



A REVIEW ON DEEP LEARNING BASED DYNAMIC INDIAN SIGN LANGUAGE RECOGNITION SYSTEM

Manojkumar Patel¹, Narendra Patel²

¹Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India.

²Birla Vishvakarma Mahavidyalaya, Gujarat, India.

¹<https://orcid.org/0009-0008-5795-1655>^{id}, ²<https://orcid.org/0000-0002-1495-1136>^{id}

Email: manojpatel.25@gmail.com, nmpatel@bvmengineering.ac.in

ARTICLE INFO

Article History

Received: March 13, 2026

Reviewed: April 16, 2026

Accepted: July 16, 2026

Published: August 31, 2026

Keywords:

Indian sign language (ISL),

Vision-based approach,

Classification,

Deep Learning,

Reduced Inequalities.

ABSTRACT

An alternative form of communication for those who are hard of hearing or speech is sign language (SL). SL differs from nation to nation. Indian Sign Language (ISL) is a substitute form of communication used by individuals in India. There are about 63 million deaf and hard of hearing persons living in India. But there are no accessible or reasonably priced human interpreters for ISL identification. Consequently, for many years, researchers have been working on creating a system that can recognize sign language. Compared to other SL, the construction of an Indian Sign Language Recognition System (SLRS) is exceedingly challenging. The field of Indian sign language recognition systems has not seen many developments. Consequently, this discipline has very few systematic literature reviews. We reviewed the most recent (post-2017) literature on dynamic Indian sign language recognition using deep learning technique on various parameters in an attempt to tackle this problem. The objective of this work is to present a comprehensive analysis of various deep learning methods used in dynamic Indian sign language recognition.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

A different form of communication for people who have hearing or speech impairments is sign language (SL). Instead of speaking and using sound patterns, people with hearing impairments communicate by making signs in visual space [1]. In the 1970s, research on sign language linguistics was started. It offers linguistic data, including different letters and symbols. Typically, there are two categories of signs: (1) static and (2) dynamic. The nonverbal language known as SL makes use of body parts to visually transmit patterns. These include non-manual components like lip movement and face expression as well as manual elements like hand and finger positions. Articular points like the joints in the fingers, wrist, elbow, and shoulder are crucial for differentiating SL. It also contains a variety of movements, including forward, backward, left, and right, as well as lip movements and facial expressions like happiness, rage, and confusion. The different sign elements are shown in fig. 1. Correct execution of all sign components is required to generate a precise sign word [1], [2]. Due to its variation among nations and even geographical areas, SL is not a standardized language. As a result, several sign languages (SLs) were created, including Chinese Sign Language (CSL), American Sign Language (ASL) and Indian Sign Language (ISL). [3] Designing a sign language recognition system (SLRS) for ISL is a more difficult task for academics since ISL is distinct from other SLs in that it involves both single handed and double handed gestures, along with a huge vocabulary of related movements. Motion-based Indian sign language referred to as dynamic Indian sign language, as it depicts motion with dynamic signs [4].

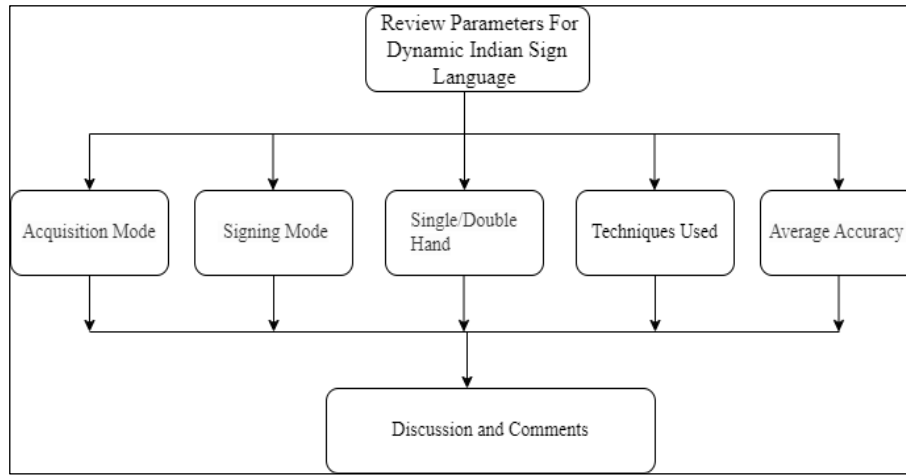


Figure 1: Sign language Elements.
Source: Authors, (2026).

I.1 MOTIVATION

In India, deafness is a persistent issue. The estimated number of deaf individuals residing in India is around 63 million. The deaf and dumb in India utilize ISL, one of the country's living languages. Sign language interpreters can facilitate communication between the regular public and the deaf and dumb. Human interpreters are neither readily available or affordable. It makes it hard to close the huge communication gap between regular people and those who are deaf or dumb. There are lots of work done on static Indian sign language. On Indian dynamic sign language, some research has been done. Therefore, the several deep learning techniques for the identification of Dynamic ISL are the main emphasis of this study.

I.2 LITERATURE SEARCH METHODOLOGY

Finding, categorizing, and contrasting the body of prior research on SLR is the aim of the systematic review. For the literature review, we look at content available as of 2017. A summary of significant studies on Dynamic ISLR methods is given in this survey study. There is discussion on research papers from conferences, workshops, symposiums, journals, and periodicals. The figure displays the inclusion/exclusion criteria after the investigations were reviewed. We conducted a comprehensive search using the Google search engine. The research papers in the specified databases have been searched using the keyword "Dynamic Indian Sign Language Recognition". from the databases of Elsevier, Springer, IEEE, and ACM, various research publications on dynamic Indian sign language recognition were found through our search. Based on the titles, abstracts, and conclusions, the papers were excluded. After reading the entire content, 40 research papers have been selected.

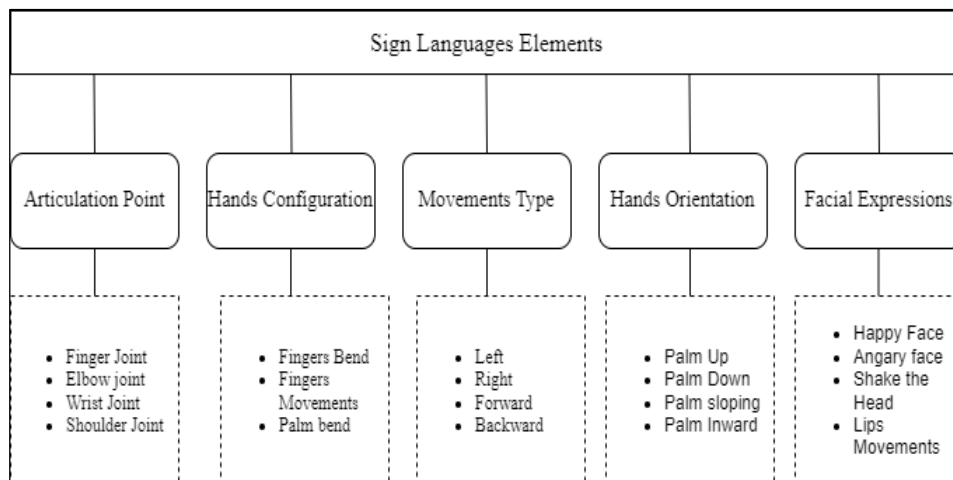


Figure 2: Review Parameters.
Source: Authors, (2026).

A study article may compare things using a different standard. Previous review studies have only examined various approaches and accuracy. On the other hand, in order to assess and comprehend the literature, we have adopted the methodology of [1], [5]. We reviewed the research on dynamic Indian sign language recognition system (DISLR) using those features. Fig. 2 shows the comparative parameter that is used in this review. The parameters of the strategy we have chosen for the Systematic Literature Review are acquisition mode, signing mode, single/double hand signs, techniques used, and average accuracy. Based on these features [1], [5] a review of dynamic Indian sign language recognition (DISLR) has been carried out and recorded.

II. LITERATURE REVIEW ON DISLR - DYNAMIC INDIAN SIGN LANGUAGE RECOGNITION

II.1 GENERAL PROCEDURE OF SIGN LANGUAGE RECOGNITION



Figure 3: SLR General Flow.

Source: Authors, (2026).

A nonverbal communication technique called sign language makes use of the hands, arms, head, and facial expressions, among other body parts, to communicate meaning [6]. SLR is the process of identifying and recognizing signs and translating them into words that have semantic meaning. The steps involved in the SLR include data gathering, pre-processing, extraction of features, and classification. Fig. 3 shows the SLR block diagram along with the overall workflow [5], [7], [8]. Sign language recognition (SLR) is the process of identifying, recognizing, and translating signs into semantically significant words. The SLR system converts the data from signs into comprehensible text or audio by decoding and understanding the data. There are two methods used to perform the function of sign recognition: the hardware based or sensor-based approach and the vision-based approach. In Vision-based method, a camera is used to capture sign images or video. There is various hardware is used to capture sign data in hardware-based approach like data gloves, Microsoft kinect sensor, etc. The signers have difficulty to handle hardware device. So that we focused on vision-based approach in this study [9-14]. After image or video is capture, key frame is identify using various technique and frame is preprocessed. The preprocessed frames or images are processed to find features for classification. The extracted features are given to classifier to classified into various sign. Training and test images are categorized using the retrieved features in the classification stage. It implies that the two methods—image acquisition, extraction of features and classification—are mostly responsible for the rate of gesture recognition. There has been a lot of work on various features extraction and classification techniques in this literature [15]. Over the past few decades, noteworthy developments in accuracy have been made in sign language identification with the use of deep learning, a vision-based technique. Dynamic signs address issues with video input or visual problems, whereas static sign use the model's modality. Moreover, the dynamic signs are the most complex signs [16].

II.2 DISLR APPROACHES

II.2.1 Sensor-Based Deep Learning Approach

By [17] created a technique for recognizing sign language using sensors. Using Kinect and the jump motion device, they gathered a total of 7500 samples of 50 different sign words. The location and orientation of the fingertip characteristics were retrieved using the Leap Motion API. HMM and a bidirectional long short-term memory neural network (BLSTM-NN) were used to classify the system. According to the outcomes, the general accuracy was 95.60% and 84.57% respectively. In turn [18] proposed a modified version of the LSTM, a continuous gesture sequence model for continuous SLR that identifies several related motions. It functions by using neural networks to simulate continuous signals by dissecting them into smaller components. To take pictures, the leap motion sensor is used. Total 942 Indian sign language signed phrases have been used to test the proposed technique. Thirty-five different sign words are used to identify these sign sentences. The average accuracy on isolated sign words and signed phrases was 89.5% and 72.3%, respectively. According to [19] proposed a hand gestures recognition system that operates in real time using information from the Microsoft Kinect RGB-D camera. Computer vision methods such as affine transformation and 3D construction are utilized to provide one to one mapping between the RGB camera and depth sensor pixels. The hand motions were separated from the background noise using segmentation. A total of thirty-six static gestures associated with the alphabets and numbers of Indian Sign Language (ISL) were used to trained convolutional neural networks (CNN). Using training data consisting of 45,000 RGB and 45,000 depth images, the model's accuracy was 98.81%. Ten ISL dynamic word gestures were trained using convolutional LSTMs, and after training of 1080 videos, an accuracy of 99.08% was achieved. By [20] developed a system for the identification of sign sequences in sentences that are constantly signed in Indian sign language with a small amount of labelled sentence data and enough labelled data of isolated signs. Multiple six degree of freedom inertial measurement units (IMUs) are used on the signer's two hands to gather the data. In order to provide end to end phrase identification without the need of sign boundary information, a suggested deep learning model comprises of a convolutional neural network (CNN), two bidirectional long short-term memory (Bi-LSTM) layers, and connectionist temporal classification (CTC).

II.2.2 Vision-Based Deep Learning Approach

In turn [21] developed a vision-based deep learning solution for the video classification problem. They examine the use of Open Pose in conjunction CNN to identify the dynamic sign language videos of twenty medical words in ISL. A standard smartphone camera is used to record all of the videos. CNN was employed for the classification process. At 85% accuracy, it performs excellently. According to [22] developed a SLRS which U-Net for Semantic segmentation of hands. ResNet 101 was used with U-Net as the backbone. The use of an RGB-D Kinect camera is entirely unnecessary since semantic segmentation makes use of information from a regular RGB camera. Using training data, the system demonstrated strong performance in multiclass semantic segmentation, producing a F1 score of 0.9957 and an IOU score of 0.9920. By [10] developed the Indian Lexicon Sign Language Dataset (INCLUDE) is a ISL dataset that contains 0.27 million frames from 4,287 videos capture for 263-word signs from 15 distinct word categories. To define INCLUDE-50, a subset of 50 sign word from various word categories is selected for quick assessment of SLR methods with various hyperparameter adjustment. The authors assess a number of deep neural networks by integrating various feature extraction techniques, augmentation techniques, encoding and decoding.

On the INCLUDE-50 and the INCLUDE dataset, the top-performing model obtains accuracy rates of 94.5% and 85.6%, respectively. In turn [11] suggested a system based on a vision-based approach for interpreting signs in Indian sign language to words in English language translation. The Resnet50 model is used to find out features from RGB images, and the LSTM is used to analyze their sequences. The dataset contains 7 isolated sign words for days. The system achieved 73.60% accuracy. According to [13] presented a system that used a 3D CNN model to recognize ISL words. The model is trained on a dataset that contains 20 sign words. The system obtained 88.35% accuracy. By [14] developed a system that recognizes sign language words. The system is based on the CNN-RNN architecture. The Inceptionv3 is used to extract features from frames, and after that, the LSTM model is used to analyze sequences of frames. The proposed system is trained on the CasTalk ISL dataset, which contains 50 sign words. During testing with the test dataset, the system achieved 99.46% top-3 accuracy and 95.99% top-1 accuracy. In turn [23] proposed a model for hand gestures detection for ISL that deaf farmers frequently use. The spatial attributes are determined using the Googlenet Model. The LSTM receives the spatial characteristics in order to determine the temporal features. Using the suggested dataset of eight ISL terms from the agriculture field. An average classification accuracy of 76.21% has been attained by the model. According to [24] suggested INSIGNVID, the first Indian Sign Language video dataset. Using transfer learning, a unique method is described that uses this dataset as input to convert the ISL sentences in the video into proper English sentences. The mobilenetv2 is used to find out spatial features. After that, a sequence of spatial features is given to LSTM to extract temporal features and classify them. The system has achieved 91% accuracy during training. By [25] proposed a DL model that uses a person's gestures to identify and detect words. ISL video frames are divided into isolated segments, and signs are identified using the LSTM and GRU model. With the IISL2020 dataset, four distinct sequential combinations of LSTM and GRU—since there are two layers of LSTM and two levels of GRU—were applied. With single level of LSTM and GRU layer, the suggested model achieved approximately 97% accuracy across 11 distinct signs.

In turn [26] proposed the MOPGRU (integrated Media Pipe optimized gated recurrent unit) model for the recognition of ISL. The authors have enhanced the conventional GRU cell's update gate by multiplying it by the reset gate in order to eliminate redundant data from previous screenings in a single screening. By getting feedback from the consequent of the reset gate, more attention is paid to the current input. Furthermore, In the output layer of the GRU cell, use SoftMax activation instead of Softsign activation and exponential linear unit activation in place of the hyperbolic tangent activation seen in conventional GRUs. The model is trained on a dataset contains 13 daily sign words. The system achieved 95% accuracy. According to [27] presented a DL model for motion-based recognition of ISL. OpenCV is used to record the gesture, MediaPipe is used to identify the important points, and an LSTM model that has been trained to predicts the sign. There are forty sign words in the dataset. 92% accuracy on average was attained. In turn [28] presented two approaches to solve the sign language translation problem: BERT Transformer and Bidirectional LSTM. The suggested algorithm's performance has been examined using the INCLUDE-50 dataset. A variety of data augmentation techniques and mediapipe based feature extraction is used to evaluate the deep neural network. On the Dataset INCLUDE 50, the top-performing model obtained an accuracy of 89.5%. This model makes use of an encoder, a decoder, and a pre-trained feature extractor. By [29] suggested a method for ISL word and character recognition. With a 99 percent accuracy rate, the Bag of Visual Words (BoVW) approach can identify ISL character (0-9, A-Z). ISL recognition at the word level is done using the INCLUDE-50 dataset. The spatial and temporal features of the video are trained using the Inception V3 model and LSTM, respectively. With CNN predictions as the RNN's input, authors were able to attain 86.7% accuracy. In turn [30] suggested the Hybrid CNN-BiLSTM SLR (HCBSLR) model, a vision-based SLRS that addresses the issue of over-preprocessing. The suggested model eliminates unnecessary or redundant frames using a key frame extraction technique based on Histogram Difference (HD), which increases the system's accuracy and efficiency. To extract spatial features, the HCBSLR system uses VGG-19, and to extract temporal features, it uses BiLSTM. The accuracy of the suggested HCBSLR system is 87.67% on average. According to [31] proposed the Automated ISLRS for Emergency Words (AISLRSEW), a SLRS designed to identify ISL words commonly used in emergency situations.

The problem of identifying ISL words with same hand position and multiple viewing angle is addressed by the proposed AISLRSEW, which enhances identification accuracy by combining CNN with locally created features. BiLSTM is used to extract temporal features. On a dataset with eight sign-words, the suggested model had an average accuracy of 94.42%. By [32] developed a LiST (Lightweight Sign language Translation) system, a lightweight SLT framework that takes into account several modalities at once, including hand gestures, hands orientation, and facial expression, from an ISL video. A two-layered LSTM architecture processes the feature map created by the InceptionV3 handling the features linked to various signer modalities. Continuous ISL statements from the INCLUDE dataset were used to test the model. The experimental findings demonstrate that the LiST framework attained a high prediction accuracy of 95.9% and translation accuracy of 91.2%. In turn [33] proposed a vision-based system for ISL hand gesture recognition. Transfer learning is employed in conjunction with two distinct CNN models—Google Net and VGG16—to extract spatial characteristics. These CNN models are combined with the LSTM model to determine temporal features. Three distinct datasets of videos with 11, 64 and 8 classes are used to assess the system. On the three datasets, CNN-LSTM architecture with Google Net achieves remarkably recognition accuracy of 93.18%, 97.50% and 96.65%, respectively. According to [34] suggested a study on automatic ISL identification that could help deaf patients and Indian healthcare professionals communicate more effectively. A new dataset of dynamic hand gestures for ISL terms that deaf COVID-19 positive patients frequently employ for emergency communication is suggested. The suggested hand movements were recognized using a hybrid deep convolutional LSTM network model, which produced an average accuracy of 83.36%. In turn [35] has suggested a deep learning-based model that consists of trainable CNN and trainable stacked 2 bidirectional long short-term memory (S2B-LSTM) to recognize the dynamic gestures of ISL. The spatial features in the input video data have been extracted using the CNN architecture as a feature extractor. S2B-LSTM is used to extract the temporal relationship between the successive frames of video. Training and testing are conducted on dataset, which consists of 360 videos of ISL dynamic movements. The recognition accuracy of the CNN-S2B-LSTM model is 97.6%. To identify ISL terms for emergency scenarios, a brand-new spatial attention (SA) based SLRS called the Spatial Attention based Sign Language Recognition Module (SASLRM) is proposed by [36]. ISL word recognition is achieved by SASLRM through a combination of attention features from a SA module and convolution features from a pretrained VGG-19 model. BiLSTM is used to extract temporal features.

On the ISL dataset, which contains eight emergency sign words, the suggested model achieved an average accuracy of 95.627%. By [37] proposed a system for automatic sign language identification that could help deaf patients and Indian healthcare professionals communicate more effectively. A new dataset of dynamic hand gestures for ISL of six medical terms that deaf patients frequently employ for emergency communication is developed. The hand movements were recognized using a hybrid deep convolutional LSTM network model, which produced a training accuracy of 100%. In turn, [38] suggested to use the MediaPipe holistic pipeline for feature extraction and LSTM network to recognize isolated words of ISL. INCLUDE and INCLUDE-50 are used to test the suggested approach. On INCLUDE-50 and INCLUDE, the accuracy achieved by the proposed method is 94.8% and 87.4%, respectively. Multi Featured Deep Network (MF-DNet), an occlusion tolerant SLRS, is suggested for identifying ISL words by [39]. The custom dataset of 8 emergency word is used. To eliminate unnecessary frames, the proposed MF-DNet employs a histogram difference based keyframe selection method. Pose features are combined with CNN features in the proposed MF-DNet to address the issues of occlusion, identical hand gestures, and multiple viewing angles. The suggested method makes use of a BiLSTM network for classification. On the ISL dataset, the average classification accuracy for the proposed SLRS was 96.88%. By [40] suggested a novel method for ISL recognition that integrates Visual Transformers (ViT), Keras, and advanced data augmentation approaches. The approach focuses on creating a Vision Transformer model that has been trained on a large include dataset and supplemented with mediapipe's keypoint information and image data. The validation approach yielded remarkable accuracy rate of 97.52% and a promising assessment loss of 0.2941. Summary of literature review of sensor based DISLR approach in tabular form is illustrated in Table 1. Summary of literature review of vision based DISLR approach in tabular form is illustrated in Table 2.

Table 1: Summary of review of Sensor based approach.

Author	Acquisition	Dataset/ No. of signs	Single/ double handed	Signing mode	Spatial Feature extraction method	Temporal Feature extraction method	Accuracy (%)
Kumar et al. [17]	Kinect and leap motion sensor	50 words	Both	Isolated	-	HMM and BLSTM-NN	95.60% for all signs
Mittal et al. [18]	Leap motion sensor	35 words	Single	Both	CNN	Modified LSTM	72.3%-signed sentence 89.5%-isolated sign
Bhagat et al., [19]	Microsoft Kinect camera	36 static sign character and 10 dynamic sign words	Both	Isolated	CNN	LSTM	99.08%
Sharma et al. [20]	inertial measurement units (IMUs)	Isolated sign-50 words, Continuous signed sentences-40	Both	Both	CNN	BILSTM, CTC	90.00%

Source: Authors, (2026).

Table 2: Summary of review of Vision based approach.

Author	Acquisition	Dataset/ No. of signs	Single/ double handed	Signing mode	Spatial Feature extraction method	Temporal Feature extraction method	Accuracy (%)
Uchil aditya P. and Jha, [21]	Smartphone	19 dynamic and one static words	Single	Isolated	Open Pose	CN	85
Jayadeep et al. [9]	camera	20 words	Both	Isolated	Inceptionv3	LSTM	100
Likhar et al. [22]	camera	10 words	Both	Isolated	Resnet 101 + U-Net	LSTM	99.08
Sridhar et al. [10]	camera	Include, Incude-50 words	Both	Isolated	MobileNetv2	BILSTM	85.6, 94.5
Mistry et al. [11]	camera	7 words	Double	Isolated	Restnet50	LSTM	99.9
Singh [13]	camera	20 words	Both	Isolated	-	3D CNN	83.30
MuthuMari appan & Gomathi [14]	camera	50 words	Both	Isolated	Inceptionv3	LSTM	top-1 :- 95.9, top-3:- 99.46
Venugopalan & Reghunadhan [23]	camera	13 words	Both	Isolated	GoogleNet	LSTM	76.21

Mistree et al. [24]	camera	55 ISL Sentences	Both	Continuous	MobileNetv2	LSTM	91.00
Kothadiya et al. [25]	camera	11 words	Both	Isolated	InceptionResNetV2	LSTM and GRU	97
Subramanian et al. [26]	camera	13 words	Both	Isolated	Mediapipe	GRU	95
Ganpatye & Mane [27]	camera	40 words	Both	Isolated	Mediapipe	LSTM	92
Prathap et al. [28]	camera	Include 50 words	Both	Isolated	Mediapipe	BILSTM, BERT TRANSFORMER	89.50
Katti et al. [29]	camera	Include 50 words	Both	Isolated	Inceptionv3	LSTM	86.70
S. Das et al. [30]	camera	20 words	Both	Isolated	VGG16	BILSTM	87.67
S. Das et al. [31]	camera	8 words	Both	Isolated	SIFT and BOF, VGG19	BILSTM	94.42
K et al. [32]	camera	words	Both	Isolated	Inception V3	LSTM	95.9
Pradip et al. [33]	camera	11, 64 and 8 words	Both	Isolated	GoogleNet	LSTM	93.18, 97.50, and 96.65
Venugopalan Reghunadhan [34]	camera	17 words	Both	Isolated	VGG16	BILSTM	83.36
Sharma & Singh [35]	camera	18 words	Both	Isolated	CNN	Stacked BiLSTM	97.6
S. Das et al. [36]	camera	8 words	Both	Isolated	VGG19+CSA(Con. Spatial attention)	BILSTM	95.62
H. V Das et al. [37]	camera	6 words	Both	Isolated	CNN	LSTM	100
Khartheesvar et al. [38]	camera	Include ,Include 50 words	Both	Isolated	Mediapipe	LSTM	87.4 ,94.8
S. Das et al. [39]	camera	8 words	Both	Isolated	(VGG19+Mediapipe)	BILSTM	96.8
Singla et al. [40]	camera	Include dataset	Both	Isolated	Mediapipe	Transformer	97.5

Source: Authors, (2026).

III. DISCUSSION

Below is a report on the dynamic ISL recognition system outcomes for our review with respect to research topics. Cameras were used in 83% of the study on dynamic ISL, as can be seen in Fig. 4 for ISL data acquisition. 3% of the study used Microsoft Kinect, 4% used leap motion sensors, 4% used a smartphone, 3% used inertial measurement units, 3% used kinect and leap motion-based sensor.

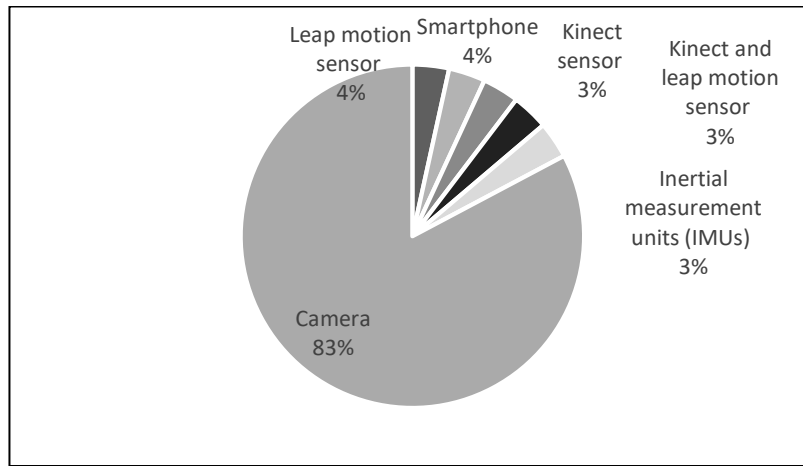


Figure 4: Data Acquisition Summary.
Source: Authors, (2026).

Fig. 5 illustrates that Dynamic ISL work has been performed 7% of the time on single-handed signs, 90% on both-handed signs, and 3% on double-handed signs.

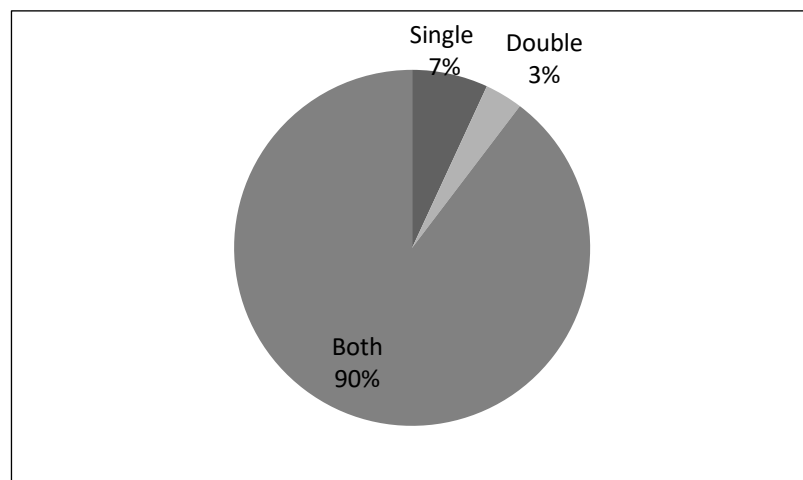


Figure 5: Hand Used Summary.
Source: Authors, (2026).

As illustrated in Fig. 6, isolated sign mode accounts for 90% of the signing modalities used for dynamic ISL research activities, while continuous sign mode accounts for 3% and both for 7%. LSTM has been used in 48% of the work on dynamic ISL as shown in Fig. 7, followed by BILSTM in 31%, CNN in 7%, GRU in 4%, HMM and BILSTM in 4%, LSTM and GRU in 3%, Transformer in 3%.

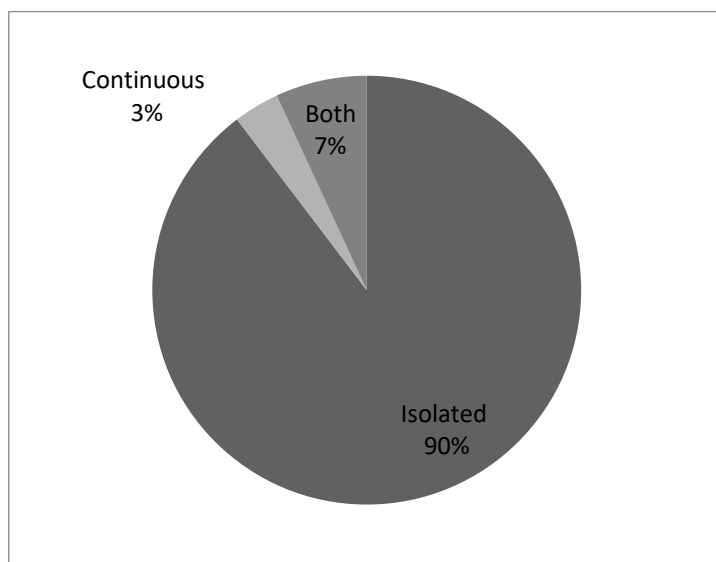


Figure 6: Signing Mode Summary.
Source: Authors, (2026).

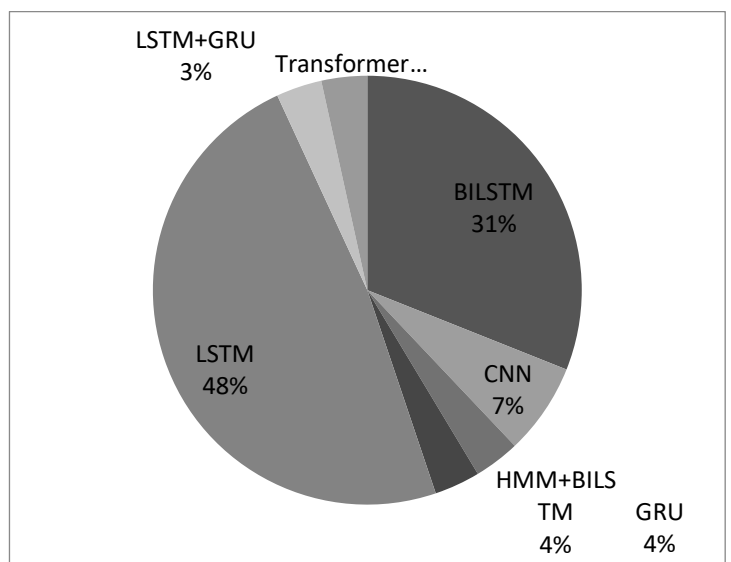


Figure 7: Deep learning Technique Summary.
Source: Authors, (2026).

According to Fig. 8, 69% of sign language recognition systems for dynamic ISL were able to reach an average accuracy of greater than 90%, whilst 24% of systems were able to achieve an accuracy of between 80 and 89%. Less than 7% of systems have an accuracy below 80%. There are two primary types of Dynamic Indian sign language recognition: continuous sign language recognition and isolated sign language recognition. One significant distinction between the two types is the supervisory information. Continuous ISL recognition addresses both the recognition challenge and the precise alignment between the video segments and the associated sentence level labels, whereas isolated ISL recognition is comparable to the action recognition field.

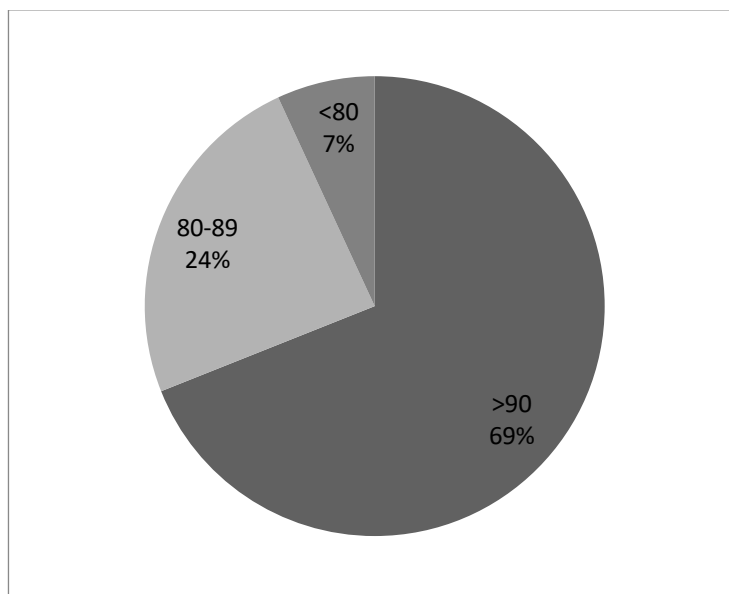


Figure 8: Accuracy Summary.

Source: Authors, (2026).

Recognition of continuous ISL is generally harder than recognition of isolated ISL. It is possible to think of isolated ISL recognition as a subset of continuous ISL recognition. A continuous ISL recognition system may perform better if it extracts more discriminative and descriptive characteristics from the video frames. Continuous dynamic signs are utilized at the sentence level, whereas isolated dynamic signs are used at the word level. There is still much space for enhancing performance in this field, even though current models in isolated and continuous sign language recognition show an increasing trend in model performance depending on DL skills in computer vision. Some potential future directions in this field include taking into account the attention mechanism, learning structured spatio-temporal patterns (like Graph Neural Networks models), utilizing multiple input modalities to benefit from multi-channel information, and applying prior knowledge on sign language. Sign Language recognition is more than just word, number, and alphabet recognition. The SLRS must recognize signs and create meaningful sentences from them in real time. To do this, it must recognize sign words and determine the context of the words. While SL recognition is a difficult task in and of itself, ISL introduces additional difficulties that make it harder for researchers to create an SLRS. The identification of a language's grammar rules is essential for language recognition; however, the grammar rules for ISL are not well established, making the process of creating an SLRS for ISL recognition difficult for academics. Another issue with ISL recognition is the lack of a consistent dataset with a broad vocabulary, since building an SLRS requires sufficient and reliable data. ISL alphabets and words use both single- and double-handed motions, unlike other SL. Signers occasionally overlap their hands when executing double-handed motions, which increases the difficulty of the segmentation and recognition challenge compared to other sign languages.

IV. CONCLUSION AND FUTURE WORK

Vision Based ISL enables people who are deaf or mute to communicate their thoughts and experiences. There is a growing trend in the usage of ISL recognition software for people. Using the keywords "Indian Sign Language Recognition," "Dynamic Indian Sign Language Recognition," and "Dynamic Indian Sign Language Recognition using Deep learning," only research publications published after 2017 were included in this analysis. Only online databases were searched for research papers for this investigation. There were no large-scale systems implemented that could process a broad vocabulary of signs in real time. In order to provide a more meaningful understanding of the technique, our literature aims to provide an overview of the research based on five parameters such as: accuracy rate, single/double-handed signs, data collection methodologies, signing style, and classification method. While this review is not intended to be exhaustive, it does offer a fair amount of information and illustrates the prevalence of research on the topic. The most widely used data acquisition tools, according to the statistics, were cameras (83%) and Kinect Senor (3%). A large portion of research on dynamic Indian sign language recognition systems has been conducted on isolated signs 90% and 90% of the studies done both single and double handed signs, with 69% of the systems having an accuracy rate of greater than 90%. It has been found that there are different sign language systems; LSTM has been used to complete 48% of the work. For large-scale systems, there is a higher effort requirement to identify dynamic Indian sign language. Research on deep learning in dynamic Indian sign language identification systems is limited. The majority of research in the field of dynamic Indian sign language recognition has been done on isolated signs. It is therefore important to use deep learning neural networks for the recognition of continuous dynamic Indian sign language.

V. AUTHOR'S CONTRIBUTION

Conceptualization: Manojkumar Patel.

Methodology: Manojkumar Patel, Narendra Patel.

Investigation: Manojkumar Patel.

Discussion of results: Manojkumar Patel.

Writing – Original Draft: Manojkumar Patel.

Writing – Review and Editing: Manojkumar Patel, Narendra Patel.

Resources: Manojkumar Patel.

Supervision: Narendra Patel.

Approval of the final text: Manojkumar Patel, Narendra Patel.

VI. REFERENCES

- [1] A. Wadhawan and P. Kumar, "Sign language recognition systems: a decade systematic literature review," *Arch Comput Methods Eng*, vol. 28, 2021, doi: 10.1007/s11831-019-09384-2.
- [2] R. Rastgoo, K. Kiani, and S. Escalera, "Word separation in continuous sign language using isolated signs and post-processing," *Jul*. 2022.
- [3] S. Kr. and C. M. and P. B. Das Soumen and Biswas, "Intelligent Indian Sign Language Recognition Systems: A Critical Review," in *ICT Systems and Sustainability*, S. and J. A. Tuba Milan and Akashe, Ed., Singapore: Springer Nature Singapore, 2022, pp. 703–713.
- [4] T. Raghuveera, R. Deepthi, R. Mangalashri, and R. Akshaya, "A depth-based Indian sign language recognition using microsoft kinect," *Sādhanā*, vol. 45, 2020, doi: 10.1007/s12046-019-1250-6.
- [5] J. M. Patel Dhaval U. and Joshi, "Review of Indian Dynamic Sign Language Recognition System," in *Proceedings of International Conference on Communication and Computational Technologies*, S. and P. S. D. and P. M. Kumar Sandeep and Hiranwal, Ed., Singapore: Springer Nature Singapore, 2023, pp. 17–31.
- [6] D. H. Neiva and C. Zanchettin, "Gesture recognition: a review focusing on sign language in a mobile context," *Expert Syst Appl*, vol. 103, 2018, doi: 10.1016/j.eswa.2018.01.051.
- [7] M. Manoharan and P. Roy, "A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets," *Jul*. 2022.
- [8] S. Das, S. kr. Biswas, M. Chakraborty, and B. Purkayastha, "A Review on Sign Language Recognition (SLR) System: ML and DL for SLR," in *2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISST)*, 2021, pp. 177–182. doi: 10.1109/ICISST52025.2021.00045.
- [9] G. Jayadeep, V. V. V. Venugopal, V. S. and G. M., "Mudra: Convolutional Neural Network based Indian Sign Language Translator for Banks," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1228–1232, 2020, [Online]. Available: <https://api.semanticscholar.org/CorpusID:219990092>
- [10] A. Sridhar, R. G. Ganesan, P. Kumar, and M. Khapra, "INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition," in *Proceedings of the 28th ACM International Conference on Multimedia*, in *MM '20*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1366–1375. doi: 10.1145/3394171.3413528.
- [11] P. Mistry, V. Jotaniya, P. Patel, N. Patel, and M. Hasan, "Indian Sign Language Recognition using Deep Learning," in *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, 2021, pp. 1–6. doi: 10.1109/AIMV53313.2021.9670933.
- [12] M. Varsha and C. S. Nair, "Indian Sign Language Gesture Recognition Using Deep Convolutional Neural Network," in *2021 8th International Conference on Smart Computing and Communications (ICSCC)*, 2021, pp. 193–197. doi: 10.1109/ICSCC51209.2021.9528246.
- [13] D. K. Singh, "3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling," *Procedia Comput. Sci.*, vol. 189, pp. 76–83, 2021, doi: <https://doi.org/10.1016/j.procs.2021.05.071>.
- [14] H. MuthuMariappan and Gomathi, "Indian Sign Language Recognition through Hybrid ConvNet-LSTM Networks," *EMITTER International Journal of Engineering Technology*, 2021, [Online]. Available: <https://api.semanticscholar.org/CorpusID:237383878>
- [15] A. Tyagi and S. Bansal, "Feature Extraction Technique for Vision-Based Indian Sign Language Recognition System: A Review," 2021, pp. 39–53. doi: 10.1007/978-981-15-6876-3_4.
- [16] R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition: A Deep Survey," *Expert Syst. Appl.*, vol. 164, p. 113794, 2021, doi: <https://doi.org/10.1016/j.eswa.2020.113794>.
- [17] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "A multimodal framework for sensor based sign language recognition," *Neurocomputing*, vol. 259, 2017, doi: 10.1016/j.neucom.2016.08.132.
- [18] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion," *IEEE Sens. J.*, vol. 19, no. 16, pp. 7056–7063, 2019, doi: 10.1109/JSEN.2019.2909837.
- [19] N. K. Bhagat, Y. Vishnusai, and G. N. Rathna, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning," in *2019 Digital Image Computing: Techniques and Applications (DICTA)*, 2019, pp. 1–8. doi: 10.1109/DICTA47822.2019.8945850.
- [20] S. Sharma, R. Gupta, and A. Kumar, "Continuous sign language recognition using isolated signs data and deep transfer learning," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 3, pp. 1531–1542, 2023, doi: 10.1007/s12652-021-03418-z.
- [21] S. and S. B. G. Uchil Aditya P. and Jha, "Vision Based Deep Learning Approach for Dynamic Indian Sign Language Recognition in Healthcare," in *Computational Vision and Bio-Inspired Computing*, J. M. R. S. and B. V. E. and I. A. M. Smys S. and Tavares, Ed., Cham: Springer International Publishing, 2020, pp. 371–383.
- [22] P. Likhari, N. K. Bhagat, and R. G. N., "Deep Learning Methods for Indian Sign Language Recognition," in *2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin)*, 2020, pp. 1–6. doi: 10.1109/ICCE-Berlin50680.2020.9352194.
- [23] A. Venugopalan and R. Reghunadhan, "Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists," *Expert Syst. Appl.*, vol. 185, p. 115601, 2021, doi: <https://doi.org/10.1016/j.eswa.2021.115601>.

- [24] K. Mistree, D. Thakor, and B. Bhatt, "Towards Indian Sign Language Sentence Recognition using INSIGNVID: Indian Sign Language Video Dataset," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 8, 2021, doi: 10.14569/IJACSA.2021.0120881.
- [25] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, "Deepsign: Sign Language Detection and Recognition Using Deep Learning," *Electronics (Basel)*, vol. 11, no. 11, 2022, doi: 10.3390/electronics11111780.
- [26] B. Subramanian, B. Olimov, S. M. Naik, S. Kim, K.-H. Park, and J. Kim, "An integrated mediapipe-optimized GRU model for Indian sign language recognition," *Sci. Rep.*, vol. 12, no. 1, p. 11964, 2022, doi: 10.1038/s41598-022-15998-7.
- [27] A. Ganpatye and S. Mane, "Motion Based Indian Sign Language Recognition using Deep Learning," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, 2022, pp. 1–6. doi: 10.1109/CONIT55038.2022.9848275.
- [28] K. B. Prathap, G. D. Swaroop, B. P. Kumar, V. Kamble, and M. Parate, "ISLR: Indian Sign Language Recognition," in *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, 2023, pp. 1–6. doi: 10.1109/PCEMS58491.2023.10136062.
- [29] R. K. Katti, S. C. P. Desai, and S. G., "Character and Word Level Gesture Recognition of Indian Sign Language," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, 2023, pp. 1–6. doi: 10.1109/I2CT57861.2023.10126314.
- [30] S. Das, S. Kr. Biswas, and B. Purkayastha, "A deep sign language recognition system for Indian sign language," *Neural Comput. Appl.*, vol. 35, no. 2, pp. 1469–1481, 2023, doi: 10.1007/s00521-022-07840-y.
- [31] S. Das, S. K. Biswas, and B. Purkayastha, "Automated Indian sign language recognition system by fusing deep and handcrafted feature," *Multimed. Tools Appl.*, vol. 82, no. 11, pp. 16905–16927, 2023, doi: 10.1007/s11042-022-14084-4.
- [32] A. K. P. P. and R. C. Poonia, "LiST: A Lightweight Framework for Continuous Indian Sign Language Translation," *Information*, vol. 14, no. 2, 2023, doi: 10.3390/info14020079.
- [33] P. Patel and Narendra Patel, "Dynamic Hand Gesture Recognition for Indian Sign Language using Integrated CNN-LSTM Architecture," *International Journal of Next-Generation Computing*, Nov. 2023, doi: 10.47164/ijngc.v14i4.1039.
- [34] A. Venugopalan and R. Reghunadhan, "Applying Hybrid Deep Neural Network for the Recognition of Sign Language Words Used by the Deaf COVID-19 Patients," *Arab. J. Sci. Eng.*, vol. 48, no. 2, pp. 1349–1362, 2023, doi: 10.1007/s13369-022-06843-0.
- [35] S. Sharma and S. Singh, "A Spatio-Temporal Framework for Dynamic Indian Sign Language Recognition," *Wirel. Pers. Commun.*, vol. 132, no. 4, pp. 2527–2541, Oct. 2023, doi: 10.1007/s11277-023-10730-8.
- [36] S. Das, S. Kr. Biswas, and B. Purkayastha, "An Expert System for Indian Sign Language Recognition Using Spatial Attention-based Feature and Temporal Feature," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 23, no. 3, Mar. 2024, doi: 10.1145/3643824.
- [37] H. V. Das, K. Mohan, L. Paul, S. Kumaresan, and C. S. Nair, "Transforming consulting atmosphere with Indian sign language translation," *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 13543–13555, 2024, doi: 10.1007/s11042-023-15214-2.
- [38] G. Khartheesvar, M. Kumar, A. K. Yadav, and D. Yadav, "Automatic Indian sign language recognition using MediaPipe holistic and LSTM network," *Multimed. Tools Appl.*, vol. 83, no. 20, pp. 58329–58348, 2024, doi: 10.1007/s11042-023-17361-y.
- [39] S. Das, S. Kr. Biswas, and B. Purkayastha, "Occlusion robust sign language recognition system for indian sign language using CNN and pose features," *Multimed. Tools Appl.*, Apr. 2024, doi: 10.1007/s11042-024-19068-0.
- [40] V. Singla, S. Bawa, and J. Singh, "Enhancing Indian sign language recognition through data augmentation and visual transformer," *Neural Comput. Appl.*, vol. 36, no. 24, pp. 15103–15116, Aug. 2024, doi: 10.1007/s00521-024-09845-1.