.CNN using image stacking,2.CNN +LSTM 3.LSTM with OpenPose	Experimented on the dynamic dataset which is of health related, emergency and everyday term words using 3 different models. Accuracy: 83,55, 95% respectively for the three algorithms mentioned.

Table 2 gives the information about the deep learning approaches used for the SLR and its ummarizesthefollowing:

- Commonly used techniques are CNN, RNN, LSTM and the hybrid models.
- Work done is on manual gestures only.
- Accuracy ranges from 54% to 99%
- Many authors have created their own dataset.

Figure 4 represents the accuracy obtained by different techniques in recognizing the gestures which is explained in detail in section 4 of this paper.

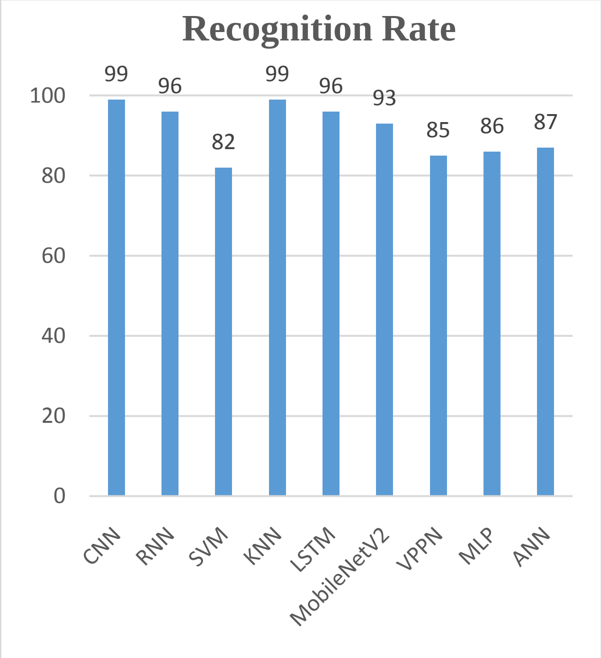

Fig.4 ISL Recognition Rate obtained by various approaches used since 2012

6. CHALLENGES & FUTURE RESEARCH DIRECTIONS

The challenges which may occur in SLR process are explained and the future research directions which could be carried out to overcome the existing drawbacks are explained here.

6.1 Challenges

- Non availability of huge benchmark datasets (For example: Numerals other than 0-9, various isolated and continuous gestures) in complex background and different lighting conditions which are not available in the public dataset repositories mentioned in the Table 1.
- Ability to track the dynamic gestures with various backgrounds and different lighting conditions.
- Gestures vary with signer to signer.
- As ISL uses both the hands, occlusion problem [31] occurs.

6.2 Future research directions

Most of the methods developed so far use uniform backgrounds and uniform lighting conditions, restricted set of gestures, and restricted number of signers. Majority of the work done till now is on static gestures which includes alphabets and numbers only from 0-9. Few researchers have excluded the alphabets H and J as it contains movement of the hand. Very limited work is carried out on sentence and word level gestures. More research should be done on isolated, continuous and non-manual gestures which includes spatial and temporal data with different background and lighting conditions. In order to increase the efficiency of the gestures recognition, more work is to be carried out in recognizing and classifying the time series data. Algorithms like supervised, unsupervised and semi supervised approaches, neural network techniques with better capabilities, having the potential to recognize a wide number of classes should be developed which makes the machine capable to analyze the patterns in a better way. A standard private dataset should be created as there is a problem of non-availability of large ISL dataset.

7. CONCLUSION

This review highlights the different approaches like digital image processing, machine learning and deep learning techniques which are used for the SLR system. It showcases several works and a comparative analysis of those works carried out in recognizing ISL gestures. Most of the authors have used SVM, ANN, K-NN, CNN, LSTM, pre-trained models and the hybrid models which have worked well on static and dynamic gestures of ISL. This paper gives the information about the publicly available datasets. The impaired assistive system helps people with disabilities to communicate with those who are not disabled without isolating them in the society, so an automated system that could translate various ISL gestures into text/speech is required which benefits the impaired people as well as the people with no knowledge about the sign language.

8. REFERENCES

[1] Suharjito, R. Anderson, F. Wiryana, M. C. Ariesta, and G. P. Kusuma, “Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output,” *Procedia Comput. Sci.*, vol. 116, pp. 441–448, 2017, doi: 10.1016/j.procs.2017.10.028.

[2] D. Deora and N. Bajaj, “Indian sign language recognition,” *Proc. 2012 1st Int. Conf. Emerg. Technol. Trends Electron. Commun. Networking, ET2ECN 2012*, 2012, doi: 10.1109/ET2ECN.2012.6470093.

[3] V. Adewale and A. Olamiti, “Conversion of Sign Language To Text And Speech Using Machine Learning Techniques,” *J. Res. Rev. Sci.*, vol. 5, no. 1, 2018, doi: 10.36108/jrrslasu/8102/50(0170).

[4] A. Bhattacharya, V. Zope, K. Kumbhar, P. Borwankar, and A. Mendes, “Classification of Sign Language Gestures using Machine Learning,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 8, no. 12, pp. 97–103, 2020, doi: 10.17148/IJARCCE.2019.81219.

[5] M. Jayaraju, “Indian Sign Language Recognition System-Review,” 2014.

[6] M. Saraswat and K. V Arya, “in an Image Sequence of Non-manual Indian,” pp. 267–275.

[7] K. Shenoy, T. Dastane, V. Rao, and D. Vyavaharkar, “Real-time Indian Sign Language (ISL) Recognition,” *2018 9th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2018*, pp. 1–9, 2018, doi: 10.1109/ICCCNT.2018.8493808.

[8] Sharma, S., & Singh, S. (2022). Recognition of Indian Sign Language (ISL) Using Deep Learning Model. *Wireless Personal Communications*, *123*(1), 671–692. https://doi.org/10.1007/s11277-021-09152-1

[9] A. Divkar, R. Bailkar, and D. C. S. Pawar, “Gesture Based Real-time Indian Sign Language Interpreter,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3307, pp. 387–394, 2021, doi: 10.32628/cseit217374.

[10] V. Adithya and R. Rajesh, “Hand gestures for emergency situations: A video dataset based on words from Indian sign language,” *Data Br.*, vol. 31, p. 106016, 2020, doi: 10.1016/j.dib.2020.106016.

[11] K. Mistree, D. Thako, and B. Bhatt, “Towards Indian

Sign Language Sentence Recognition using Insignvid: Indian Sign Language Video Dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 8, pp. 697–707, 2021, doi: 10.14569/IJACSA.2021.0120881.

[12] M. Suneetha, P. M.V.D., and K. P.V.V., "Multi-view motion modelled deep attention networks (M2DA-Net) for video based sign language recognition," *J. Vis. Commun. Image Represent.*, vol. 78, no. April, p. 103161, 2021, doi: 10.1016/j.jvcir.2021.103161.

[13] R. B. Mapari and G. U. Kharat, "Indian Sign Language Alpha-Numeric Character Classification using Neural Network," vol. 02, no. 06, pp. 1–8, 2016.

[14] M. Geetha and P. V. Aswathi, "Dynamic gesture recognition of Indian sign language considering local motion of hand using spatial location of Key Maximum Curvature Points," *2013 IEEE Recent Adv. Intell. Comput. Syst. RAICS 2013*, pp. 86–91, 2013, doi: 10.1109/RAICS.2013.6745452.

[15] P. R. V. Chowdary, M. N. Babu, T. V. Subbareddy, B. M. Reddy, and V. Elamaran, "Image processing algorithms for gesture recognition using MATLAB," *Proc. 2014 IEEE Int. Conf. Adv. Commun. Control Comput. Technol. ICACCCT 2014*, pp. 1511–1514, Jan. 2015, doi: 10.1109/ICACCCT.2014.7019356.

[16] R. R. Itkarkar, A. Nandi, and B. Mane, "Contour-based real-time hand gesture recognition for Indian sign language," *Adv. Intell. Syst. Comput.*, vol. 556, pp. 683–691, 2017, doi: 10.1007/978-981-10-3874-7_65.

[17] K. Tripathi and N. B. G. C. Nandi, "Continuous Indian Sign Language Gesture Recognition and Sentence Formation," *Procedia Comput. Sci.*, vol. 54, pp. 523–531, 2015, doi: 10.1016/j.procs.2015.06.060.

[18] M. E. Scholar, D. System, and R. Shau, "Available Online at www.ijeecse.com Indian Sign Language Recognition Available Online at www.ijeecse.com," vol. 2, no. 3, pp. 56–59, 2015.

[19] K. M. B, A. Godbole, and S. Belhe, "Using Kinect Sensor," pp. 528–535, 2015, doi: 10.1007/978-3-319-20801-5.

[20] G. A. Rao and P. V. V. Kishore, "Selfie video based continuous Indian sign language recognition system," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 1929–1939, 2018, doi: 10.1016/j.asej.2016.10.013.

[21] C. J. Sruthi and A. Lijiya, "SigNet: A deep learning based indian sign language recognition system," *Proc. 2019 IEEE Int. Conf. Commun. Signal Process. ICCSP 2019*, pp. 596–600, 2019, doi: 10.1109/ICCSP.2019.8698006.

[22] N. K. Bhagat, Y. Vishnusai, and G. N. Rathna, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning," *2019 Digit. Image Comput. Tech. Appl. DICTA 2019*, pp. 1–8, 2019, doi: 10.1109/DICTA47822.2019.8945850.

[23] Sharma, A., Mittal, A., Singh, S., & Awatramani, V. (2020). Hand Gesture Recognition using Image Processing and Feature Extraction Techniques. *Procedia Computer Science*, *173*, 181–190. https://doi.org/10.1016/J.PROCS.2020.06.022

[24] Wadhawan, A., & Kumar, P. (2020). Deep learning-based sign language recognition system for static signs. *Neural Computing and Applications*, *32*(12), 7957–7968.https://doi.org/10.1007/S00521-019-04691-Y.

[25] P. Likhar, N. K. Bhagat, and G. N. Rathna, "Deep Learning Methods for Indian Sign Language Recognition," IEEE Int. Conf. Consum. Electron. - Berlin, ICCE-Berlin, vol. 2020-Novem, pp. 16–21, 2020, doi: 10.1109/ICCE-Berlin50680.2020.9352194.

[26] S. Dhivyasri, K. B. Krishnaa Hari, M. Akash, M. Sona, S. Divyapriya, and V. Krishnaveni, "An efficient approach for interpretation of Indian sign language using machine learning," *2021 3rd Int. Conf. Signal Process. Commun. ICPSC 2021*, no. May, pp. 130–133, 2021, doi: 10.1109/ICSPC51351.2021.9451692.

[27] R. Dhiman, G. Joshi, and C. Rama Krishna, "A deep learning approach for Indian sign language gestures classification with different backgrounds," *J. Phys. Conf. Ser.*, vol. 1950, no. 1, 2021, doi: 10.1088/1742-6596/1950/1/012020.

[28] M. Suneetha, P. M.V.D., and K. P.V.V., "Multi-view motion modelled deep attention networks (M2DA-Net) for video based sign language recognition," *J. Vis. Commun. Image Represent.*, vol. 78, no. April, p. 103161, 2021, doi: 10.1016/j.jvcir.2021.103161.

[29] J. Gandhi, P. Gandhi, A. Gosar, and S. Chaudhari, "Video recognition techniques for indian sign language in healthcare domain," *2021 2nd Int. Conf. Emerg. Technol. INCET 2021*, pp. 1–5, 2021, doi: 10.1109/INCET51464.2021.9456116.

[30] M. Vazquez-Enriquez, J. L. Alba-Castro, L. Docio-Fernandez, and E. Rodriguez-Banga, "Isolated sign language recognition with multi-scale spatial-temporal graph convolutional networks," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 3457–3466, 2021, doi: 10.1109/CVPRW53098.2021.00385.

[31] R. Rastgoo, K. Kiani, and S. Escalera, *Sign Language Recognition: A Deep Survey*, vol. 164. Elsevier Ltd., 2021.

[32] https://islrtc.nic.in/sites/default/files/Interpreter%20Directory.pdf