Received: 11 August 2024 | Revised: 22 November 2024 | Accepted: 10 January 2025

IET Image Processing

DOI: 10.1049/ipr2.70000

The Institution of Engineering and Technology WILEY

REVIEW

An integrative survey on Indian sign language recognition and translation

Rina Damdoo[1] | **Praveen Kumar**[2]

[1]Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Ramdeobaba University, Nagpur, India

[2]Department of Computer Science and Engineering, Visvesvaraya National Institute of Technology, Nagpur, India

Correspondence
Rina Damdoo, Shri Ramdeobaba College of Engineering and Management, Ramdeobaba University, Nagpur, 440013, India.
Email: damdoor@rknec.edu

Abstract

Hard of hearing (HoH) people commonly use sign languages (SLs) to communicate. They face major impediments in communicating with hearing individuals, mostly because hearing people are unaware of SLs. Therefore, it is important to promote tools that enable communication between users of sign language and users of spoken languages. The study of sign language recognition and translation (SLRT) is a step forward in this direction, as it tries to create a spoken-language translation of a sign-language video or vice versa. This study aims to survey the Indian sign language (ISL) interpretation literature and gives pertinent information about ISL recognition and translation (ISLRT). It provides an overview of recent advances in ISLRT, including the use of machine learning based, deep learning based, and gesture-based techniques. This work also summarizes the development of ISL datasets and dictionaries. It highlights the gaps in the literature and provides recommendations for future research opportunities for ISLRT development.

1 | INTRODUCTION

Sign language (SL) is a visual language used by millions of deaf, hard-of-hearing, and speech disability individuals. According to the World Federation of the Deaf (WFD), around 70 million people use SL. Due to audism, hard of hearing (HoH) individuals are often ignored, and compelled to use alternate communication means. Consequences of this are, to present signs HoH people need to write down or type message, or are forced to use special gloves [1], with which HoH individuals are disquieted. Secondly, SL is an effective technique to address the problems faced by HoH people and enable communication between two signers, removing the impediment posed by verbal language. However, it fails to solve the issue when

(i) Everyone uses a different SL to communicate (among the worldwide 300 different SLs estimated by WFD)
(ii) Someone is uncommunicative through signs, creating a broad barrier among signers and between signers and non-signers.

Thus non-invasive recognition of SL is an important problem that has practical applications in fields such as education, communication, and human-computer interaction. Automatic machine translation, and speech translation, which translate between different spoken languages, have made significant advancements. This has made communication between users across various spoken languages possible, by translating spoken languages into text or audio. They have been incorporated into our lives as commonly used apps on our smartphones or through online browsers, like Google Translator, a pre-built model trained by Google, that enables automatic translation between two languages. By including SLs in such models as source or target languages, this strategy can be extended for SLs as well, to translate

(i) Spoken language text to SL;
(ii) SL to spoken language text;
(iii) Speech to SL and
(iv) SL to speech

in an end-to-end fashion [2]. Towards this aim, there also exist some android-based applications like Hand Talk Translator (translates text and audio to American sign language (ASL) avatar), Sanket (converts English words/sentences to ISL and plays them using WebGL-based avatar) [3].

The proposed study provides an overview of recent advances in ISLRT, including background information, development of ISL datasets, various ISLRT techniques, and challenges and future directions in ISLRT research. This study aims to achieve the following major contributions:

1. We compile and synthesize existing knowledge on Indian sign language (ISL), offering a holistic view of current research and advancements.
2. We highlight the unique challenges faced by the research community in India regarding ISL recognition and translation.
3. We identify significant gaps in existing ISL research, providing insights that can guide future research and development.

The remainder of this paper is structured as follows: Section 2 introduces the relevant terms and details about SL, sign language recognition (SLR), and sign language translation (SLT) tasks. Section 3 describes the conducted survey. Section 4 presents the background, evolution, and overview of ISL. Section 5 discusses various data acquisition methods. Section 6 details the datasets and dictionaries for ISLRT found in the literature. Section 7 provides an overview of the work on ISL since 2011. Finally, Section 8 concludes the paper by providing potential directions for future research.

2 | SIGN LANGUAGE

HoH people face communication difficulties that prevent them from fitting into society. It can affect them in many ways and cause social withdrawal, loneliness, isolation, or even frustration. Symptoms like mental and emotional issues or lack of confidence can be common among them. Sometimes it can get even worse due to the potential failure in school and the increasing unemployment rates. All of these are just a small symbol of their suffering and pain. Even though there are some ways to help HoH people integrate with others as discussed earlier (in Section 1), many problems remain to be solved. With the recent breakthroughs in artificial intelligence (AI) and the rapidly growing technology, the power of AI has enabled researchers to find solutions to these problems [4]. It is significant to mention that India is the leading country (ranked 1) contributing to the growth of SLR systems research over the past two decades [5]. In the subsections, we illustrate the SLRT terminologies from the literature that the readers are expected to get acquainted.

2.1 | Sign language translation

The translation of SL to text/speech is known as sign language translation (SLT). SLs are region-specific; for instance, the SL used in India is different from the SL used in America. Additionally, they might not be related to the regional language. For instance, in the case of English, a spoken language in the United States, the United Kingdom, and Australia, each has a unique SL. Overall, this emphasizes the importance of researching both SLT algorithms and regional signing customs [2].

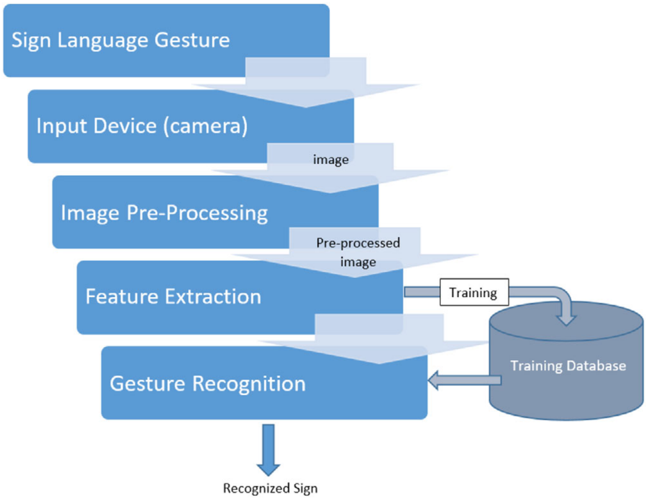

FIGURE 1 Sign language translation

SLT involves several stages of converting sign language into spoken or written language and vice versa. These stages may vary depending on the complexity and goals of the SLT system. Figure 1 presents the stages in the process of SLT. The process begins with capturing sign language gestures and movements. This can be done using various techniques, including cameras, depth sensors, or wearable devices. Raw input data may contain noise or irrelevant information. Pre-processing involves data cleaning and filtering to ensure accuracy. In the feature extraction stage, relevant features of the sign language gestures, such as hand shape, palm orientation, location, and motion, are extracted from the input image. In the sign language recognition stage, machine learning algorithms or computer vision techniques are used to recognize and classify sign language gestures. The system identifies the signs and their meanings. The system may analyse the context of the signs, including facial expressions, body movements, and non-manual signals, to improve accuracy and disambiguate meanings. Finally, once the sign language gestures are recognized, translation algorithms are applied to convert the signs into spoken or written language. These algorithms may involve natural language processing (NLP) techniques or rule-based systems. Contextual information may be considered during translation to ensure the output is contextually appropriate and grammatically correct. The translated text is synthesized into spoken language using text-to-speech (TTS) technology or written form. Overall, the goal of SLT is to create a seamless and effective communication tool for sign language users [6–9].

2.2 | Sign language recognition

Sign language recognition (SLR) refers to a process for recognizing and interpreting sign language gestures made by a human and converting the signals into machine-readable language for use in various applications [10–14]. SLR system

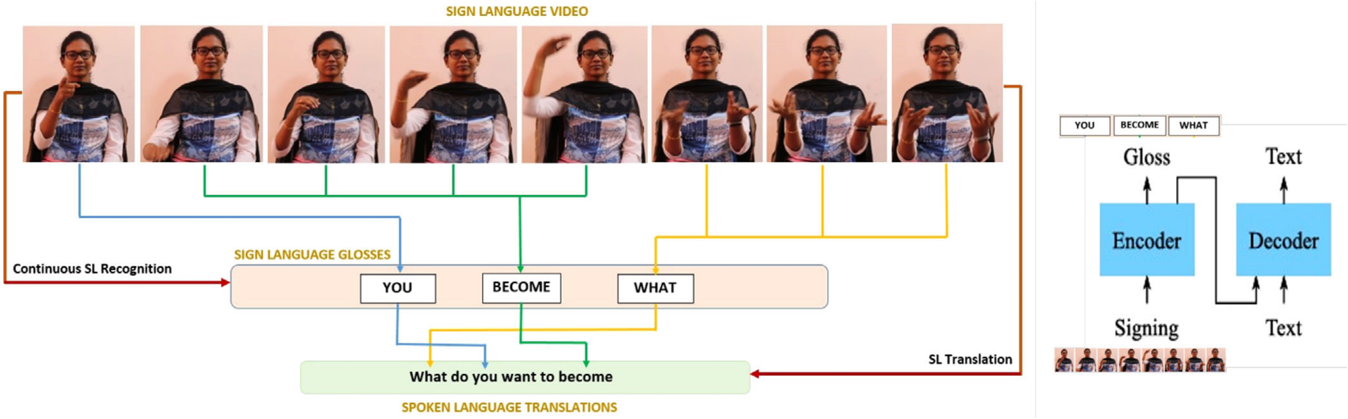

FIGURE 2 (a) Sign language recognition and sign language translation (b) joint end-to-end sign language recognition and translation.

identifies static or dynamic hand gestures of SL using machine learning, convolutional neural networks (CNN), artificial neural networks (ANN), recurrent neural networks (RNN), long short-term memory (LSTM)[15–17], etc. SLR systems also exploit techniques like computer vision, depth sensors, or wearable devices. The distinction between translation and recognition is frequently ambiguous within the context of scientific literature [18]. Sign language recognition is commonly categorized into two main types: isolated recognition and continuous recognition. The process of isolated sign recognition involves the classification of individual signs. Continuous sign language recognition, on the other hand, takes into account sequences including two or more signs. For example, the application of sign language recognition involves transcribing sign language videos into sign language glosses. The input, consisting of videos, and the output glosses, are both associated with the same language. In SLT, however, sequences of signs are translated into a different language. The translation precision depends on the recognition precision, so it is a crucial phase in SLT. So, in the proposed work we emphasize the study of recognition systems in ISL literature and summarise algorithms in Section 7.

2.3 | Joint end-to-end sign language recognition and translation

Joint end-to-end SLRT typically involves training a neural network model to directly map the input video of sign language gestures to the output text or speech translation. The training uses deep learning techniques, such as Transformers [19]. It is a more advanced approach to recognizing and translating SL gestures because recognition and translation serve in a single integrated system. One of the key advantages of the joint end-to-end SLRT is that it can be more accurate than separate recognition and translation stages (2-stage), as the system can learn to recognize and translate SL gestures simultaneously, taking into account the context and meaning of the gestures. One of the SL's intermediate stages in the 2-stage translation is its transcription as glosses, a text-based representation. For instance, given the sentence, 'What do you want to become?,' the gloss sequence in ISL would be 'You become what' (Figure 2). However, due to gloss's inability to effectively reflect the information found in SLs, glosses alone may not be sufficient to generate appropriate sentences and sometimes may act as an information bottleneck. So, they are often used in research as an intermediate stage in the automatic translation process only to guide the translation system's learning. According to Kollar et al. [19], a joint approach, generates better results. However, the biggest challenge in developing effective joint end-to-end recognition and translation systems for SLs is the need for a large amount of training data. As per the author's knowledge, no work has implemented the ISLRT system for sentence translation using a joint end-to-end approach before the publication of this work.

Apart from the above mentioned, it is worth mentioning sign language notation, and sign language generation (SLG). Sign language notation transcribes signs like phonetic alphabets for spoken languages. It is proficient in representing all signs used in every SL. It is used as a standard since it does not depend on regional variances in SLs. There are various ways found in the literature for representing signs, for instance, HamNoSys [20, 21], SiGML [21], Stokoe Notation [22, 23], and SignWriting [24, 25](more details in Appendix A). Sign language generation [22, 26] is about creating SL utterances or animations from written or spoken language input (details in Appendix B). SLG entails creating an avatar or skeleton that mimics the necessary gestures, facial expressions, and signs (details in Appendix B).

3 | LITERATURE REVIEW METHOD

3.1 | Search strategy

Since the ISLRT technology has significantly developed over the past 10 years, this survey was conducted by searching articles published between 2011 and 2024. To present an in-depth review of sound ISLR research, we adhere to the following standards during our literature search. Using a set of keywords, the study searched the research area of automated ISL recognition and translation in three scientific journal databases: Web of

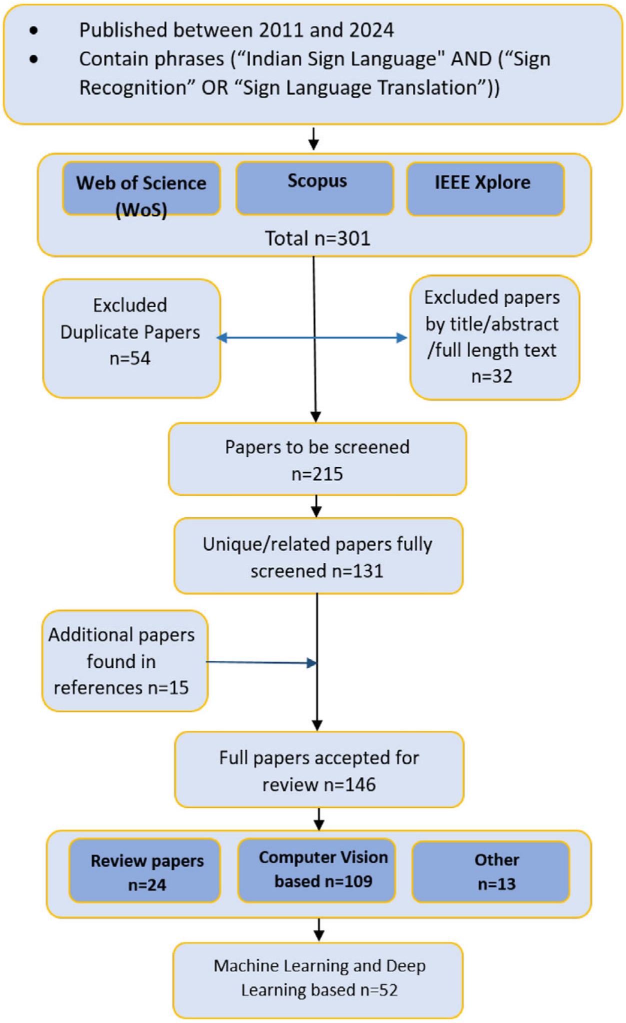
Published between 2011 and 2024
Contain phrases ("Indian Sign Language" AND ("Sign Recognition" OR "Sign Language Translation"))
Web of Science (WoS)
Scopus
IEEE Xplore
Total n=301
Excluded Duplicate Papers n=54
Excluded papers by title/abstract /full length text n=32
Papers to be screened n=215
Unique/related papers fully screened n=131
Additional papers found in references n=15
Full papers accepted for review n=146
Review papers n=24
Computer Vision based n=109
Other n=13
Machine Learning and Deep Learning based n=52

FIGURE 3 Literature search of study selection.

Science (WoS), Scopus, and IEEE Xplore. To find related contributions, search engines were queried for papers published on or after 2011 that contained specific key phrases ('Sign Language' AND ('Indian Sign Recognition' OR 'Sign Language Translation')) in their titles or abstracts. We include reputed journal articles as well as conference papers, and any paper that is included proposes, implements, and evaluates a sign language machine translation system from an Indian sign language to a spoken language. Our research primarily focuses on the non-intrusive translation of Indian sign languages into text, so the articles that do not primarily focus on the computer vision-based ISLRT were excluded. Each paper was reviewed and agreed upon by at least one researcher before inclusion. After searching, there were 301 articles found from the search engines, and we shortlisted 215 of them following the criteria presented in Figure 3.

3.2 | Study selection

After a full-text screening for relevance to the topic, this review included 146 published articles. The most recent update to the included papers was in November 2024. All accepted articles were categorized based on their approaches, namely, survey/review, computer vision-based, and other (sensor/glove-based). Additionally, a few articles were included for their contributions to knowledge about ISL background, ISL datasets/dictionaries, data acquisition methods, and regional ISL language works. Ultimately, the study focused on 52 full articles that addressed machine learning and deep learning approaches with a computer vision perspective towards ISLRT, as these are the most prominent methods.

4 | INDIAN SIGN LANGUAGE BACKGROUND

ISL is an officially recognized sign language used in India and some neighbouring regions of India. The estimated number of ISL users is 6 million in India, and 6.815 million in neighbouring countries [27, 28]. India has a rich and diverse sign language landscape, with multiple SLs used by the deaf community across different regions [29]. Some of the major SLs used in India include Tamil sign language (TSL), Bangla sign language (BdSL), Kannada sign language (KSL), and Marathi sign language (MSL) [30]. These SLs have evolved, drawing from the local cultures and languages, and have unique linguistic features, grammar, and vocabulary. Various factors like the history of deaf education, local languages, and the development of the deaf community in India influenced the evolution of SL. The first school for the deaf was established in India, with the first recorded use of SL in deaf education and taught since 2001. With the growth of the deaf community in India, a unique deaf culture has emerged, with ISL playing a central role in preserving and promoting this culture. However, due to the low level of English comprehension among Indian HoH people, they have limited access to spoken and written content. So, to help HoH people at institutions and locations where communication between HoH and hearing people occurs, ISL interpreters are urgently needed. However, India has less than 300 certified interpreters, indicating a severe shortage compared to the HoH population. With the advent of new technologies, there has been a growing interest of researchers in the development of ISLRT systems, which have the potential to advance the recognition and use of ISL [31]. The objective of the proposed work is to consolidate the existing ISL datasets and recognition algorithms with their advantages and limitations. Also, it provides future direction to create datasets for unconstrained continuous SL translation.

4.1 | Classification of ISL signs

In communication, SLs are expressed using articulators. The body's movable and visually perceivably perceptible parts, such as the head, upper body, and facial features, are known as SL articulators. These articulators are used to produce, restrict, and contrast words with one another to portray prosody. Five fundamental factors of the ISL system mentioned below are dependent on these articulators:

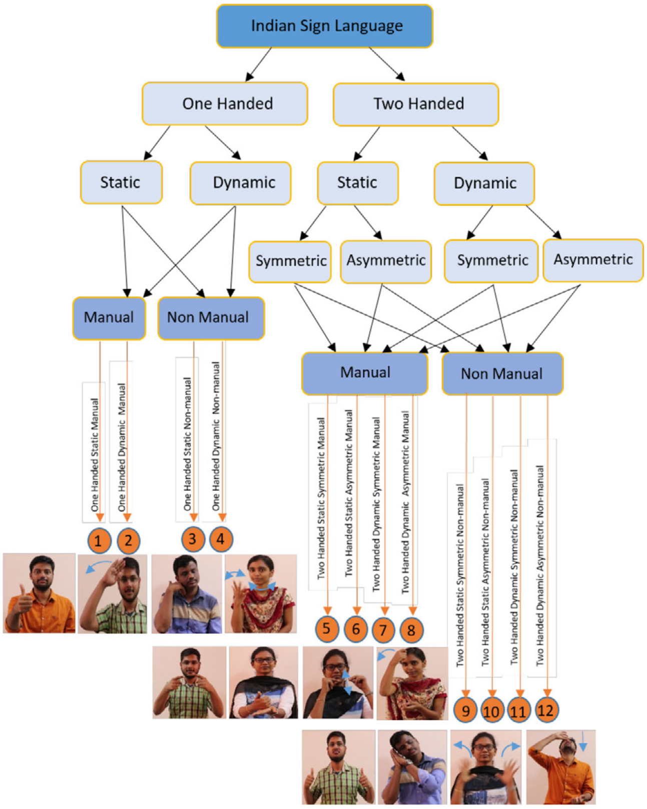
Indian Sign Language
One Handed
Two Handed
Static
Dynamic
Static
Dynamic
Symmetric
Asymmetric
Symmetric
Asymmetric
Manual
Non Manual
Manual
Non Manual
One Handed Static Manual
One Handed Dynamic Manual
One Handed Static Non-manual
One Handed Dynamic Non-manual
Two Handed Static Symmetric Manual
Two Handed Static Asymmetric Manual
Two Handed Dynamic Symmetric Manual
Two Handed Dynamic Asymmetric Manual
Two Handed Static Symmetric Non-manual
Two Handed Static Asymmetric Non-manual
Two Handed Dynamic Symmetric Non-manual
Two Handed Dynamic Asymmetric Non-manual
1
2
3
4
5
6
7
8
9
10
11
12

FIGURE 4 Hierarchical classification of ISL signs.

(i) hand and head recognition
(ii) hand and head orientation
(iii) hand and head movement
(iv) location of hand and head
(v) shape of the hand

As proposed by Kumar et al. [6], ISL articulators can be divided into two categories: word representation and finger spelling. The first one enables a signer to use the articulators to convey the meaning of words. The second one serves as a means of representing the spoken alphabet and can be used to spell out the names of people, places, and other things. Classification of ISL signs involves categorizing signs into specific classes based on various articulators and non-manual signals. The classification is essential for organizing and understanding the vocabulary of ISL. Figure 4 presents the hierarchical classification of ISL signs [7, 32]. Table 1 lists terminologies used as the basis for the classification [33]. Table 2 presents 12 categories of ISL sign and their sample articulations, followed by some additional observations about ISL in Table 3 [34–37].

TABLE 1 Terminologies used in ISL classification.

Terminology	Description
1. Manual sign	Manual signs comprise hand(s) shape, motion, and location parameters. These hand gestures, which can be made with a hand or both hands, do not express additional emotional information. The most prevalent instances of manual signs are ISL numerals, finger-spelled words, and alphabets.
2. Non-manual	Non-manual gestures, can be made with or without hand motions, such as shoulder shrugs, torso bending, and movements of the eyes, brows, cheeks, and mouth. For instance, in ISL, words like 'fear' 'not' and 'sleep' also include a facial emotion and a manual sign.
3. One-handed sign	One-handed signs are presented by only a single dominant hand. These signs are further split into static or dynamic signs. Manual and non-manual signs are further categorized into static and dynamic signs.
4. Two-handed sign	Two-handed signs are presented by both hands. They are also classified into two categories static and dynamic. The static and dynamic signs are further split into symmetric and asymmetric classes. Symmetric and asymmetric signs are classified further into manual and non-manual.
5. Symmetric sign	Signs where both hands are in an active position.
6. Asymmetric sign	Signs where the dominating hand is more active than another hand.
7. Static sign	Signs that involve holding a handshape or gesture for some time without movement.
8. Dynamic sign	Signs involve movement or change in hand shape or gesture to convey meaning.

5 | DATA ACQUISITION METHODS FOR SIGN LANGUAGE RECOGNITION

The latest research fields use a deep understanding of human behaviours from visual input devices. It is motivated by fascinating scientific challenges and the rise in social demands for diverse software applications, such as virtual environments, surveillance systems, medical support systems, etc. Choudhury et al. [38] highlight various hand gesture recognition techniques and their applications in automatic SLR. One of the most popular representations of SLs often records the signer's utterances in a video while simultaneously recording all the manual and non-manual elements. However, videos are not the only medium available to portray SLs, and the camera is not the only acquisition tool. For SLR, the input data is obtained using varied methods [39]. We divide the same into five categories glove-based, armband sensor [40], leap motion [17, 41], Kinect-based, and vision-based, as shown in Figures 5 and 6. In the glove-based method, various elements similar to the hand's colour, orientation (palm facing forward), configuration (fingers bent, fingers moment, and palm bent), and site of articulation (finger joints, wrist, elbow, or shoulder) are extracted, using different kinds of gloves and sensors. The glove-based approaches are further categorized as, cyber-glove [42], data glove [43, 44], and colour glove [45–47]. Das et al. [48] discuss glove-based ISL interpretation techniques. A sensor armband is a wearable technology to measure and track various physiological data, such as heart rate, blood oxygen levels, and activity levels [49]. In a leap motion controller, a device tracks the hand's movement, and through a USB cable, signals are transferred to a computer and translated into commands [50]. In a Kinect-based method, a depth sensor acquires skeleton data and records the depth

TABLE 2 Categories of ISL signs and their example articulations.

#Sign category	Sign category	Number of hands	Example sign	Head movement	Eye movement
1	One handed static manual	1	Good	No	No
2	One handed dynamic manual	1	Hi	No	No
3	One handed static non-manual	1	Fever	Yes-down on dominant side	Yes- closed
4	One handed dynamic non-manual	1	Not	Yes-rotation	No
5	Two handed static symmetric manual	2	Chat	No	No
6	Two handed static asymmetric manual	2	Help	No	No
7	Two handed dynamic symmetric manual	2	Speak	No	No
8	Two handed dynamic asymmetric manual	2	Comb	No	No
9	Two handed static symmetric non-manual	2	Dare	Yes-move back	Yes- bigger in size
10	Two handed static asymmetric non-manual	2	Sleep	Yes-down on dominant side	Yes- closed
11	Two handed dynamic symmetric non-manual	2	Afraid	Yes-move back	Yes- bigger in size
12	Two handed dynamic asymmetric non-manual	2	Medicine	Yes-down on back side	No

TABLE 3 Few observations about ISL.

Observation	Example signs/Description
1. The hand shape for signs is the same but they move in opposite directions	PASS and FAIL
2. Signs have the same hand shape but a different place of articulation and movement pattern	MONEY and PAY
3. The place of articulation is the head which is the same for multiple signs	THINK and KNOW
4. There is no temporal conjugation.	Signs for BEFORE, THEN, and AFTER are used to represent the past, present, and future, respectively.
5. Interrogative phrases with question terms like WHAT, WHERE, WHICH, HOW, etc. are placed at the end	English: 'What do you want to become' ISL: 'You become what'
6. ISL is described as subject-object-verb (SOV) language	English: 'He likes fruits' ISL: 'He fruit like'
7. SL does not use articles, linking verbs, or prepositions	No signs for do, is, for, by, during, etc.

and RGB image. In a vision-based system, a camera is used to capture the movements of the user's hands [51]. This survey focuses on a vision-based system, so it is further elaborated in Section 5.1. For the literature on other acquisition methods (Figure 6) refer to Appendix C.

5.1 | Vision-based approach for sign language recognition

ISL is a visual-gestural language that expresses meaning through body movement, facial expressions, and hand gestures. Vision-based ISL systems [52–55] use smart/web/mobile cameras as the device to capture/record images/videos. A computer vision system analyses the signer's actions using computer vision techniques to translate ISL gestures into text or speech. It is the richest data acquisition method, as it allows the expression of all the manual and non-manual aspects and usage of the surrounding space around the signer. Signers also use the surrounding area for various purposes, such as positioning an object at some point so that they can later refer to it from that location. In

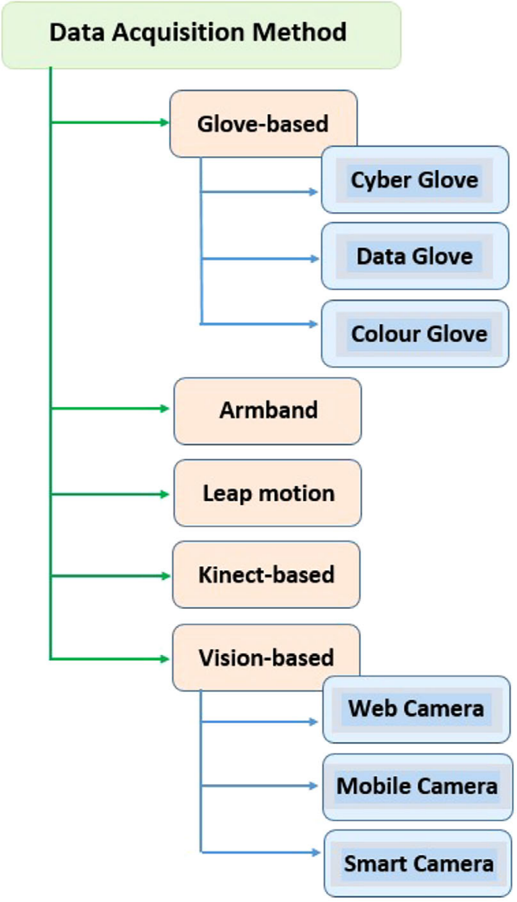

FIGURE 5 Data acquisition methods used in sign language translation.

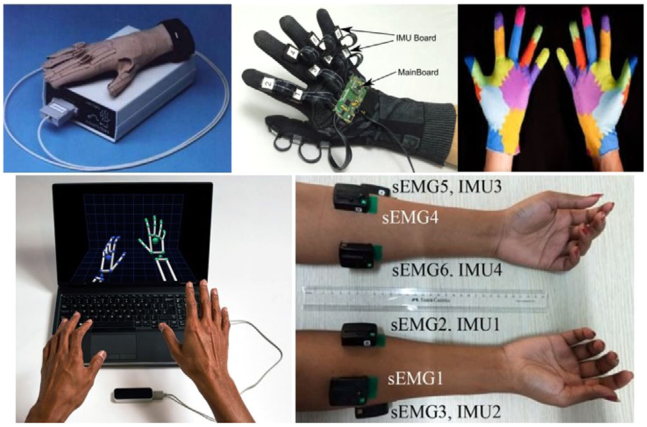

FIGURE 6 (a) Cyber glove, (b) data glove, (c) colour glove, (d) leap motion device and (e) sensors Armband.

vision-based ISL systems, users are not required to wear any uncomfortable external devices and only need to use their hands inside the camera collection range. A camera is economical, convenient, and simple to use.

Ahmed et al. [56] proposed a vision-based double-handed dynamic sign recognition system for 24 isolated word signs. The authors extracted features from faces and hands using trajectory tracking of the moving hand. The authors obtained an accuracy of 90% using dynamic time warping (DTW). Rehman et al. [57] proposed a computer vision-based system for recognizing Bengali sign language gestures in real-time. The system recognizes 30 consonants and six vowels in Bengali. The authors used a colour-based skin segmentation algorithm to extract the hand region from the video stream and apply a series of image processing techniques to detect and recognize the hand gestures. They use a combination of Haar-like feature-based cascaded classifier, and KNN classifier to recognize the gestures. The authors report an average recognition accuracy of 98.17% (vowels) and 94.75% (consonants) for their system. The colour glove-based approach also uses a vision-based acquisition method, so the colour glove-based approach can be considered a hybrid category. Refs. [57–63] use a colour-based skin segmentation approach. Amrutha et al. [64] present a comparison of sensor-based and vision-based methods. Ahmed et al. [1] give a detailed review of systems-based sensory gloves for SLR. Apart from many benefits, a vision-based SLR approach has a few limitations. Environmental elements like light, skin tone, background conditions, and occlusion have a significant impact on its performance. Also, several image processing techniques are needed for which the user must always have a camera. Table C1 in Appendix C presents the pros and cons of using various acquisition methods.

Apart from the well-known data acquisition methods that Figure 5 presents, there are a few state-of-the-art, such as crowdsourcing platforms like Amazon Mechanical Turk (MTurk), which is used to collect sign language data from a huge number of individuals. This data can be used to develop and train machine learning models for SLR.

6 | ISL DATASETS AND DICTIONARIES

Compared to American sign language (ASL), fewer models have been proposed to recognize ISL. So, ISL is a field where much more research is yet to be done. Many of the proposed ISL models can perceive some characters, digits, words, or simple sentences but not continuous/unconstrained ones. It is mainly due to the unavailability of standard datasets, which consist of nearly every basic word and their usage in sentences that can be used for ISL continuous translation. Many researchers created their custom datasets for characters, digits, and static/dynamic word ISL signs. Recently, a small number of datasets with videos for limited ISL sentences were created, in around 2021. Two such datasets are the Indian sign language video dataset (INSIGNVID) and the Indian sign language dataset for continuous sign language translation and recognition (ISL-CSLTR). This section presents available datasets (Section 6.1) and dictionaries (Section 6.2) for ISL recognition and translation.

6.1 | Indian sign language datasets

This section summarizes the available ISL datasets to the best of our knowledge. The majority of these are created specifically for SLR (containing images and glosses), whereas some are suitable for SLT (including videos and translations in English text/audio).

Listed here are some of the prominent datasets:

- IIITA-ROBITA ISL gesture database: The IIITA-ROBITA dataset [65] includes videos of 23 distinct word gestures recorded with a Sony Handycam at 30 fps and 320 by 240 pixels. The dataset consists of static and dynamic gestures with the same background with varying lighting conditions. It is a public dataset (https://robita.iiita.ac.in/dataset.php).
- Indian lexicon sign language dataset (INCLUDE): The research paper [66] presents a dataset of ISL gestures, INCLUDE. The INCLUDE dataset consists of over 4287 videos of 263 ISL word signs performed by seven experienced student signers. The videos were recorded in varying lighting conditions, and no constraints were imposed on the surroundings or attire. The INCLUDE dataset is publicly available. (https://zenodo.org/record/4010759#.ZClcdXZBxPZ). A subset of INCLUDE is INCLUDE-50, including 50 words for easy testing of the recognition model.
- Indian sign language 3D dataset (ISL3D): A dataset of 500 ISL word signs with 59 joints was created by Kumar et al. [67] using a 3D motion capture system (VICON Motion System Ltd., Oxford, UK). To precisely replicate 3D human activities, the motion capture system (MOCAP) utilizes multiple cameras to display each skeletal joint location in 3D space. Spatial points in 3D space represent the human joints' 3D skeleton. The authors developed 50,000 3D word sign videos for 500 words with 100 signs each.

- Indian sign language video dataset (INSIGNVID): INSIGNVID is an ISL video dataset of 55 frequently used sentences of 5 s duration each, along with their English translations [92]. To make background subtraction easier, the authors recorded videos on green backgrounds.
- Indian sign language dataset for continuous sign language translation and recognition (ISL-CSLTR): ISL-CSLTR is the sentence-level completely labelled dataset. It is created by Elakkiya et al. [8] with a vocabulary of 1036 words forming 100 sentences, resulting in 700 videos' dataset (18,863 sentence level frames). The colour videos of seven signers are recorded. It is a public dataset (https://data.mendeley.com/datasets/kcmpdxky7p/1).
- The corpus for Indian sign language recognition (CISLR): CISLR [95] consists of 4765 words in the form of 7050 videos for word-level recognition in ISL. The authors created CISLR by scraping and curating data from two public internet sources, the first being the Indian Sign Language Research and Training Centre (ISLRTC) [36], and the second, the Indian Sign Language Dictionary [35]. The authors scraped public ISL videos from YouTube and annotated them with an English word.
- Indian sign language dataset (ISLD): ISLD is a character-level dataset containing greyscale sign images on dark backgrounds for 26 alphabets and nine digits (excluding 0). It is a public dataset (https://www.kaggle.com/datasets/vaishnaviasonawane/indian-sign-language-dataset).
- Indian sign language dataset (ISL): It is a Complete ISL dataset on a character level containing coloured images on light backgrounds for 26 alphabets and nine digits (excluding 0). It is a public dataset (https://www.kaggle.com/datasets/prathumarikeri/indian-sign-language-isl).
- Indian sign language dataset (recognized by ISRTC): This dataset includes all the alphabets and numerals in Indian hand recognition provided by ISRTC (Indian Sign Research and Training Centre). This dataset is in black and white background for faster computing and for getting better accuracy while training. It is a public dataset (https://www.kaggle.com/datasets/kshitij192/isl-dataset).
- Indian sign language (ISLRTC referred): It is an ISL dataset including alphabets and digits. The collection has 1000 images for each class with different lighting conditions and backgrounds, producing 36,000 images. The images were captured with a 720p laptop camera. It is a public dataset [106]. (https://www.kaggle.com/datasets/atharvadumbre/indian-sign-language-islrtc-referred).
- Hand gesture recognition: This dataset contains videos of 50 unique ISL word signs, each containing 40 videos of 20 frames. Colour images with variable backgrounds are captured. It is a public dataset (https://www.kaggle.com/datasets/kshitij192/action-recognition).
- Indian traffic sign detection benchmark dataset YOLO format: Jaseer et al. [107] developed the Indian traffic sign detection dataset. This dataset primarily applies to projects involving traffic sign detection using the YOLO algorithm. The dataset is formatted according to the YOLO framework. It contains 1264 images, all of which have been fully annotated using the labelling tool. The dataset comprises images depicting three distinct categories of Indian traffic signs https://ieee-dataport.org/documents/indian-traffic-sign-detection-benchmark-dataset-yolo-format).
- ISLTranslate: The ISLTranslate dataset [99] uses ISLRTC's [36] educational sign videos. The targeted viewers for these videos are school children and parents, and the range of words covered in the videos is beginner-level. The authors also use another resource from Deaf Enabled Foundations (DEF) [101]. The ISLTranslate consists of 2685 videos from DEF, and the remaining 28537 videos are from ISLRTC.
- iSign Dataset: The iSign dataset [100] is a benchmark for Indian Sign Language Processing. The authors used two publicly available datasets ISLTranslate [99] having 31k video-sentence pairs, and CISLR [95] having 7k video-word pairs for isolated sign language recognition. The iSign dataset comprises 118k video-English sentence/phrase/word pairs and is free for research. https://huggingface.co/datasets/Exploration-Lab/iSign.

Apart from these prominent datasets on ISL, many are created by researchers. Kaur et al. [80] created a dataset of 72 signers for 26 ISL alphabets on a uniform background with position, scale, and rotation variations. There are a total of 1865 images. In the second dataset, alphabetic signs appear on complex backgrounds. 100 samples per alphabet are captured, with four different background modifications. There are 2600 images in the second dataset. Adithya et al. [87] created a dataset of hand gestures for emergency situations, based on ISL. The authors noted that few ISL datasets focus specifically on emergencies. They created a video dataset for eight emergency words, such as 'help', 'fire', and 'police'. They recorded videos of eight people performing these gestures, using a standardized procedure to ensure consistency. Their dataset includes approximately 50 videos each for raw and cropped categories showing one of the selected gestures. The dataset created by Wadhawan et al. [88] includes both coloured and greyscale images for static signs. There are 100 different sign gestures, which include 67 regularly used words (e.g., tongue, food, etc), 0–9 digits, and 23 alphabets. It consists of 35,000 images, containing 350 for each sign that varies in size, colour, and lighting condition to help the classifier make more accurate generalizations. Gangrade et al. [90] collected an ISL dataset of alphabets and numbers gestures with cluttered backgrounds from 12 distinct signers using the Microsoft Kinect sensor. It includes 43200 RGB images. Sharma et al. [91] created a dataset for alphabets by applying various data augmentation techniques, different lighting conditions, and the same dark background. The dataset consisted of more than 150,000 images of all 26 finger-spelled alphabets, including approximately 5500 samples of each. Sharma et al. [94] created a dataset of 5408 static gesture images of 26 alphabets with the help of 65 signers. To motivate research in this field, Mistree et al. [9] generated an ISL corpus of 13,720 sentences and a video dataset of 47,880 ISL videos. However, it is not yet available in the public domain to the best of our knowledge. Table 4 provides a summary of the ISL datasets, detailing the

TABLE 4 Indian sign language datasets.

Reference	Dataset name	Sign category	Size (Images/Videos)	# Signers	Modality
Nandy et al. (2010) [65]	IIITA-ROBITA ISL gesture database	23 words	600 videos	—	RGB
Lilha et al. (2011) [68]	—	26 alphabets	2340 images	3	RGB
Kishore et al. (2012) [69]	—	80 alphabets, numbers and words	800 images	10	RGB
Katoch et al.(2012) [70]	—	60 words	720 images	12	RGB
Sahoo et al. (2014) [71]	—	26 single handed alphabets	2600 images	10	RGB
Sharma et al.(2014) [72]	—	Ten single handed digits dark background	5000 images	100	RGB
Tripathi et al. (2015) [73]	—	Ten sentences	500 videos	5	RGB
Mehrotra et al. (2015) [74]	—	37 words	2775 images	15	RGB
Ansari et al. (2016) [75]	Indian Kinect	140 alphabets, digits and words	5041 images	18	RGB
Sharma A et al. (2016) [76]	—	26 alphabets	1068 images (light background)	4	RGB
Kharate et al. (2016) [77]	—	36 alphabets, digits	36,000 images with black background	40	RGB
Prasad et al. (2016) [78]	—	80 alphabets, digits and words	720 videos	9	RGB
Kumar et al. (2017) [79]	—	50 words	7500 videos	10	RGB
Kaur B et al. (2017) [80]	ISL complex background	26 alphabets	1865 (images uniform background) 2600 (images complex background)	72	RGB
Kumar et al. (2017) [81]	—	30 words	2700 images	10	RGB
Kishore et al. (2018) [82]	—	500 words	18,000 images	4	RGB-D
Rao et al. (2018) [83]	—	200 words	60,000 images	5	RGB
Ravi et al. (2019) [84]	BVCSL3D	200 words	20,000 videos	10	RGB-D
Athira et al. (2019) [85]	—	26 alphabets, 11 single-handed dynamic words	900 images 700 videos	7	RGB
Marippan et al. (2019) [86]	—	80 words, 50 sentences	800 word and 500 sentences videos	10	RGB
A. Sridhar et al. (2020) [66]	INCLUDE	263 words	4292 videos	7	RGB
A. Sridhar et al. (2020) [66]	INCLUDE50	50 words	958 videos	7	RGB
Kumar et al. (2020) [67]	ISL3D	500 words	50,000 videos	5	RGB-D
V. Adithya et al. (2020) [87]	—	8 words	824 videos	26	RGB
Wadhawan et al. (2020) [88]	—	100 alphabets, digits, words	35,000 images	—	RGB and Grey-scale
Raghuveera et al. (2020) [89]	—	140 alphabets, numbers, words	4600 images	21	RGB-D
Gangrade et al. (2020) [90]	—	36 alphabets, digits	43,200	12	RGB
Elakkiya et al. (2021) [8]	ISL-CSLTR	100 sentences	700 videos	7	RGB
Sharma A et al. (2021) [91]	—	26 alphabets	150,000 images	—	RGB
Mariappan et al. (2021) [15]	CasTalk-ISL Dataset	50 words	5000 videos	10	RGB
Mistree et al. (2021) [92]	INSIGNVID	55 sentences	1289 videos	4	RGB
Sharma S et al. (2021) [93]	—	43 alphabets, words	2150 images	50	RGB
Sharma S et al. (2022) [94]	—	26 alphabets	5408 images	65	RGB
A. Joshi et al. (2022) [95]	CISLR	4765 words	7050 videos	71	RGB
Subramanian et al. (2022) [96]	—	13 words	11,700 images	—	RGB
Katoch et al. (2022) [97]	—	36 alphabets, digits	36,000 images with white background	3	RGB
Rajalakshmi et al. (2023) [98]	Indian Isolated Word Sign Dataset (IIWS)	500 words	3000 videos	7	RGB
Joshi et al. (2023) [99]	ISLTranslate	11K words, 31K sentences	31K videos	—	RGB
Joshi et al. (2024) [100]	iSign	40K words, 118K sentences	118K videos	—	RGB

publication year, sign category, number of videos or images, number of signers, and the modality used.

6.1.1 | Evaluation of the existing datasets

A few datasets like ISLD, ISL, IS-RTC, and ISLRTC are available from varied internet sources. They focus on alphabet and numerals recognition. ISL complex background [80] is the most diversified dataset for ISL alphabets with uniform and complex backgrounds participated by 72 signers. IIITA-ROBITA [65] includes videos recorded in varied lighting conditions but with consistent backgrounds. However, INCLUDE [66] removes the barrier of imposing consistent surroundings or the signer's attire. Predominantly these datasets focus on limited isolated words or short phrases rather than full, continuous sentence-level signing. CISLR [95] was created in the year 2022 by scraping and curating data from two public internet sources, the first being the ISLRTC [34], and the second being the Indian sign language dictionary [33]. It includes 7050 videos of 71 signers and has 4765 words of vocabulary. It is the largest diversified dataset to be used in ISL word recognition. However, this dataset restricts the scope of research for more complex translation tasks. BVCSL3D [84], and ISL3D [67] are multimodal 3D word datasets captured in a controlled environment using Microsoft Kinect sensors, and MOCAP systems, respectively. This limits the dataset's applicability to the model's generalization to different backgrounds, lighting, and noise conditions encountered in real-life scenarios. INSIGNVID [92] is the first ISL sentence-level video dataset for 55 sentences. The dataset was created with four right-handed signers of different ages and genders, dressed in black against a green background for clarity. The dataset was authenticated by an ISLRTC-certified interpreter, ensuring its consistency and reliability for ISL translation. ISL-CSLTR dataset [8] features seven signers of which two are native signers from a deaf school, and four are student volunteers from SASTRA Deemed University in Thanjavur, Tamil Nadu, India. The INSIGNVID and ISL-CSLTR datasets are structured with sentence-level annotations that link video-recorded gestures to short, spoken language sentences. These datasets serve as valuable resources for researchers aiming to evaluate their architectures against state-of-the-art models in gesture recognition and sign language translation tasks. However, generative models like the Transformers [19] cannot use these datasets. Recently, two datasets, the ISLTranslate [99] with 31K sentences, and the iSign [100] with 118K sentences were published. These datasets are the fusions of existing public datasets like DEF [101], ISLRTC [36], and CISLR [95]. The iSign [100] can be used in various generative ISL tasks. An assistant professor of sign language linguistics at ISLRTC translated and validated a random selection of 100 sentences by watching the videos and producing English translations.

Generative models require extensive data from diverse sources. To accelerate research in the Indian Sign Language (ISL) domain, larger continuous datasets can be developed by merging various existing datasets such as ISL-CSLTR [8], ISLTranslate [99], and iSign [100]. Crowdsourcing platforms like Amazon MTurk, which involves gathering services or content from a broad group of people, often through online communities, can be leveraged for tasks like data collection, annotation, labelling, and transcription.

6.2 | Indian sign language dictionary

A dictionary is an initial step in meeting the communicative needs of the deaf community. Any natural language that has thousands of native users need to have a well-documented database of its structural and pragmatic descriptions. Indian sign language is no exception. SignDict is an online sign language dictionary [108] that provides translations of English words and phrases into ISL, along with video demonstrations of the signs. There are currently very few comprehensive dictionaries of ISL available online that can help researchers curate the ISL dataset. Here, signs are made available in the form of videos. ISLRTC dictionary was developed by the Indian Sign Language Research and Training Centre (ISLRTC), an autonomous organization under the Ministry of Social Justice and Empowerment, Government of India. It includes over 10,000 signs, with videos and descriptions, and is available on the ISLRTC website [36]. A comprehensive ISL dictionary is urgently required to improve communication, build a primary database, and promote extensive use of ISL. Many hearing people like family, friends, and colleagues of HoH person are interested in learning ISL. Professionals in the field, like teachers and interpreters working with HoH people, often need to look up signs. HoH people are also interested in learning English or Hindi words for particular signs. Therefore, to meet this requirement of both HoH and hearing people, the objective of ISH Shiksha [103] is to provide online interactive, accessible education for deaf and hearing people and build a bridge between the two worlds. The primary objective of ISH News [102] is to disseminate news and information in a format that is easily accessible to the estimated 18 million HoH individuals in India. Table 5 lists public dictionaries and their reference links. This information can be useful for researchers interested in developing their datasets and evaluating ISLR systems. For instance, Mariappan et al. [15] created a word dataset named CasTalk–ISL Dataset using the ISL dictionary. Dumbre et al. [106] created a custom dataset with 1000 images per character using the signs demonstrated on the ISLRTC website. Joy et al. [108] used signs from ISLRTC to include them in a smartphone app. Subhash et al. [109] created a custom dataset where the signer replicated the signs from examples available online at [35] that were displayed on a monitor.

7 | RESEARCH WORK ON INDIAN SIGN LANGUAGE RECOGNITION SYSTEMS

This section presents a comprehensive survey of research works published in the past for more than a decade (since 2011) when the development of ISL recognition systems gained momentum. Before this period, research in sign language recognition

TABLE 5 Indian sign language public dictionaries.

Dictionary name	Reference
ISL Dictionary (by Ramakrishna Mission Vivekananda University)	indiansignlanguage.org [35]
ISLRTC Dictionary (by Indian Sign Language Research and Training Centre)	islrtc.nic.in [36]
ISL Dictionary (by Mook Badhir Mandal)	indiandeaf.org/ISL [37]
DEF ISL (by Deaf Enabled Foundation)	def.org.in/ [101]
ISH News (2018)	indiasigninghands.com/projects/ishnews/ [102]
ISH Shiksha (2021)	ishshiksha.com/ [103]
Text to Indian Sign Language conversion (2018) Thapar Institute of Engineering and Technology, Patiala	islfromtext.in [22, 104]
Talking Hands	talkinghands.co.in [105]

was relatively limited, owing to the challenges posed by the complexity of sign languages and the availability of appropriate technologies. However, technological advancements in computer vision and machine learning in the last decade paved the way for more sophisticated approaches to ISL recognition that catered to the unique gestures and linguistic nuances of ISL [110]. This section meticulously categorizes various research efforts by focusing on diverse approaches employed for recognition and classification tasks within the ISL domain. Ranging from recent machine learning and deep learning methodologies to more traditional approaches like motion tracking and graph-based techniques, this section offers a panoramic view of the advancements achieved in ISL recognition, shedding light on not only noteworthy accomplishments but also the attendant challenges and the promising paths ahead.

7.1 | Machine learning based approach for sign language recognition

Machine learning (ML) is used in a wide range of applications. Perhaps, one of the most well-known domains of machine learning is NLP. Sign language is a natural language, so many researchers have used ML techniques for SLR systems (Table 6).

Rekha et al. [111] demonstrated an ISL double-handed alphabet recognition system. The authors acquired a dataset of 26 alphabet signs, 23 of which are static and three are dynamic. The authors used YC_bC_r skin colour model to segment and detect the hand region, a support vector machine (SVM) to classify static signs, and dynamic time warping to classify dynamic signs. The method using the histogram of edge frequency and SVM used by Lilha et al. [68] reports an accuracy of 98% for recognizing 26 alphabets. They used a non-skin colour wristband to identify the palm region.

Agrawal et al. [112] demonstrated a system for recognizing double-handed signs. They used a camera to capture 235 images of 36 signs. The authors used Otsu's algorithm for segmentation and fused shape descriptors, histogram of oriented gradients (HOG) descriptors, and scale-invariant feature transform (SIFT) features to compute a feature vector. In the recognition phase, the authors used a multi-class SVM (MSVM) to classify ISL signs. The authors concluded that the system performs better with a fusion of extraction methods. Dixit et al. [70] proposed an approach to recognize double-handed word signs. They extracted structural shape descriptors using Hu invariant moment for scale and position-invariant pattern identification. MSVM was used for training and recognizing signs to obtain 96.23% accuracy. Adithya et al. [113] recognized single and double-handed signs for 36 ISL alphabets and digits. The authors segmented collected sign images using skin colour in YC_bC_r colour space and used ANN for the classification to achieve 91.11% accuracy. Subhash et al. [109] used MSVM for the classification of ten static and 11 dynamic words. The signer replicated the signs from available online examples that were played on a monitor and recorded using a digital camera [35]. They reported 95.3% recognition accuracy. Sahoo et al. [71] implemented an ISL alphabet recognition system. The authors extracted structural features, local histogram features, and direct pixel values of greyscale gestural images. KNN and neural network classifiers achieved a 95.30% (single-handed alphabet dataset) and 96.37% (double-handed alphabet dataset) recognition rate, respectively. A similar method is used by Sharma et al. [72] using centroid techniques and direct pixel value to extract features for ISL numbers static signs. They reported a recognition accuracy of 97.10% using MATLAB nprtool. Tripathi et al. [73] proposed a system for ISL sentence recognition with invariant backgrounds. They captured 500 samples for selected ten sentences, including single-handed and double-handed dynamic signs. Key frame extraction using discrete wavelet transform (DWT) is accomplished during the feature extraction process to speed up training and testing for the hidden Markov model (HMM). Signs are classified, with an overall accuracy of 91%. The authors tested the results obtained by the proposed model against several parameters such as Euclidean distance, Mahalanobis distance, City block distance, Chessboard distance, Cosine distance, and Correlation distance [123]. Their model obtained maximum classification accuracy with Euclidean distance and Correlation distance-based classifiers.

Kharate et al. [77] presented a comparative analysis over three classifiers (nearest mean classifier, K-nearest neighbour classifier, and Naive Bayes classifier) and feature descriptors like Fourier descriptor, 7 Hu moments, shape matrix, and chain code for the classification of alphabets and digits. The authors

TABLE 6 Machine learning based ISLR systems.

Reference	Year	Acquisition mode	# of hands	Static/ Dynamic	Signing mode	Recognition accuracy
Rekha et al. [111]	2011	Camera	2	Both	Isolated	SVM 86.3% (static) and DTW 77.2% (dynamic)
Lilha et al. [68]	2011	Camera	2	Static	Isolated	SVM 98.1%
Agrawal et al. [112]	2012	Camera	2	Static	Isolated	SVM 93%
Dixit et al. [70]	2013	—	Both	Static	Isolated	SVM 96.23%
Adithya et al. [113]	2013	Camera	Both	Static	Isolated	ANN 91.11%
Subhash et al. [109]	2014	Camera	Both	Both	Isolated	SVM 95.3%
Sahoo et al. [71]	2014	Camera	Both	Static	Isolated	ANN 95.30% (single handed alphabet) and ANN = 96.37%(double handed alphabet)
Sharma et al. [72]	2014	Camera	1	Static	Isolated	ANN 97.10%(numerals)
Tripathi et al. [73]	2015	Camera	Both	Both	Continuous	HMM 91%
Mehrotra et al. [74]	2015	Kinect sensor	2	Both	Isolated	SVM 86.16%
Gupta et al. [114]	2016	Camera	Both	Static	Isolated	KNN 90%
Kumar et al. [115]	2016	Mobile Camera	1	Dynamic	Isolated	ANN 90%
Kumar et al. [116]	2016	Camera	1	Both	Isolated	SVM 93% (static) and 100% (dynamic)
Kharate et al. [77]	2016	Web camera	Both	Static	Isolated	KNN 99.61%
Rokade et al. [60]	2017	Internet sources	Both	Static	Isolated	SVM 92.12%, ANN 94.37%
Kaur et al. [80]	2017	Camera	Both	Static	Isolated	Uniform background DHM 98.3%, KM 97.9% Complex background DHM 75.9%, KM 72.6%
Kumar et al. [81]	2017	Kinect sensor	Both	Dynamic	Isolated	Single-handed dynamic words HMM 81.29%, Double-handed dynamic words HMM 84.81%
Kumar et al. [117]	2017	Kinect sensor and leap motion	1	Dynamic	Isolated	Coupled HMM 90.80%
Dutta et al. [118]	2017	Camera	Both	Static	Isolated	ANN (Single handed alphabets 94.88%, double handed alphabet 95.84%)
Rao et al. [119]	2018	Mobile Camera	1	Dynamic	Isolated	ANN 90.00%
Rao et al. [120]	2018	Mobile Camera	1	Dynamic	Isolated	ANN 91%
Joshi et al. [121]	2018	Camera	Both	Static	Isolated	SVM 94.5%
Athira et al. [85]	2019	Web Camera	Both	Both	Isolated	Finger spelling alphabets SVM 91%. Single-handed dynamic words = 89%
Marippan et al. [86]	2019	Camera	Both	Both	Both	FCM 75%
Gangrade et al. [90]	2020	Kinect sensor	Both	Static	Isolated	SVM 88.24%
Adithya et al. [87]	2020	Camera	2	Static	Isolated	SVM 96.25%
Raghuveera et al. [89]	2020	Kinect sensor	Both	Both	Both	SVM 71.85%
Joshi et al. [122]	2020	Camera	Both	Static	Isolated	SVM 92%
Katoch et al. [97]	2022	Camera	Both	Static	Isolated	SVM 99.17%

observed that the fusion descriptors have a high rate of recognition (99.6%) because they incorporate shape information from contour and region-based descriptors. Gupta et al. [114] first divided static alphabet signs into single-handed and double-handed. Further, using a combination of HOG and SIFT descriptors, both the data are sent to a KNN classifier. Using HOG features alone, the authors reported 78.84% accuracy when tested on 26 signs. An accuracy of 80% was recorded for the same classification strategy applied to SIFT features, while the fusion of two types of descriptors yielded 90% accuracy.

Kumar et al. [115] proposed a CSLR system in which they used the front camera of a mobile phone for collecting signs. The authors created a formal database of 18 ISL signs with Ten different signers. Pre-filtering, segmentation, and contour detection are performed with Gaussian filtering, and Sobel with adaptive block thresholding and morphological subtraction, respectively. Hand and head contour energies are computed from discrete cosine transform. The authors used a feed-forward ANN classifier and achieved an accuracy of 90%. Kumar et al.'s algorithm [116] uses skin colour segmentation to extract alphabet signs from video sequences with less cluttered and dynamic backgrounds. Certain features of the gesture are obtained depending on whether it is static or dynamic. The authors detect the fingertips, by finding all the convexity defects of the hand using its contour and convex hull. Pre-trained SVM classifiers are used to categorize signs. Static and dynamic gesture recognition succeeded with approximately 93% and 100% accuracy.

To detect the hand shape, Rokade et al. [60] first applied a segmentation phase based on skin colour. Then, the detected region is transformed into a binary image. Next, the binary image goes under the Euclidean distance transformation, row and column projection. For feature extraction, central moments with HU moments are used. For alphabets classification, the authors compared ANN, and SVM varying the numbers of feature sets. The authors concluded that ANN (94.37%) gives better accuracy even with fewer feature set than SVM (92.12%). Kaur et al. [80] observed that most signs need both hands, which adds complexity. In ISL, some alphabets like ‘A’, ‘B’, ‘P’, ‘Q’, ‘U’, ‘W’, ‘M’, and ‘N’ have identical shapes. ‘M’, ‘N’, and ‘W’ have high occlusion of hands. Few, like ‘I’, and ‘K’ are the super-gesture of others. The authors demonstrated the performance of dual-Hahn moments, and Krawtchouk moments for the ISL database on uniform and complex backgrounds [93]. For the ISL uniform background database dual-Hahn moments, and Krawtchouk moments, achieved an accuracy of 98.3% and 97.9%, respectively. On complex backgrounds, dual-Hahn moments and Krawtchouk moments recorded respectively, with 75.9% and 72.6% accuracy. Kumar et al. [81] present a robust position invariant SLR framework. A Kinect sensor was used to obtain the signer’s skeleton information. The framework is capable of recognizing occluded sign gestures and has been tested on a dataset of 2700 gestures. The recognition process has been performed using HMM, and the results showed an efficiency of 83.77% on occluded gestures. Using coupled fusion using the coupled hidden Markov model (CHMM), Kumar et al. [117], achieved an accuracy of 90.80% in word recognition. CHMM is considered a special type of dynamic Bayesian network, where the architecture is capable enough to capture the asynchronous and temporal inter-modal dependencies between two different information channels. CHMM outperforms conventional HMM in recognition of gestures as CHMM is a collection of multiple HMMs, where one HMM corresponds to one data stream. Using a technique developed by Dutta et al. [118], first the background of the sign image was subtracted by RGB filtering and thresholding. Next, the authors performed edge detection, hand posture extraction, scale to a standard size, and compared the processed image with the database images using an error matrix. Using ANN, the accuracy for single-handed alphabets was 94.88% and the same for double-handed alphabets was 95.84%.

Joshi et al. [121] tested the effectiveness of the HOG descriptor and SVM classifier using the ISL uniform and complex background database [80] and attained 95% overall accuracy. In their further work, Joshi et al. [122], proposed that a combined TOPSIS-Taguchi-based decision-making technique is effective, in picking the right parameter combination to get high performance. The authors tuned the proposed/selected set of parameters for ISL Dataset Complex Background [80] and then tested it on the publicly available ISL dataset [80]. Their proposed multi-level HOG features set (a smaller set of 280) achieved an accuracy of 92% using SVM with poly-kernel. Rao et al. [119, 120] proposed a CSLR system to bring the usage of SL closer to real-time deployment on mobile platforms using a video ISL database created with a mobile front camera in selfie mode. They used discrete cosine transform (DCT) to extract features from the collected signs. Using ANN, the authors obtained an accuracy of 90%. Marippan et al. [86] proposed a model where authors used fuzzy c-means clustering (FCM). This model is trained on a dataset containing samples for 40 words and 25 sentences and reported 75% accuracy.

Gangrade et al. [90] proposed a technique for the ISL alphabets and numbers recognition using HOG + SVM. For a dataset with cluttered backgrounds, an accuracy of 88.24% was recorded. To evaluate the effectiveness of the Emergency Situations dataset, Adithya et al. [87] conducted experiments to show that an average accuracy of 96.25% using the SVM classifier is achieved. The authors suggest that the dataset can be useful in developing new tools and technologies to assist people in emergencies, such as a smart phone app that could recognize the gestures and provide appropriate assistance. Ensemble is the process of combining a diverse set of features (individual models) to improve the stability and predictive power of the system. Raghuveera et al. [89] combined all the predictions from the HOG, SURF (speeded up robust features), and local binary patterns (LBP) features together. The three most popular methods for combining the predictions from different models are bagging (building multiple models from different subsamples of the training dataset), boosting (building multiple models each of which learns to fix the prediction errors of a prior model in the chain) and stacking (building multiple models and supervise model that learns how to best combine the predictions of the primary models). Raghuveera et al. [89] recorded 71.85% accuracy using SVM.

Katoch et al. [97] used SURF, a local feature detector and descriptor. SURF is robust against rotation, variance, and point-of-view occlusion. It also provides operators with box filters for fast computation. The authors demonstrated with SVM and CNN classifiers for 36 signs (26 alphabets and ten numerals) and observed that the two classifiers produced comparable accuracy of more than 99%. Elakkiya [124] depicts the vital role of machine learning methods in the automatic recognition of sign languages and addresses the need for subunit sign modelling for continuous sign language.

Apart from ISL, few researchers contributed to developing regional ISL recognition systems. The authors of [58, 125, 126] proposed a system for recognizing Bangla sign language gestures using machine learning techniques. Amitkumar et al. [62, 63], presented a method for recognizing 43 Marathi signs consisting of vowels and consonants using computer vision techniques. The authors used pattern matching to classify the extracted features into different signs. The authors evaluated their method, achieving an accuracy of 90%. Sumaiya et al. [127] proposed a system for recognizing and translating Kannada sign language (KSL) gestures of alphabets using machine learning techniques. The authors collected a dataset of KSL gestures from deaf and HoH individuals and used it to train and test several machine-learning models, including SVM and KNN.

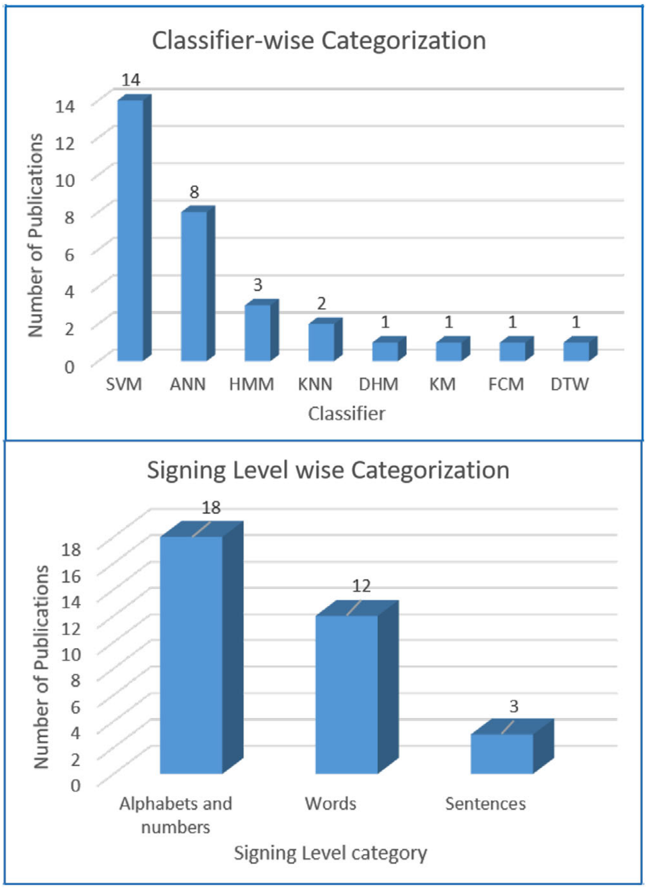

FIGURE 7 Analysis of machine learning based ISLRT systems: (a) Classifier wise categorization and (b) signing level wise categorization.

They achieved the highest accuracy, with a recognition rate of 98.90%.

7.1.1 | Discussion on the performance of machine learning based approach for SLR

Sign language recognition has three parts: Alphabet and number recognition, word recognition, and sentence interpretation. Alphabets and integers are usually one-or two-handed static signs in SL. Words are symbols with manual and non-manual gestures and movements. Sentences in SL are a combination of symbolic signs with grammar.

Figure 7 presents the analysis of machine learning-based ISLRT systems. After the analysis, we conclude:

- Support vector machine (SVM) is the most popular classifier among the researchers (achieving an accuracy above 99%).
- A significant work on isolated ISL signs (alphabets, and digits) recognition has been done as compared to word recognition and continuous (sentence) sign interpretation.
- A significant portion of the ML-based literature has concentrated solely on the manual aspects. Even though hands are predominant, signing allows signers to express far more information by combining them with non-manual aspects.
- There are several biases in the datasets that are currently available (training datasets are of small size, less number of signers).
- Only, Kaur et al. [80], Joshi et al. [121, 122] have used the common dataset (ISL complex background) to evaluate various SLR models. [75] used the extended dataset of the dataset created by [89]. Apart from this, other researchers have created their datasets and haven't made them public, so their work transparently can not be compared with others.

Significant work on isolated ISL sign (alphabet, digits, and words) recognition has been done as compared to continuous (sentence) sign interpretation. Continuous sentence interpretation is a challenging task and it is essential to work now in this domain. From the ISL literature, we found three significant works that use a machine-learning approach for continuous sentence interpretation. We compare them in Table 7.

The detection and segmentation of hands from images with complex background conditions, by different people is a problem addressed by sign recognition. Another major challenge machine learning techniques face is accurately interpreting patterns from the raw input data. So, few researchers have used an ensemble of feature extraction techniques [89] resulting in better accuracy.

7.2 | Deep learning based approach for sign language recognition

Recently emerged deep learning techniques outperform the traditional approach to hand gestures and ISL recognition (Table 8). It eliminates the need for standard pre-processing, segmentation, and complex hand-crafted feature extraction from images. For instance, CNNs automate the feature extraction step by learning the high-level abstractions in the image and capturing the most discriminative feature values employing a hierarchical architecture [128].

Deep learning models such as LSTMs, GRUs [15, 17, 79, 129, 130], and Transformers [19, 131, 132] are specifically designed to handle sequential data by modelling temporal dependencies, making them indispensable for CSLRT. Continuous sign language comprises a stream of gestures and movements where each gesture's meaning is influenced by the context provided by preceding and subsequent gestures. Unlike isolated gesture recognition, this task requires a comprehensive understanding of temporal relationships to construct meaningful sentences. These models excel at extracting patterns and relationships within temporal data. LSTMs and GRUs, as types of RNNs, process sequences incrementally, retaining critical information across time through hidden states. LSTMs use gates (input, forget, and output gates) to control the flow of information through time steps and GRUs use a simpler gate mechanism than LSTMs by combining the forget and input gates into an update gate. The gate mechanism in LSTMs and GRUs enables them to selectively remember essential features and forget irrelevant ones, ensuring an effective understanding of the gesture sequence. On the other hand, transformers

TABLE 7 Comparison of machine learning based sentence interpretation ISL works.

Reference	Methodology used	%Accuracy	Advantages	Limitations
Tripathi et al. [73]	• Hand segmentation • Key frame extraction using the gradient method • Extract feature using orientation histogram and variance (DWT) • Classification using HMM	91%	• System works with dynamic backgrounds • Pioneer work for continuous ISL translation with good accuracy	Small size of the dataset: only ten sentences, five signers
Marippan et al. [86]	• Depict the three contours/ROI covering the face and both hands • Extract the features such as the number of points in the convex hull, number of defect points, and distance from the centre to each finger • Classification using FCM	75%	• System works with dynamic backgrounds • Sufficiently larger dataset: 50 sentences, ten signers	FCM requires more computation time
Raghuveera et al. [89]	• Hand segmentation using K-mean clustering • Extraction of three (SURF, HOG, and LBP) features • Prediction of output label using ensembling • Sentence interpretation using inverted index • Classification using SVM	71.85%	System can recognize alphabets, digits, words, and sign sentences constituting 140 ISL gestures (Alphabets, digits, and words) for 21 signers	By using Microsoft Kinect, the system can maintain its performance independent of external factors, but this makes system impractical for real-time applications

TABLE 8 Deep learning based ISL recognition systems

Reference	Year	Acquisition mode	# of hands	Static/Dynamic	Signing mode	Recognition accuracy
Dour et al. [76]	2016	Web Camera	2	Static	Isolated	ANFIS 75.39%
Prasad et al. [78]	2016	—	Both	Static	Isolated	BPNN 92.34%
Kumar et al. [79]	2017	Kinect and leap motion	Both	Dynamic	Isolated	HMM and BLSTM 95.60%
Rao et al. [83]	2018	Mobile camera	1	—	Isolated	CNN 92.88%
Kaur et al. [133]	2019	Camera	1	Static	Isolated	FFBPNN 99.43%
Ravi et al. [84]	2019	Kinect sensor	Both	Dynamic	Isolated	CoT4CNN 86.66%
Wadhawan et al. [88]	2020	Web Camera	Both	Static	Isolated	CNN 99.72% (coloured) and 99.90% (Greyscale)
Gangrade et al. [90]	2020	Kinect sensor	Both	Static	Isolated	CNN 99.93%
Kumar et al. [67]	2020	MOCAP	Both	Dynamic	Isolated	2 stream CNN (2CNN) 92.14%
Kinjal et al. [16]	2020	Camera	Both	Both	Isolated	MobileNetV2 + RNN 99.57 %
Sharma A. et al. [91]	2020	Camera	Both	Static	Isolated	SVM-VGG16 One hand 98.52%, Two hand 97%
Marippan et al. [15]	2021	Camera	Both	Dynamic	Isolated	CNN + LSTM-RNN 95.99%
Santhosh et al. [45]	2021	Camera	—	Both	Isolated	CNN + LSTM-RNN 96%
Kumar A. et al. [46]	2021	Web Camera	Both	Static	Isolated	ELM-NN 80.76%
Varsha et al. [134]	2021	Camera	Both	Both	Isolated	Inception V3 93%
Sharma S.et al. [93]	2021	Camera	Both	Static	Isolated	G-CNN 94.83%(alphabets), 99.96%(words)
Sharma S. et al. [94]	2022	Camera	Both	Static	Isolated	CNN 92.43%
Natarajan et al. [129]	2022	Camera and web camera	Both	Both	Continuous	Hybrid CNN-LSTM 98.56%
Katoch et al. [97]	2022	Camera	Both	Static	Isolated	CNN 99.64%
Dumbre et al. [106]	2022	Camera	Both	Both	Isolated	CNN 99.90%
Areeb et al. [130]	2022	Camera	Both	Static	Isolated	Pre-trained VGG-16+LSTM 98%
Prasath et al. [135]	2023	Camera	Both	Both	Isolated	MLCNN 87.50 %
Rajalakshmi et al. [98]	2023	Camera	Both	Both	Isolated	hDNN 99.17%

leverage self-attention mechanisms to process all sequence elements simultaneously, capturing both short and long-range dependencies with exceptional efficiency. In LSTMs, and GRUs sequential computation makes parallelism difficult. On the contrary, Transformers are highly parallelizable.

Dour et al. [76] proposed a method for recognizing the ISL alphabets using adaptive neuro-fuzzy inference systems (ANFIS) and Sugeno fuzzy neural network. The authors presented an approach involving the segmentation of hand gestures and the extraction of relevant features. A 1 × 64 matrix vector was obtained by dividing the input gesture image into 64 subblocks, followed by the proposed fuzzy neural network to recognize the sign language alphabets with an average accuracy of 75.39%. Prasad et al. [78] proposed an approach that first extracts the edges of the hand region using a Canny edge detector and then further processes these edges using a fusion-based operator that combines the results of four different edge detection algorithms. The resulting edge map is used to extract features for word sign recognition. Using a back propagation neural network (BPNN), an accuracy of 92.34% was reported. Kumar et al. [79] suggested a multimodal framework for SLR systems that uses a Kinect sensor and Leap motion. The authors collected a dataset consisting of 50 dynamic ISL words. Both of the devices simultaneously recorded these signs. The authors achieved 95.60% accuracy using fused features and an HMM-BLSTM (bidirectional long short-term memory) classifier combination. Rao et al. [83] demonstrated their work using CNN, for 200 ISL words and recorded an accuracy of 92.88%. They recorded 60,000 images using a mobile camera.

Kaur et al. [133] created their dataset for alphabets and digits with 130 subjects (https://github.com/jasminek247/SLR/tree/master). The authors used SIFT as a descriptor. It extracted the features that train the feed forward back propagation neural network (FFBPNN) reaching alphabet and digit classification accuracy of 99.43%. Patil et al. [136], used a similar approach to classify 26 ISL alphabets. Their work mainly focused on analyzing the time required for implementing various SIFT algorithm phases.

Ravi et al. [84], proposed multi-modal spatio temporal co-trained CNNs (CoT4CNN) and tested them on a self-created BVCSL3D dataset having 200 ISL word-sign classes. The authors estimated the performance of the proposed architecture against different spatiotemporal fusion strategies like product fusion, sum fusion, PCA fusion, concatenation, and averaging fusion. The average recognition with cross-subject testing on the ISL dataset was 86.66%.

Wadhawan et al. [88] deal with the modelling of static SLR using CNN. In their work, a total of 35,000 sign images of 100 static signs are collected from different users. The proposed system is evaluated on approximately 50 CNN models using various optimizers. The authors achieved the highest training accuracy of 99.72% and 99.90% on colour, and greyscale images, respectively. A technique for the ISL alphabets and numbers recognition utilizing a five-layer CNN model was proposed by Gangrade et al. [90] to achieve an average accuracy of 99.93%.

Kumar et al. [67] proposed a method for recognizing 3D sign language words using a two-stream CNN and achieved an accuracy of 92.14% for the ISL3D dataset. This method extracts spatiotemporal features from the depth and colour information using a joint distance topographic descriptor (JDTD) and joint angle topographical descriptor (JATD), which encodes distance and angular information in a colour image captured with the MOCAP system. In their further work [137], the authors built six CNNs using the complex multi-channel CNN architecture. They trained them using images via joint angular displacement maps (JADM) and joint distance maps (JDM). The authors compared the performance of the proposed model against the baseline models by using the ISL3D dataset and reported 92% performance.

According to Kinjal et al. [16], a process that uses frame sequence generator and image augmentation techniques enhances dataset size and decreases overfitting. SLR performance is greatly enhanced by employing basic image editing techniques and batches of shifted frames of videos. The authors achieved 99.57% accuracy in the classification of 21 ISL words using MobileNetV2 + RNN.

To recognize ISL gestures, Sharma et al. [91] carried out an analytical evaluation of three deep learning-based approaches (the pre-trained VGG16 model, the natural language-based output network, and the hierarchical network). The accuracy of the hierarchical model is 98.52% for one-hand gestures and 97% for two-hand gestures, significantly outperforming the other two models. Occlusions arise by using a single hand in some gestures and both hands in other gestures. To address this issue, their model first determines the number of hands in the input gesture.

Marippan et al. [15] proposed a model that uses a combination of CNN and LSTM-RNN to get increased recognition performance. The model is trained with their proposed CasTalk-ISL dataset of 100 words and achieved 95.99% accuracy. The author's goal product is to translate hand gestures into speech. A similar approach was taken by Santhosh et al. [45] to achieve 96% accuracy for only isolated signs. Kumar et al. [46] proposed recognizing an extreme learning machine (ELM) learning algorithm to train neural networks for the ISL alphabets. In ELM, hidden nodes are chosen randomly and never updated. The authors reported an average accuracy of 80.76%. Varsha et al. [134] implemented an ISL word recognition model using deep CNN (Inception V3 model) and achieved an accuracy of 93%. This project used the IIITA-ROBITA ISL gesture database [65]. For the system implementation, the authors have used 20 classes from the dataset. The images in the dataset are reprocessed and then passed into the inception model. Sharma et al. [93] collected a dataset from multiple signers under different light and background conditions (uniform and complex). This dataset consists of RGB images of 43 classes of ISL performed by 50 signers, resulting in 2150 gesture images. This dataset is subdivided into alphabets and isolated words. The proposed model is termed Gesture CNN (G-CNN). Accuracy of 94.83%, 93.60%, and 93% are attained for alphabet recognition using G-CNN, VGG-11, and VGG-16, respectively. Accuracy of 99.96%, 97.87%, and 97% are obtained for isolated word recognition using G-CNN, VGG-11, and VGG-16, respectively.

Sharma et al. [94] implemented a system for ISL gesture recognition using CNN. The authors evaluated the proposed work on a dataset of 26 ISL alphabet gestures and achieved 92.43% accuracy. Natarajan et al. [129] implemented a deep learning framework that is able to recognize and translate SL in real-time and demonstrated it on ISL-CSLTR dataset [8]. The framework consisted of three main components: SL recognition, translation, and video generation. The SL recognition and translation components use the MediaPipe library and a hybrid CNN + bi-directional long short term memory (CNN + Bi-LSTM) model. The video generation component uses a generative adversarial network (GAN) to generate videos of a person signing the translated text. Using an emergency ISL dataset [87], Areeb et al. [130] suggest classification and detection models. The detection model is based on the YOLO v5, whereas the optimal classification model combines pre-trained VGG-16 and LSTM. The classification model's accuracy of 98% was recorded. The focus of Prasath et al.'s [135] research is modelling a multi-layered CNN (ML-CNN) with an encoder where the features are learned to increase prediction accuracy. The predicted ML-CNN performs effectively when applied over three diverse datasets known as DEVISIGN, SLR, and ROBITA dataset [65] where the hand gestures are predicted with an accuracy of 87.50%. For the recognition of static Indian signs, the authors Kothadiya et al. [131] have implemented a vision transformer with 99.29 % accuracy. The dataset used in the simulation was prepared from the collection of publicly available Indian sign language dataset (static) [138], which includes gestures of numbers (0-9) and the alphabet. The dataset consists of RGB images of a total of 36 classes with more than 1000 images per class. Rajalakshmi et al. [98] used a hybrid deep neural net (hDNN). Multi-semantic feature extraction unit comprises the manual articulation tracking (MAT) and non-manual element tracking (NMET) submodules for feature detection and the spatial component detection (SCD) and sequential-temporal component extraction (STCE) submodules for feature learning. Semantic manual articulation tracking is done in the MAT submodule. This module uses the sign frame images to track the hand, palm, and entire pose skeleton using Holistic MediaPipe. From the sign gestures, the Holistic Mediapipe generates landmark pixel coordinates and skeleton posture. The SCD module then receives the skeletal posture as input to extract spatial characteristics. To extract the temporal and sequential components, the landmark coordinates produced by the Holistic MediaPipe are passed to the STCE component. The authors compared the proposed framework with other baseline models using the word-level American sign language (WLASL) database to evaluate the scalability of the proposed model to record 99.87% accuracy on the ISL dataset.

7.2.1 | Discussion on the performance of deep learning based approach for SLR

Deep learning has seen remarkable advancements in recent years and has revolutionized the ISL field, but it also comes with challenges, such as the need for substantial computing resources and a large amount of labelled data. Figure 8 presents the analysis of deep learning-based ISLRT systems. After the analysis, we conclude:

- A significant work on isolated ISL sign (alphabet, digits, and words) recognition has been done as compared to continuous (sentence) sign interpretation (only [129]).
- A significant portion of the DL-based literature has concentrated on the manual and non-manual aspects.

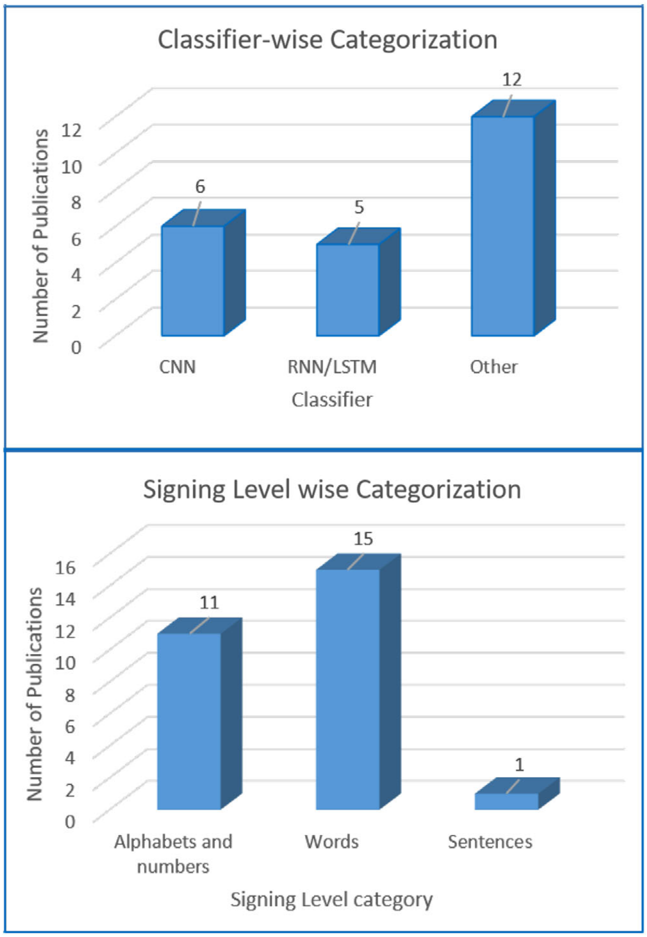

FIGURE 8 Analysis of deep learning based ISLRT systems: (a) Classifier wise categorization and b) signing level wise categorization.

- There are several biases in the datasets currently available (training datasets are of small size and less number of signers).
- Only, [134, 135] have used the common dataset (IIITA-ROBITA ISL Gesture Database) to evaluate various SLR models. Apart from this, other researchers have created their custom datasets for the evaluation of SLR systems.

7.3 | Motion tracking based approach for sign language recognition and translation

Motion tracking, also known as motion capture, is a technology used to monitor and record the movement of objects or individuals in a three-dimensional space. It involves capturing the position, orientation, and often the velocity of objects as they move over time. Motion tracking is being widely used in ISL scientific research because every signer while presenting the sign can have different velocities. The primary areas of study in automatic sign language recognition have concentrated on the recognition of isolated signs (Table 9). In isolated sign recognition, static or dynamic single signs are recognized without the continuation of any other sign. In continuous sign recognition, different signs are performed one after another to

TABLE 9 Motion tracking based ISL recognition and interpretation systems.

Reference	Year	Acquisition mode	# of hands	Static/ Dynamic	Signing mode	Sign categories	Recognition accuracy
Kishore et al. [139]	2016	Camera	Both	Dynamic	Isolated	(58 ISL words)	BPNN 90.17%
Kishore et al. [82]	2018	MOCAP	Both	Both	Isolated	(500 ISL words)	Tracking-based, Motionlets-based 98.9%
Shenoy et al. [140]	2018	Camera	Both	Dynamic	Continuous	(5 ISL words, 7 ISL sentences)	HMM 97.23%
Mistree et al. [92]	2021	Camera	Both	Dynamic	Continuous	(55 ISL sentences)	MobilNetV2 94%

recognize sign language words or sentences [33]. Bhuyan et al. [141] proposed a model for dynamic hand gesture recognition systems that incorporated the development of SLR systems. It is based on finite state representation and gesture summarization (object-based video abstraction) using key video object planes (VOPs). A hand is considered a video object (VO). For the moving hand, a binary model is used to track in subsequent frames. The Hausdorff distance measure is used to select key VOPs, to transform an entire video sequence into a small number of candidate frames that are sufficient to represent a particular gesture sequence. For gesture recognition, a finite state machine (FSM) consisting of a finite number of keyframes with respective keyframe durations is used. Kishore et al. [139] presented a system for recognizing ISL sentences. The dataset comprised of sentences formed from 58 words. For hand tracking Horn Schunck optical flow algorithm is used. Shape features are extracted in each frame with an active contour level set model. Horn Schunck optical flow is the tracker for hands in the video frames. The algorithm uses two consecutive frames and computes the velocity vectors in the x and y directions. From these vectors, the position of the hand in the frame is computed. Shape features are used to train the backpropagation neural network. The identified signs are then mapped to text and further converted to voice commands with a Windows text-to-speech application programmable interface. They achieved a recognition rate of 90.17%. The sign recognition experiment of Nanivadekar et al. [59] includes A to Z, numerals 1 to 10, and a few phrases that are important in everyday conversation and an emergency. The authors introduced one lefty signer and one signer with an extra thumb to get variations. Their sign recognition algorithm takes into consideration motion tracking by implementing a difference image algorithm, pattern recognition by employing an edge detection algorithm, and hand tracking by employing a skin colour-based model. The authors have not mentioned recognition accuracy. Elakkiya et al. [142] suggested a novel subunit sign modelling methodology, to address uncertainties during the segmentation of the hand and identify epenthesis movement in lengthy video sequences. Their work uses clustering techniques based on spatial and temporal data along with enhanced dynamic programming and dynamic time warping. Here, epenthesis movement identification and subunit modelling of the lengthy video sequence are carried out using subunit multi-stream parallel HMM (SMP-HMMs) and minimum entropy clustering. Elakkiya et al. [143] present a new framework for continuous sign language recognition using subunit sign modelling. The proposed framework aims to address this issue by breaking down the sign gestures into subunits and modelling them individually. Their proposed framework consists of three main stages: subunit sign extraction, subunit sign modelling, and continuous sign recognition. In the first stage, the sign gestures are segmented into subunits using a combination of hand shape detection and motion detection. In the second stage, each subunit is modelled using HMMs, which are commonly used for modelling temporal data. In the third stage, the subunits are combined to recognize continuous sign gestures using subunit multi-stream parallel HMM (SMP-HMM). The authors of [142, 143] have not evaluated the proposed techniques using any ISL datasets. Kishore et al. [82] contributed Motionlet matching with adaptive kernels. In this work, phase 1 deals with hand tracking. Initially, each frame is segmented to produce motion joints and non-motion joints. After extraction of motion joints, it is categorized into one of four identified classes. Phase 2 deals with motionlet which is intra-finger variations. Joint angle measurements and finger joint relative distance are used for extracting the shape and orientation of 3D motionlet. Three feature kernels based on finger shape orientation and trajectories are constructed for each sign to compare the similarity between a query sign and a database sign. For sign recognition, they assert a 98.9% accuracy.

Shenoy et al. [140] classified 33 ISL hand poses with an accuracy of 99.7%. The system is able to classify 12 gestures with an average accuracy of 97.23%. The approach uses an HMM chain for gesture recognition and a k-NN model to classify hand pose. The model proposed by Mistree et al. citec12 uses a pretrained model with the INSIGNVID ISL video dataset. Using this model the videos of signs for 55 most frequently used sentences are converted to English sentences. The green background is used while recording the video to make background removal an insignificant task.

7.4 | Graph-based approach for sign language recognition

Graphs can be utilized to represent the gestures at various states in sign sequences. Kumar et al. [144] proposed a method that uses a graph-based approach to recognize sign language gestures. In this approach, each sign gesture is represented as a graph, where 57 nodes represent the joints of the hand and 56 edges represent the connections between them. The graphs are then compared with a pre-defined set of gesture graphs using graph matching (GM) techniques that divide the recognition process into two stages. In the first stage, the system makes an early estimation of the gesture from the input data using a simplified graph representation. If the estimated gesture is not a match, the system proceeds to the second stage, which uses the full graph representation of the input data for more accurate recognition. The authors recorded 98% recognition accuracy.

7.5 | Geometric estimation based approach for sign language recognition

Nikam et al. [145] in convexity hull algorithm initially computes maximum and minimum x and y coordinate points and by joining those points forms a bounding rectangle that contains the hull. Like hulls there are other points, convex defects of the hand, which are present in between the valley of two fingers. Further, by taking an average of all such defect points, we get a centre of the palm. So, the radius of the palm is considered as the depth of the palm. The ratio of palm radius and distance of the hull point from the centre point of the palm is used to determine the finger opening and closing position. accuracy is not mentioned. A similar approach by Hussain et al. [146] uses palm and fingertip position estimation based on hand contour. Geetha et al. [147] approach for identifying ISL relies on piece-wise polynomial functions to approximately model contours/surfaces using a sparse set of control points. The authors used B-Spline approximation [148] for shape matching of ISL alphabets and numerals. To obtain the control points (curvature points), boundary tracing of the sign gesture is done. The maximum curvature points contributed to the shape of the gesture. The 2D space of the gesture is then divided into eight octants. The feature vector has a set of eight values each corresponding to the count of maximum curvature points (MCP). Lastly, the SVM classifier was used for recognition. The designed system achieved a recognition rate of about 80% for the static alphabets and numerals. Athira et al. [85] use a method for co-articulation elimination in finger spelling for both static and dynamic alphabets. According to the authors, the static and dynamic gestures from the collection of input frames can be separated by their centroid changes between frames. Based on the lack of centroid change between consecutive frames during the signing phase, the static gestures are detected (gradient of acceleration approach). A static gesture is one in which the centroid does not move after N frames. Otherwise, it is classified as either a dynamic gesture or a co-articulation region. The authors tested the proposed work under constant background and proper illumination conditions. Using SVM, for classification, they recognized alphabets with an accuracy of 91% and single-handed dynamic words with 89% accuracy.

Apart from this, there are also other approaches. For instance, Subha et al. [149] proposed a method that provides the conversion of a set of $31(2^5 - 1)$ binary static and dynamic images into decimal numbers by using a binary-to-decimal conversion algorithm and then the relevant Tamil letters for the decimal numbers are generated. Subramanian et al. [96] proposed a MediaPipe-optimized gated recurrent unit (MOPGRU) model. They upgraded the gate of the standard GRU cell by multiplying its output by the reset gate. With an improved update gate mechanism, the output of the reset gate rescreens the information. It eliminates the undesired information in the data (1662 landmark key points extracted from the frames), thereby giving more attention to the important information. Their suggested MOPGRU captured the full information dependency in time series data with an average recognition of 95% and a faster convergence speed. Ghotkar and Kharate [150] developed a rule-based and DTW-based method for ISL word recognition. The DTW-based method gives better accuracy than the rule-based method.

7.6 | End-to-end learning based approach for sign language recognition and translation

The introduction of neural machine translation (NMT) in 2014 marked a significant shift from traditional statistical methods to neural network-based approaches. NMT directly translates sequences, such as sentences, without relying on intermediate representations or manual feature extraction. It leverages encoder–decoder architectures (Figure 11) where the encoder converts input data into a context vector, and the decoder generates the translated output. This method, enhanced by the use of word embeddings, encodes both semantic and syntactic information, allowing models to understand and process word relationships more effectively. Attention mechanisms, which enable the model to focus on relevant parts of the input during translation, further enhance contextual accuracy. Without gloss attention models (Figure 11a), directly translate from input sign videos to text output without intermediate gloss representation. Models with gloss attention (Figure 11b), explicitly translate input sign videos into glosses as an intermediate step before generating text output.

NMT models rely heavily on large, high-quality parallel datasets, with the quality and size of the data being critical factors influencing performance. The advent of Transformer models [19] in 2017 revolutionized NMT by introducing self-attention mechanisms, which capture global dependencies in sequences while allowing for parallelization, thus improving efficiency and scalability. Transformers excel at handling long-range dependencies and variable-length sequences, making them particularly effective for rare or unseen phrases. Camgoz et al. [19] approached the task as a spatiotemporal NMT problem, integrating temporal dynamics and spatial cues using the PHOENIX14T dataset. Recent work by Gogoi et al. [132] introduces an End-to-End Sign Transformer Network for translating Indian sign language (ISL) sentences into text using ISL-CSLTR dataset [8], and INCLUDE50 dataset [66]. The authors implement the joint end-to-end recognition and translation approach introduced by Camgoz et al. [19] and achieve a baseline 15.50 BLEU-4 score on the combined dataset of ISL-CSLTR and INCLUDE50. To the best of our knowledge, this is the first study to demonstrate the use of a transformer for interpreting unconstrained ISL sentences.

TABLE 10 Comparison of machine learning and deep learning in sign language recognition and translation.

Key points	Machine learning (ML)	Deep learning (DL)
Feature extraction	Requires manual extraction of features like hand shape, motion, and position.	Automatically extracts features from raw video or images.
Model complexity	Utilizes simpler models like HMMs, SVMs, or ANN for gesture recognition.	Employs complex models like CNNs for spatial features and RNNs/LSTMs, or Transformer for temporal sequences.
Data requirement	Can work with smaller, labelled datasets; often needs pre-processing.	Requires large, diverse datasets for training to learn complex patterns effectively.
Performance	Effective for recognizing predefined gestures with limited variability.	Excels in handling complex, continuous sign language sequences with higher variability.
Training time and computational power	Generally faster to train due to simpler models and has lower computational needs; can run on standard hardware.	Longer training time due to deep architectures and large datasets, with high computational needs, often requiring GPUs.
Application areas	Suitable for basic gesture recognition tasks, such as recognizing individual signs or isolated words from a limited set.	Ideal for continuous sign language recognition, translation, and generation tasks, such as translating full sentences in sign language to text or speech.

7.7 | Summarizing notable key studies within the ISL domain

This section summarizes notable empirical results from key studies within the domain. Machine learning and deep learning play vital roles in sign language translation due to their ability to handle the intricate and multimodal characteristics of sign languages. ML depends on manually crafted features, such as motion vectors, angles, or histograms, requiring significant domain knowledge. On the contrary, DL automatically extracts features directly from raw data, minimizing the need for manual feature design and excelling at identifying patterns in large, high-dimensional datasets. Table 10 compares machine learning and deep learning approaches, summarizing their advantages and limitations, emphasizing distinctions such as feature extraction, model complexity, data requirements, performance, training time and computational power, and their applicability in SLRT areas.

We found that numerous techniques have been reported for ISL alphabets and digit recognition. In 2011, Rekha achieved over 86.3% accuracy for static and dynamic signs, and Lilha et al. [68] achieved over 98% accuracy for only static 26 ISL double-handed alphabet signs using SVM. Dixit et al. reported 96.23% accuracy with Hu invariant moments and MSVM for static word sign recognition in 2013. Subhash et al. achieved an increased accuracy of 95.3% for static and dynamic words using MSVM in 2015. The work by Tripathi et al. [73] is the first in the ISL domain to translate limited ISL sentences into English. The authors achieved 91% accuracy for ISL sentences using an HMM and multiple distance measures.

From 2015 to 2019 few researchers demonstrated collecting sign images from various sources like internet sources [74], using Mobile Camera, [83, 119, 120], using Web camera [76, 77], using Kinect sensors [60, 79, 84], and superimposing the signs on complex backgrounds [80] to generalize their proposed models for dynamic word gesture recognition, resulting into an average accuracy of around 80%. Concurrently Marippan et al. [86], and Raghuveera et al. [89] progressed in ISL sentence translation to achieve an average accuracy of 75%. In 2021 CNN and its variants like RNN, and LSTM achieved an accuracy of 96% for isolated dynamic word signs [15, 45]. In 2022, Natarajan et al. [129] implemented a hybrid deep learning framework to recognize and translate smaller-length ISL sentences using the ISL-CSLTR dataset [8] achieving an accuracy of 98.56%. Currently, few researchers Kothadiya et al. [131] using a vision transformer and Rajalakshmi et al. [98] using a hybrid Deep Neural Net (hDNN) with MediaPipe-based feature extraction have achieved almost 100% accuracy for isolated ISL signs. Recently, a few researchers Mistree et al. [9], Joshi et al. [103], Gogoi et al. [132] created larger ISL continuous sign translation datasets as transformers and their variants need datasets of larger sizes.

8 | CONCLUSIONS AND FUTURE DIRECTIONS

Recent developments in deep learning and computer vision have produced encouraging breakthroughs in several motion detection and gesture recognition techniques. The field of human-computer interaction has made significant progress recently, leading to the development of sign language recognition. A real-time automatic system for recognizing SL will be a priceless gift for HoH people, enabling them to communicate with the spoken language people easily. Despite the accomplishment of numerous research projects in the SLR field, the need for creating a real-time automated system persists [151]. In this study, we consider the literature on ISL datasets, ISL recognition, and ISL translation since 2011. Figure 9 presents the categorization of ISL works as static/dynamic sign and isolated/continuous sign. Figure 10 presents yearly progress in ISL research for two major contributors in the literature that is, ML using computer vision and DL using computer vision. From this visual analysis, we conclude:

1. Significant work on isolated ISL sign recognition has been done (92.59%) as compared to continuous sign interpretation (7.41%).

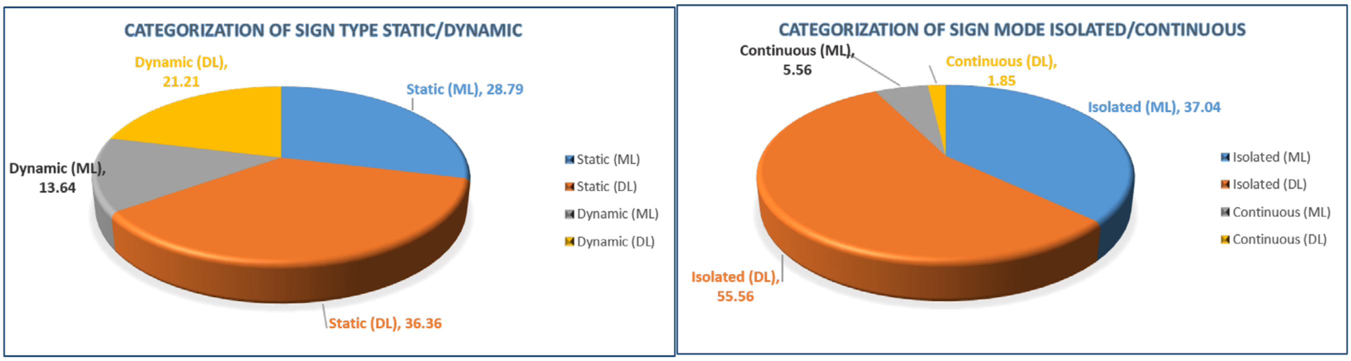

FIGURE 9 (a) Categorization of ISL works as static or dynamic sign. (b) Categorization of ISL works as isolated or continuous sign (Number of published papers using ML/DL and computer vision = 52).

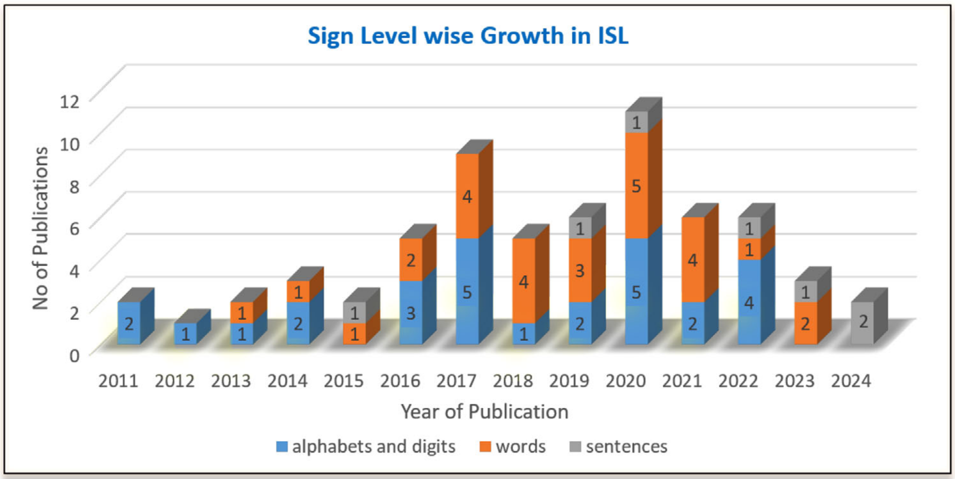

FIGURE 10 Sign-level wise yearly growth in ISL (Number of published papers using ML/DL and computer vision =52)

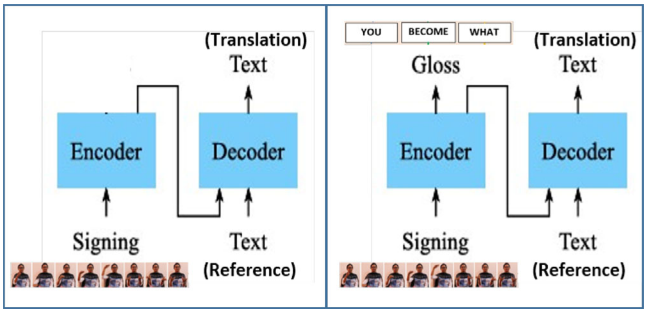

FIGURE 11 End-to-end sign language translation: (a) without gloss attention and (b) with gloss attention.

2. Many researchers have worked on static ISL sign recognition (65.15%) as compared to dynamic sign recognition (34.85%).
3. For continuous sign interpretation using DL (accuracy near 98% using Hybrid CNN-LSTM [129]) is found to be a better approach than ML (accuracy near 91% using HMM [73]).

Apart from these conclusions, we found:

4. ISL recognition is a challenging problem for researchers, due to its complex and dynamic nature and the variations in signing style, lighting conditions, background conditions, and camera viewpoints [152].
5. Few datasets and dictionaries are available for the isolated static/dynamic signs. Perhaps, many authors created their custom datasets and evaluated their SLR systems using the same.
6. Constrained small datasets are available for sentence-level ISL interpretation. These datasets include a limited set of sentences and are useful if ISLR is considered a multiclass classification problem.
7. For sentence-level translation, interpreted data are used, and spoken language influence is included in the dataset. Just native signer data are used, so most are of the same ethnicity. Hence, SLRT systems can not reach high quality and generalization.

8.1 | Future directions

This work summarizes the development of ISL datasets and dictionaries since 2011. It is observed that isolated static sign recognition is much of the researchers' interesting domain, and as compared to continuous SL interpretation, significant work is done in the isolated SLR domain. The research in continuous domain lags, as very few datasets are available. Therefore, the availability of high-quality datasets, similar to PHOENIX14T (RWTH-PHOENIX-Weather-2014T) [19] for the development and evaluation of ISLRT systems is essential. ISH News can be the right source to curate such a dataset. ISH News channel has various categories for instance business, sports, etc. along with their timestamped captions. The availability of unconstrained large datasets for ISL can help the research community look at it as a generative problem. Using the signs demonstrated on the ISL dictionary websites the custom datasets can be created for sentence-level ISL recognition.

Despite the recent advances in ISLR, several challenges need to be faced. One of the main challenges is the need for accurate and consistent annotation of the signs, which requires the involvement of expert signers and careful quality control to ensure the annotations' accuracy. In addition, there is a need for more user-friendly ISL recognition systems, that easily get integrated into existing communication platforms.

AUTHOR CONTRIBUTIONS

A. conceptualized the survey and designed the methodology, conducted the literature review and analysis, and wrote the initial draft. A. B. contributed to the revision and editing. A. prepared the figures and tables. B. supervised the project. Both authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors extend their gratitude to Mr. Mohsin Tigala, an employee of Reliance SMART Bazaar in Nagpur, Maharashtra, India, for assisting them in locating ISL resources beyond academic databases. His Unique Disability Identity number is MH1110420000093638.

FUNDING INFORMATION

The authors declare that they have received no funding for the current study.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated during the current study.

ORCID

Rina Damdoo https://orcid.org/0000-0002-7182-9516

REFERENCES

1. Ahmed, M.A., Zaidan, B.B., Zaidan, A.A., Salih, M.M., Lakulu, b.M.M.: A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017. Sensors 18(7), 2208 (2018). https://doi.org/10.3390/s18072208
2. Adrián Núñez-Marcos, G.L.: A survey on sign language machine translation. Expert Syst. Appl. 213, 118993 (2023). https://doi.org/10.1016/j.eswa.2022.118993
3. Bhatia, P., Verma, S., Kaur, S.: Sign language generation system based on Indian sign language grammar. ACM Trans. Asian Lang. Inf. Process. 19(4), 1–26 (2020). https://doi.org/10.1145/3384202
4. Safeel, M., Sukumar, T., Shashank, K.S., Arman, M.D., Shashidhar, R., Puneeth, S.B.: Sign language recognition techniques- A review. In: Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON), pp. 1–9. IEEE, Piscataway, NJ (2020).
5. Adeyanju, I., Bello, O., Adegboye, M.: Machine learning methods for sign language recognition: A critical review and analysis. Intell. Syst. Appl. 12, 200056 (2021). https://doi.org/10.1016/j.iswa.2021.200056
6. Kumar, S.S., Wangyal, T., Saboo, V., Srinath, R.: Time series neural networks for real time sign language translation. In: Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 243–248. IEEE, Piscataway, NJ (2018)
7. Goyal, L., Goyal, V.: Text to sign language translation system:: A review of literature. Int. J. Synth. Emotions 7, 62–77 (2016). https://doi.org/10.4018/IJSE.2016070104
8. Elakkiya, R., Natarajan, B.: ISL-CSLTR: Indian sign language dataset for continuous sign language translation and recognition. Mendeley Data, V1 (2021). https://doi.org/10.17632/kcmpdxky7p.1
9. Mistree, K., Thakor, D., Bhatt, B.: An approach based on deep learning for Indian sign language translation. Int. J. Intell. Comput. Cybern. 16(3), 397–419 (2022). https://doi.org/10.1108/IJICC-08-2022-0227
10. Rastgoo, R., Kiani, K., Escalera, S.: Sign language recognition: A deep survey. Expert Syst. Appl. 164, 113794 (2021)
11. Cheok, M.J., Omar, Z.B., Jaward, M.H.: A review of hand gesture and sign language recognition techniques. Int. J. Mach. Learn. Cybern. 10, 131–153 (2017)
12. Jiang, X., Satapathy, S., Yang, L., Wang, S.H., Zhang, Y.: A survey on artificial intelligence in Chinese sign language recognition. Arabian J. Sci. Eng. 45, 9859–9894 (2020). https://doi.org/10.1007/s13369-020-04758-2
13. Ghanem, S., Conly, C., Athitsos, V.: A survey on sign language recognition using smartphones. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 171–176. Association for Computing Machinery, New York, NY (2017)
14. Balakrishnan, S., Kumar, B.N., Subramanian, S.K., Nithiskumar, R.: A survey for recognizing sign language for deaf-mute people using image-based hand recognition. AIP Conf. Proc. 2742, 020007 (2024)
15. MuthuMariappan, H., Gomathi, V.: Indian sign language recognition through hybrid ConvNet-LSTM networks. EMITTER Int. J. Eng. Technol. 9(1), 182–203 (2021)
16. Mistree, K., Thakor, D., Bhatt, B.: Towards performance improvement in Indian sign language recognition. In: Proceedings of the 17th International Conference on Natural Language Processing (ICON), pp. 349–354. NLP Association of India (NLPAI), Indian Institute of Technology Patna, India (2020)
17. Mittal, A., Kumar, P., Roy, P.P., Balasubramanian, R., Chaudhuri, B.B.: A modified LSTM model for continuous sign language recognition using leap motion. IEEE Sens. J. 19, 7056–7063 (2019)
18. Coster, M.D., Shterionov, D., Herreweghe, M.V., Dambre, J.: Machine translation from signed to spoken languages: State of the art and challenges. Univ. Access Inf. Soc. 23, 1305–1331 (2022)
19. Camgoz, N.C., Koller, O., Hadfield, S., Bowden, R.: Sign language transformers: Joint end-to-end sign language recognition and translation. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10020–10030. IEEE, Piscataway, NJ (2020)
20. Hanke, T.: HamNoSys-representing sign language data in language resources and language processing contexts. Paper presented at the 4th International Conference on Language Resources and Evaluation, Lisbon, 25 May 2004
21. Kaur, K., Kumar, P.: HamNoSys to SiGML conversion system for sign language automation. Procedia Comput. Sci. 89, 794–803 (2016)
22. Sugandhi, S., Bhatia, P., Kaur, S.: Online-multilingual-dictionary-using-Hamburg-notation-for-avatar-based-Indian-sign-language-generation-system. Int. J. Cognit. Lang. Sci. 12(8), 1118–1123 (2018)

23. Stokoe, W.C., Casterline, D.C., Croneberg, C.G.: A Dictionary of American Sign Language on Linguistic Principles. Linstok Press, Silver Spring, Maryland (1965)
24. Parkhurst, S., Parkhurst, D.: SignWriting: A Complete System for Writing and Reading Signed Languages. SignWriting Press (2007), Signwriting.org
25. Hoffmann-Dilloway, E.: Writing the smile: Language ideologies in, and through, sign language scripts. Lang. Commun. 31(4), 345–355 (2011). https://doi.org/10.1016/j.langcom.2011.05.008
26. Natarajan, B., Elakkiya, R.: Dynamic GAN for high-quality sign language video generation from skeletal poses using generative adversarial networks. Soft Comput. 26(23), 13153–13175 (2022). https://doi.org/10.1007/s00500-022-07014-x
27. Attar, R., Goyal, V., Goyal, L.: State-of-the-art of automation in sign language: A systematic review. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22(4), 1–80 (2022). https://doi.org/10.1145/3564769
28. Eberhard, D., Simons, G., Fennig, C.: Ethnologue: Languages of the World, 24th Edition. SIL International, Dallas, Texas (2021)
29. Ardiansyah, A., Hitoyoshi, B., Halim, M., Hanafiah, N., Wibisurya, A.: Systematic literature review: American sign language translator. Procedia Comput. Sci. 179, 541–549 (2021). https://doi.org/10.1016/j.procs.2021.01.038
30. Bhatia, P., Wadhawan, A.: Sign language recognition systems: A decade systematic literature review. Arch. Comput. Methods Eng. 28, 785–813 (2019). https://doi.org/10.1007/s11831-019-09384-2
31. Dhanjal, A., Singh, W.: An automatic machine translation system for multi-lingual speech to Indian sign language. Multimedia Tools Appl. 81, 1–39 (2022). https://doi.org/10.1007/s11042-021-11706-1
32. Ghotkar, A., Gajanan, K., Kharate, G.: Study of vision based hand gesture recognition using Indian sign language. Int. J. Smart Sens. Intell. Syst. 7(1), 96–115 (2014). https://doi.org/10.21307/ijssis-2017-647
33. Kakde, M.U., Nakrani, M.G., Rawate, A.M.: A review paper on sign language recognition system for deaf and dumb people using image processing. Int. J. Eng. Res. Technol. 5(3), 590–592 (2016). https://doi.org/10.17577/IJERTV5IS031036
34. Dasgupta, T., Anupam, B.: Prototype machine translation system from text-to-Indian sign language. In: Proceedings of the 13th International Conference on Intelligent User Interfaces, pp. 313–316. Association for Computing Machinery, New York, NY (2008)
35. ISL Dictionary (by Ramakrishna Mission Vivekananda University). https://indiansignlanguage.org. Accessed: 17 Nov 2024
36. ISLRTC Dictionary (by Indian Sign Language Research and Training Center). http://islrtc.nic.in. Accessed: 17 Nov 2024
37. ISL Dictionary (by Mook Badhir Mandal). http://indiandeaf.org/ISL. Accessed: 17 Nov 2024
38. Choudhury, A., Talukdar, A., Sarma, K.: A review on vision-based hand gesture recognition and applications. In: Intelligent Applications for Heterogeneous System Modeling and Design, pp. 261–286. IGI Global, PA, USA (2015)
39. Bahia, N., Rani, R.: Multi-level taxonomy review for sign language recognition: Emphasis on Indian sign language. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22(1), 1–39 (2022). https://doi.org/10.1145/3530259
40. Gupta, R., Kumar, A.: Indian sign language recognition using wearable sensors and multi-label classification. Comput. Electr. Eng. 90, 106898 (2020)
41. Naglot, D., Kulkarni, M.: ANN based Indian sign language numerals recognition using the leap motion controller. In: Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT), pp. 1–6. IEEE, Piscataway, NJ (2016)
42. Oz, C., Leu, M.: Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker. Neurocomputing 70, 2891–2901 (2007). https://doi.org/10.1016/j.neucom.2006.04.016
43. Kakoty, N., Dev Sharma, M.: Recognition of sign language alphabets and numbers based on hand kinematics using a data glove. Procedia Comput. Sci. 133, 55–62 (2018). https://doi.org/10.1016/j.procs.2018.07.008
44. Das, A., Yadav, L., Singhal, M., et al.: Smart glove for sign language communications. In: Proceedings of the 2016 International Conference on Accessibility to Digital World (ICADW), pp. 27–31. IEEE, Piscataway, NJ (2016)
45. Santhosh, K., Hoysala, S., Srihari, D., Chandra, S., Krishna, A.: Gesture Recognition of Indian Sign Language. In: Bhateja, V., Satapathy, S., Travieso-Gonzalez, C., Flores-Fuentes, W. (eds.) Computer Communication, Networking and IoT. Lecture Notes in Networks and Systems, pp. 25–36. Springer, Singapore (2021)
46. Kumar, A., Kumar, R.: A novel approach for ISL alphabet recognition using extreme learning machine. Int. J. Inf. Technol. 13, 349–357 (2020). https://doi.org/10.1007/s41870-020-00525-6
47. Joudaki, S., Mohamad, D., Saba, T., Rehman, A., Al-Rodhaan, M., Al-Dhelaan, A.: Vision-based sign language classification: A directional review. IETE Tech. Rev. 31(5), 383–391 (2014). https://doi.org/10.1080/02564602.2014.961576
48. Das, P., De, R., Paul, S., Chowdhury, M., Neogi, B.: Analytical study and overview on glove based Indian sign language interpretation technique. In: Proceedings of the Michael Faraday IET International Summit 2015, pp. 313–318. IEEE, Piscataway, NJ 2015
49. Gupta, R., Rajan, S.: Comparative analysis of convolution neural network models for continuous Indian sign language classification. Procedia Comput. Sci. 171, 1542–1550 (2020). https://doi.org/10.1016/j.procs.2020.04.165
50. Kumar, P., Saini, R., Behera, S.K., Dogra, D.P., Roy, P.P.: Real-time recognition of sign language gestures and air-writing using leap motion. In: Proceedings of the Fifteenth IEEE International Conference on Machine Vision Applications (MVA), pp. 157–160. IEEE, Piscataway, NJ (2017)
51. Oudah, M., Al-Naji, A.A., Chahl, J.: Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. J. Imaging 6, 73 (2020). https://doi.org/10.3390/jimaging6080073
52. El-Alfy, E.S., Luqman, H.: A comprehensive survey and taxonomy of sign language research. Eng. Appl. Artif. Intell. 114, 105198 (2022). https://doi.org/10.1016/j.engappai.2022.105198
53. Samal, B., Panda, M.: Integrative review on vision-based dynamic Indian sign language recognition systems. In: Proceedings of the 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), pp. 1–6. IEEE, Piscataway, NJ (2021)
54. Prasath, G., Kumarappan, A.: A review on deaf and dumb communication system based on various recognitions aspect. In: Proceedings of International Conference on Deep Learning, Computing and Intelligence, pp. 191–203. Springer Nature, Singapore (2022)
55. Patel, D.U., Joshi, J.M.: Review of Indian dynamic sign language recognition system. In: Proceedings of International Conference on Communication and Computational Technologies, pp. 17–37. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-3951-8_2
56. Ahmed, W., Chanda, K., Mitra, S.: Vision based hand gesture recognition using dynamic time warping for Indian sign language. In: Proceedings of the 2016 International Conference on Information Science (ICIS), pp. 120–125. IEEE, Piscataway, NJ (2016)
57. Rahaman, M.A., Jasim, M., Ali, M., Hasanuzzaman, M.: Real-time computer vision-based Bengali sign language recognition. In: Proceedings of the 2014 17th International Conference on Computer and Information Technology (ICCIT), pp. 192–197. IEEE, Piscataway, NJ (2014)
58. Hasan, M., Sajib, T.H., Dey, M.: A machine learning based approach for the detection and recognition of Bangla sign language. In: Proceedings of the 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec), pp. 1–5. IEEE, Piscataway, NJ (2016)
59. Nanivadekar, P.A., Kulkarni, V.: Indian sign language recognition: Database creation, hand tracking and segmentation. In: Proceedings of the 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), pp. 358–363. IEEE, Piscataway, NJ (2014)
60. Rokade, Y., Jadav, P.: Indian sign language recognition system. Int. J. Eng. Technol. 9, 189–196 (2017). https://doi.org/10.21817/ijet/2017/v9i3/170903S030
61. Reshna, S., Jayaraju, M.: Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and

SVM. In: Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 386–390. IEEE, Piscataway, NJ (2017)

62. Shinde, A.K., Kagalkar, R.: Sign language to text and vice versa recognition using computer vision in Marathi. Natl. Conf. Adv. Comput. NCAC2015(1), 23–28 (2015)
63. Shinde, A., Kagalkar, R.: Advanced Marathi sign language recognition using computer vision. Int. J. Comput. Appl. 118, 1–7 (2015). https://doi.org/10.5120/20802-3485
64. Kallingale, A., Prabu, P.: A comparative study on Indian sign language representation. In: Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), pp. 1–6. IEEE, Piscataway, NJ (2021)
65. Nandy, A., Mondal, S., Prasad, J., Chakraborty, P., Nandi, G.: Recognizing and interpreting Indian sign language gesture for human robot interaction. In: Proceedings of the 2010 International Conference on Computer and Communication Technology, pp. 712–717. IEEE, Piscataway, NJ (2010)
66. Sridhar, A., Ganesan, R.G., Kumar, P., Khapra, M.M.: INCLUDE: A large scale dataset for Indian sign language recognition. In: Proceedings of the 28th ACM International Conference on Multimedia. pp. 1366–1375. Association for Computing Machinery, New York, NY (2020)
67. Kumar, E.K., Kishore, P., Kiran Kumar, M.T., Kumar, D.A.: 3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2-stream CNN. Neurocomputing 372, 40–54 (2020). https://doi.org/10.1016/j.neucom.2019.09.059
68. Lilha, H., Shivmurthy, D.: Evaluation of features for automated transcription of dual-handed sign language alphabets. In: Proceedings of the 2011 International Conference on Image Information Processing (ICIIP), pp. 1–5. IEEE, Piscataway, NJ (2011)
69. Kishore, P., Kumar, R.P.: A video based Indian sign language recognition system (INSLR) using wavelet transform and fuzzy logic. Int. J. Eng. Technol. 4, 537–542 (2012). https://doi.org/10.7763/IJET.2012.V4.427
70. Dixit, K., Jalal, A.S.: Automatic Indian sign language recognition system. In: Proceedings of the 2013 3rd IEEE International Advance Computing Conference (IACC), pp. 883–887. IEEE, Piscataway, NJ (2013)
71. Sahoo, A., Ravulakollu, K.: Vision based Indian sign language character recognition. J. Theor. Appl. Inf. Technol. 67, 770–780 (2014)
72. Sharma, M., Pal, R., Sahoo, A.K.: Indian sign language recognition using neural networks and KNN classifiers. ARPN J. Eng. Appl. Sci. 9(8), 1255–1259 (2014)
73. Tripathi, K., Baranwal, N., Nandi, G.C.: Continuous dynamic Indian Sign Language gesture recognition with invariant backgrounds. In: Proceedings of the 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2211–2216. IEEE, Piscataway, NJ (2015)
74. Mehrotra, K., Godbole, A., Belhe, S.: Indian sign language recognition using Kinect sensor. In: Proceedings of the International Conference on Image Analysis and Recognition, pp. 528–535. Springer, Cham (2015)
75. Ansari, Z., Harit, G.: Nearest neighbour classification of Indian sign language gestures using Kinect camera. Sadhana 41, 161–182 (2016). https://doi.org/10.1007/s12046-015-0405-3
76. Dour, G., Sharma, S.: Recognition of alphabets of Indian sign language by Sugeno type fuzzy neural network. Pattern Recognit. Lett. 30, 737–742 (2016)
77. Kharate, G.K., Ghotkar, A.: Vision based multi-feature hand gesture recognition for Indian sign language manual signs. Int. J. Smart Sens. Intell. Syst. 9, 124–147 (2016)
78. Prasad, M.V.D., Kishore, P.V.V., Eepuri, K., Kumar, D.A.: Indian Sign language recognition system using new fusion based edge operator. Pattern Recognit. Lett. 88, 574–584 (2016)
79. Kumar, P., Gauba, H., Roy, P., Dogra, D.: A multimodal framework for sensor-based sign language recognition. Neurocomputing 259, 21–38 (2017). https://doi.org/10.1016/j.neucom.2016.08.132
80. Kaur, B., Joshi, G., Vig, R.: Identification of ISL alphabets using discrete orthogonal moments. Wireless Pers. Commun. 95, 4823–4845 (2017)
81. Kumar, P., Saini, R., Roy, P., Dogra, D.: A position and rotation invariant framework for sign language recognition (SLR) using Kinect. Multimedia Tools Appl. 77(7), 8823–8846 (2017). https://doi.org/10.1007/s11042-017-4776-9
82. Kishore, P., Anil Kumar, D., Sastry, A., Eepuri, K.: Motionlets matching with adaptive kernels for 3-D Indian sign language recognition. IEEE Sens. J. 18(8), 3327–3337 (2018). https://doi.org/10.1109/JSEN.2018.2810449
83. Rao, G., Syamala, K., Kishore, P.V.V., Sastry, A.S.C.: Deep convolutional neural networks for sign language recognition. In: Proceedings of the 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), pp. 194–197. IEEE, Piscataway, NJ (2018)
84. Ravi, S., Suman, M., Kishore, P., Eepuri, K., Maddala, T., Anil Kumar, D.: Multi modal spatio temporal co-trained CNNs with single modal testing on RGB–D based sign language gesture recognition. J. Comput. Lang. 52, 88–102 (2019). https://doi.org/10.1016/j.cola.2019.04.002
85. Athira, P., Sruthi, C., Lijiya, A.: A signer independent sign language recognition with co-articulation elimination from live videos: An Indian scenario. J. King Saud Univ. Comput. Inf. Sci. 34(3), 771–781 (2022). https://doi.org/10.1016/j.jksuci.2019.05.002
86. Mariappan, H.M., Gomathi, V.V.: Real-time recognition of Indian sign language. In: Proceedings of the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), pp. 1–6. IEEE, Piscataway, NJ (2019)
87. Adithya, V., Rajesh, R.: Hand gestures for emergency situations: A video dataset based on words from Indian sign language. Data Brief 31, 106016 (2020)
88. Bhatia, P., Wadhawan, A.: Deep learning-based sign language recognition system for static signs. Neural Comput. Appl. 32, 7957–7968 (2021). https://doi.org/10.1007/s00521-019-04691-y
89. Raghuveera, T., Deepthi, R., Mangalashri, R.: A Depth-Based Indian Sign Language Recognition Using Microsoft Kinect. Sādhanā 45, 34 (2020). https://doi.org/10.1007/s12046-019-1250-6.
90. Gangrade, J., Bharti, J.: Vision-based hand gesture recognition for Indian sign language using convolution neural network. IETE J. Res. 69, 1–10 (2020). https://doi.org/10.1080/03772063.2020.1838342
91. Sharma, A., Sharma, N., Saxena, Y., Singh, A., Sadhya, D.: Benchmarking deep neural network approaches for Indian Sign Language recognition. Neural Comput. Appl. 33, 1–12 (2021). https://doi.org/10.1007/s00521-020-05448-8
92. Mistree, K., Thakor, D., Bhatt, B.: Towards Indian sign language sentence recognition using INSIGNVID: Indian sign language video dataset. Int. J. Adv. Comput. Sci. Appl. 12(8), 0120881 (2021). https://doi.org/10.14569/IJACSA.2021.0120881
93. Sharma, S., Singh, S.: Vision-based hand gesture recognition using deep learning for the interpretation of sign language. Expert Syst. Appl. 182, 115657 (2021). https://doi.org/10.1016/j.eswa.2021.115657
94. Sharma, S., Singh, S.: Recognition of Indian sign language (ISL) using deep learning model. Wireless Pers. Commun. 123, 671–692 (2022). https://doi.org/10.1007/s11277-021-09152-1
95. Joshi, A., Bhat, A., S, P., et al.: CISLR: Corpus for Indian sign language recognition. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 10357–10366. Association for Computational Linguistics, Stroudsburg, PA (2022)
96. Subramanian, B., Olimov, B., Naik, S., Kim, S., Park, K.H., Kim, J.: An integrated mediapipe-optimized GRU model for Indian sign language recognition. Sci. Rep. 12, 11964 (2022). https://doi.org/10.1038/s41598-022-15998-7
97. Katoch, S., Singh, V., Tiwary, U.S.: Indian sign language recognition system using SURF with SVM and CNN. Array 14, 100141 (2022). https://doi.org/10.1016/j.array.2022.100141
98. Elangovan, R., R, E., Subramaniyaswamy, V., et al.: Multi-semantic discriminative feature learning for sign gesture recognition using hybrid deep neural architecture. IEEE Access 11, 2226–2238 (2023). https://doi.org/10.1109/ACCESS.2022.3233671
99. Joshi, A., Agrawal, S., Modi, A.: ISLTranslate: Dataset for translating Indian sign language. In: Proceedings of the Findings of the Association for Computational Linguistics ACL 2024, pp. 10466–10475. Association for Computational Linguistics, Stroudsburg, PA (2023)

100. Joshi, A., Mohanty, R., Kanakanti, M., et al.: iSign: A Benchmark for Indian Sign Language Processing. In: Proceedings of the Findings of the Association for Computational Linguistics ACL 2024, pp. 10827–10844. Association for Computational Linguistics, Stroudsburg, PA (2024)
101. DEF ISL (by Deaf Enabled Foundation): https://def.org.in/. Accessed 17 Nov 2024
102. ISH News: https://indiasigninghands.com/projects/ishnews/ (2018). Accessed 17 Nov 2024
103. Shiksha, I.S.H.: https://ishshiksha.com/ (2021). Accessed 17 Nov 2024
104. Singh, D.: Sanket–text to Indian sign language translator. http://play.google.com/store/apps/details?id=in.dsingh.sanket, (2018). Accessed 17 Nov 2024
105. Talking Hands: https://talkinghands.co.in/. Accessed 18 Nov 2024
106. Dumbre, A., Jangada, S., Gosavi, S., Gupta, J.: Classification of Indian sign Language characters utilizing convolutional neural networks and transfer learning models with different image processing techniques. In: Proceedings of the IEEE World Conference on Applied Intelligence and Computing (AIC), pp. 423–430. IEEE, Piscataway, NJ (2022)
107. KP, M.J.: Indian Traffic Sign Detection Benchmark Dataset in YOLO Format. IEEE Dataport. IEEE, Piscataway, NJ (2022). https://doi.org/10.21227/sppg-r994
108. Joy, J., Balakrishnan, K., Madhavankutty, S.: Developing a Bilingual Mobile Dictionary for Indian Sign Language and Gathering Users' Experience with SignDict. Assist. Technol. 32(3), 153–160 (2018). https://doi.org/10.1080/10400435.2018.1508093
109. Agrawal, S., Jalal, A., Bhatnagar, C.: Redundancy removal for isolated gesture in Indian sign language and recognition using multi-class support vector machine. Int. J. Comput. Vision Rob. 4, 23–38 (2014). https://doi.org/10.1504/IJCVR.2014.059361
110. Tavari, N.V., Deorankar, P.A.V., Chatur, D.P.N.: Hand Gesture Recognition of Indian Sign Language to aid Physically impaired People. Paper presented at the International Conference on Industrial Automation and Computing, Jhulelal Institute of Technology, Nagpur, 13–14 April 2014
111. Rekha, J.U., Bhattacharya, J., Majumder, S.: Shape, texture and local movement hand gesture features for Indian Sign Language recognition. In: Proceedings of the 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011), pp. 30–35. IEEE, Piscataway, NJ (2011)
112. Agrawal, S.C., Jalal, A.S., Bhatnagar, C.: Recognition of Indian sign language using feature fusion. In: Proceedings of the 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), pp. 1–5. IEEE, Piscataway, NJ (2012)
113. Adithya, V., Vinod, P.R., Gopalakrishnan, U.: Artificial neural network based method for Indian sign language recognition. In: Proceedings of the 2013 IEEE Conference on Information & Communication Technologies, pp. 1080–1085. IEEE, Piscataway, NJ (2013)
114. Gupta, B., Shukla, P., Mittal, A.: K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. In: Proceedings of the 2016 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–5. IEEE, Piscataway, NJ (2016)
115. Kumar, D.A., et al.: Selfie continuous sign language recognition using neural network. In: Proceedings of the 2016 IEEE Annual India Conference (INDICON), pp. 1–6. IEEE, Piscataway, NJ (2016)
116. Kumar, A., et al.: Sign language recognition. In: Proceedings of the 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), pp. 422–428. IEEE, Piscataway, NJ (2016)
117. Kumar, P., Gauba, H., Roy, P.P., Dogra, D.P.: Coupled HMM based multisensor data fusion for sign language recognition. Pattern Recognit. Lett. 86, 1–8 (2017). https://doi.org/10.1016/j.patrec.2016.12.004
118. Dutta, K.K., Bellary, S.A.S.: Machine learning techniques for Indian sign language recognition. In: Proceedings of the 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 333–336. IEEE, Piscataway, NJ (2017)
119. Rao, G., Kishore, P.V.V.: Selfie video based continuous Indian sign language recognition system. Ain Shams Eng. J. 9(4), 1929–1939 (2017)
120. Rao, G.A., Kishore, P.V.V., Sastry, A.S.C.S., Kumar, D.A., Kumar, E.K.: Selfie continuous sign language recognition with neural network classifier. In: Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications, pp. 31–40. Springer, Singapore (2018)
121. Joshi, G., Gaur, A., Sheenu: Interpretation of Indian sign language using optimal HOG feature vector. In: Advances in Computing and Data Sciences, pp. 65–73. Springer, Berlin, Heidelberg (2018)
122. Joshi, G., Singh, S., Renu, V.: Taguchi-TOPSIS based HOG parameter selection for complex background sign language recognition. J. Visual Commun. Image Represent. 71, 102834 (2020). https://doi.org/10.1016/j.jvcir.2020.102834
123. Tripathi, K., Nandi, N.: Continuous Indian sign language gesture recognition and sentence formation. Procedia Comput. Sci. 54, 523–531 (2015). https://doi.org/10.1016/j.procs.2015.06.060
124. Elakkiya, R.: Machine learning based sign language recognition: A review and its research frontier. J. Ambient Intell. Hum. Comput. 12, 7205–7224 (2021). https://doi.org/10.1007/s12652-020-02396-y
125. Yasir, F., Prasad, P.W.C., Alsadoon, A., Elchouemi, A.: SIFT based approach on Bangla sign language recognition. In: Proceedings of the 2015 IEEE 8th International Workshop on Computational Intelligence and Applications (IWCIA), pp. 35–39. IEEE, Piscataway, NJ (2015)
126. Uddin, M.A., Chowdhury, S.A.: Hand sign language recognition for Bangla alphabet using Support Vector Machine. In: Proceedings of the 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET), pp. 1–4. IEEE, Piscataway, NJ (2016)
127. Sumaiya, Prasad, A.R., Sukruth, M., Varsha, R., Varshith, S.: Kannada sign language recognition using machine learning. Int. Res. J. Eng. Technol. 9(7), 859–866 (2022)
128. Adithya, V., Rajesh, R.: A deep convolutional neural network approach for static hand gesture recognition. Procedia Comput. Sci. 171, 2353–2361 (2020). https://doi.org/10.1016/j.procs.2020.04.255
129. Natarajan, B.S., R, E., Elakkiya, R., et al.: Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation. IEEE Access 10, 104358–104374 (2022).
130. Areeb, M., Maryam, M., Nadeem, M., Alroobaea, R., Anwer, F.: Helping hearing-impaired in emergency situations: A deep learning-based approach. IEEE Access 10, 8502–8517 (2022). https://doi.org/10.1109/ACCESS.2022.3142918
131. Kothadiya, D., Bhatt, C., Saba, T., Rehman, A.: SIGNFORMER: DeepVision Transformer for Sign Language Recognition. IEEE Access 11, 4730–4739 (2023). https://doi.org/10.1109/ACCESS.2022.3231130
132. Gogoi, R., Kumar, P., Damdoo, R.: End to End Simple Indian Sign Language Sentence Translation Using Sign Transformer Network. In: Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–13. IEEE, Piscataway, NJ (2024)
133. Kaur, J., Rama, C.: An efficient Indian sign language recognition system using Sift descriptor. Int. J. Eng. Adv. Technol. 8(6), 1456–1461 (2019)
134. Varsha, M., Nair, C.S.: Indian sign language gesture recognition using deep convolutional neural network. In: Proceedings of the 2021 8th International Conference on Smart Computing and Communications (ICSCC), pp. 193–197. IEEE, Piscataway, NJ (2021)
135. Prasath, G., Kumarappan, A.: Prediction of sign language recognition based on multi layered CNN. Multimedia Tools Appl. 82, 1–21 (2023). https://doi.org/10.1007/s11042-023-14548-1
136. Sinha, P.G.: Distinctive Feature Extraction for Indian Sign Language (ISL) Gesture using Scale Invariant Feature Transform (SIFT). J. Inst. Eng. India Ser. B 98, 19–26 (2016). https://doi.org/10.1007/s40031-016-0250-8
137. Kumar, E.K., Kishore, P.V.V., Sastry, A.S.C.S., Kumar, M.T.K., Kumar, D.A.: Training CNNs for 3-D sign language recognition with color texture coded joint angular displacement maps. IEEE Signal Process Lett. 25(5), 645–649 (2018). https://doi.org/10.1109/LSP.2018.2817179
138. Arikeri, P.: Indian Sign Language (ISL). https://www.kaggle.com/datasets/prathumarikeri/indian-sign-language-isl (2021). Accessed 18 Oct 2022
139. Kishore, P.V.V., Prasad, M.V.D., Kumar, D.A., Sastry, A.S.C.: Optical flow hand tracking and active contour hand shape features for continuous sign language recognition with artificial neural networks. In: Proceedings of

the 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 346–351. IEEE, Piscataway, NJ (2015)

140. Shenoy, K., Dastane, T., Rao, V., Vyavaharkar, D.: Real-time Indian Sign Language (ISL) Recognition. In: Proceedings of the 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–9. IEEE, Piscataway, NJ (2018)
141. Bhuyan, M.: FSM-based recognition of dynamic hand gestures via gesture summarization using key video object planes. Int. J. Comput. Commun. Eng. 1(6), 248–259 (2012)
142. E, R., K, S.: Enhanced dynamic programming approach for subunit modelling to handle segmentation and recognition ambiguities in sign language. J. Parallel Distrib. Comput. 117, 246–255 (2017). https://doi.org/10.1016/j.jpdc.2017.07.001
143. R, E., Selvamani, K.: Subunit sign modeling framework for continuous sign language recognition. Comput. Electr. Eng. 74, 379–390 (2019). https://doi.org/10.1016/j.compeleceng.2019.02.012
144. Eepuri, K., Kishore, P., Anil Kumar, D., Maddala, T.: Early estimation model for 3D-discrete Indian sign language recognition using graph matching. J. King Saud Univ. Comput. Inf. Sci. 33(7), 852–864 (2018). https://doi.org/10.1016/j.jksuci.2018.06.008
145. Nikam, A.S., Ambekar, A.G.: Sign language recognition using image based hand gesture recognition techniques. In: Proceedings of the 2016 Online International Conference on Green Engineering and Technologies (IC-GET), pp. 1–5. IEEE, Piscataway, NJ (2016)
146. Hussain, I., Talukdar, A., Sarma, K.: Hand gesture recognition system with real-time palm tracking. In: Proceedings of the 11th IEEE India Conference: Emerging Trends and Innovation in Technology, INDICON 2014, pp. 1–4. IEEE, Piscataway, NJ (2014)
147. Geetha, M., Manjusha, U.C.: A vision-based recognition of Indian sign language alphabets and numerals using B-spline approximation. Int. J. Comput. Sci. Eng. (IJCSE) 4(3), 406–415 (2012)
148. Madathilkulangara, G., Aswathi, P.: Dynamic gesture recognition of Indian sign language considering local motion of hand using spatial location of key maximum curvature points. In: Proceedings of the 2013 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2013, pp. 86–91. IEEE, Piscataway, NJ (2013)
149. Rajam, P.S., Balakrishnan, G.: Recognition of Tamil sign language alphabet using image processing to aid deaf-dumb people. Procedia Eng. 30, 861–868 (2012)
150. Ghotkar, A., Kharate, G.: Dynamic hand gesture recognition for sign words and novel sentence interpretation algorithm for Indian sign language using Microsoft Kinect sensor. J. Pattern Recognit. Res. 10(1), 24–38 (2015). https://doi.org/10.13176/11.626
151. Eunice, R., Duraisamy, H.: A review on computational methods based automated sign language recognition system for hearing and speech impaired community. Concurrency Comput. Pract. Exper. 35(9), e7653 (2023). https://doi.org/10.1002/cpe.7653
152. Krishnan, L.: A systematic literature review on the robustness of sign language recognition methods in low-light environments. TechRxiv preprint, techrxiv.170473883.37069467 (2024)

How to cite this article: Damdoo, R., Kumar, P.: An integrative survey on Indian sign language recognition and translation. IET Image Process. 19, e70000 (2025). https://doi.org/10.1049/ipr2.70000

APPENDIX A: SIGN LANGUAGE NOTATION

- HamNoSys: HamNoSys is an online multilingual ISL dictionary developed using HamNoSys notation. Prillwitz et al. [20] developed it in 1984. HamNoSys includes manual and non-manual language-independent components that represent SL features with approximately 200 characters. It is a phonetic notation that consists of SL features like hand shape, orientation, movement, location, head, body gesture, and some non-manual features.
- SignWriting: SignWriting [24] is more pictorial. In SignWriting, body components and motions are depicted respectively with straightforward drawings and arrows. Shapes of the characters (abstract representations of the hands, face, and body) and their spatial arrangement on the page (which does not follow a sequential order) make it a highly featural and aesthetically recognizable work.
- Stokoe: Stokoe notation is the first phonemic script used for sign languages. William Stokoe et al. [23] developed it for American sign language (ASL). Stokoe uses Latin letters and numerals for the shapes they have in fingerspelling and iconic glyphs to represent the hands' position, motion, and orientation.
- SiGML: The University of East Anglia (UEA) developed SiGML for specifying signing sequences. It is like XML and uses different tags for representing their corresponding HamNoSys symbol.

APPENDIX B: SIGN LANGUAGE GENERATION

Sign language generation is, creating SL utterances or animations from written or spoken language input. SL generation involves several steps, including natural language processing, semantic analysis, and motion planning. Natural language processing includes analysing the input text to identify the linguistic structure and meaning. Semantic analysis involves interpreting the input text and mapping it to the appropriate signs and gestures in the SL. After the linguistic and semantic information extraction, the system generates the related signs and movements. This stage involves selecting the appropriate hand shapes, movements, and locations on or around the body to represent the words and grammatical structures of the input text. The system must consider factors such as spatial reference, facial expressions, and body language to convey meaning accurately. Sign language generation systems use a wide range of techniques, including rule-based approaches, statistical models, and machine learning algorithms. These systems require large annotated datasets of SL videos and linguistic information to train the models and improve their accuracy. Hence, sign language generation is a challenging task requiring a deep understanding of both natural and sign language. Avatar, also known as virtual human, was developed by Glauert in 2010 at the University of East Anglia. It inputs SiGML (generated from HamNoSys), and produces the animation frames. Singh et al. [104] generate 3D avatar animation from these frames. Sugandhi et al. [22] created a system that accepts user input in Hindi and English. The input tool uses Google API for Hindi transliteration. The input is processed through a parser and translated into an ISL sentence. For the translation purposes, morphological information extraction (for each word in the sentence) and stemming has been applied based on POS tag information to convert it into ISL sentences. After recognizing the stems of the input sentence, its HamNoSys from the Database, is fetched, and the corresponding SiGML is looked up in the cache (saved after generation). This SiGML is taken as input by an animation server that uses Web Graphics Library (WebGL) for generating 3D avatar animation (islfromtext.in) [104]. IIIT-Bangalore (cognitive.iiitb.ac.in/isli/) initiated the project 'Indian Sign Language Interpreter'. The translation module in this project transforms English into a textual gloss of ISL, and the animation module turns the ISL gloss into computer animation. The translation module uses a rule-based system. A vocabulary of 1300 signs with fewer non-manual elements is available in the animation module, while 300 naturally animated signs with non-manual features like mouthing and head movement are available. There are currently around 200 sentence pairs in the corpus. Dasgupta et al. [34] prototyped text to ISL, a rule-based transfer grammar machine translation system. The system inputs simple English sentences and parses using a minibar parser. The parsed structure is encoded to grammatical relation (more likely subject, object, verb) by lexical functional grammar and converted to the ISL word structure by applying transfer rules. The ISL sentences are displayed using a stream of pre-recorded videos/icons. The authors evaluated the system based on a set of 208 sentences, resulting in an accuracy of 89.4%. Natarajan et al. [26] passed concatenated sign gesture images and skeletal key point information as input to the generator network. The generated results are classified into a class of sign glosses group using the VGG-16 framework. Further, the authors applied intermediate frame generation techniques to create intermediary frames between sign gestures. The authors used combined perceptual loss and contextual L1 loss to predict the intermediary-frame between two sign gestures. They have demonstrated the work on ISL-CSLTR dataset [8] and achieved a structural similarity index measure (SSIM) score of 0.937.

APPENDIX C: DATA ACQUISITION METHODS FOR SIGN LANGUAGE RECOGNITION

- Cyber glove: A cyber glove is used to get the gesture's joint angles features. As seen in Figure 6a, it is a lightweight, elastic glove that includes several sensors to measure how far the fingers bend in various directions. A cyber glove detects the user's gestures and translates them into commands for a computer system. To classify 60 sign words in 2007, Oz et al. [42] used a cyber glove with 22 virtually undetectable thin and flexible sensors. They recorded an accuracy of 92.00%.
- Data glove: It is the first glove-style gadget that is commercially available for hand tracking. Analogue hand gesture signals such as finger flexion, hand orientation, and hand position are detected using a variety of sensors. The signals are transformed into digital data with the help of the analogue–digital converter(ADC). As seen in Figure 6b, data gloves comprise an accelerometer and flex sensor. Flex sensors identify all finger joints' bending points, which transmit data to the microcontroller. It is mounted in an external layer, from the affiliation joints of the fingers and palm to the fingertips. An accelerometer exists behind the glove. In 2016, Das et al. [44] used a data glove equipped with five flex sensors over the fingers and a gyroscope in the middle of the palm. The maximum number of outputs it can generate is around 200 words. The authors experimented with three basic gestures and achieved 86.67% recognition accuracy. Using a locally created data glove, Kakoty et al. [43] recorded a real-time recognition of ISL and ASL alphabets and numerals based on hand kinematics. The pre-processing unit receives data from the finger joint angle sensor in the glove. The authors used the technique for standardizing feature scaling to eliminate subjectivity and enhance the recognition outcomes. The recognized SL is translated into speech using label-matching. An average recognition rate of 96.7% is obtained for the ISL and ASL numbers. Das et al. [48] discuss glove-based ISL interpretation techniques.

  Although cyber gloves and data gloves are both types of wearable devices used for HCI, there are some differences between the two. The main difference is their intended use. Cyber gloves are primarily used for human-computer interaction and virtual reality applications, data gloves are used for a broader range of applications, including virtual reality, motion capture, and medical rehabilitation.
- Colour Gloves: Colour gloves are special gloves used in recognition systems for SL. The colour helps with easier extraction/segmentation (Figure 6c). Saba et al. [47] offered a customized colour glove and an autonomous vision-based SLT system. The palm and tips of each finger on this glove are coloured. A camera records the signer's image and videos. Using colour segmentation and image processing techniques, the authors found the centroid of the colour-segmented spot and utilized ANN to classify and translate gestures. The recorded recognition rate was over 90%. Munir et al. [51] list SLR as an application area of hand gesture recognition systems with colour-based recognition using glove markers. The authors emphasize hand gesture recognition approaches and explain their benefits and drawbacks in various situations. In addition, it concentrates on classification algorithms, hand segmentation techniques, and computer vision techniques that deal with similarity and difference points. Santhosh et al. [45] achieved 96% accuracy using colour glove and DL algorithms. They created 25 classes of ISL words that contain 50 videos for each and around 1500 videos in the dataset.
- Sensor armband: SLR can be performed using a sensor armband with numerous inertial measurement units (IMU) and a surface electromyogram (sEMG). A surface electromyogram uses sensors bound to the skin's surface to estimate the electrical potentials produced in the muscles. In addition to being used for designing prosthetic limb controllers, electromyograms are also used for muscle fatigue research, health monitoring, and rich information regarding hand motions and SLR systems. An IMU consists of a tri-axial accelerometer that detects information about linear acceleration and orientation and a tri-axial gyroscope that detects the rotation rate about an orthogonal axis (Figure 6e). Rinki et al. [40] used sensor armband in 2020. Three sEMG sensors and two 6-degree-of-freedom IMUs are placed on both forearms to recognize 100 words achieving a performance of 97.27%.

TABLE C1 Pros and cons of using different acquisition methods.

Image Acquisition device	Pros	Cons
Cyber glove	○ These devices are not impacted by the environment while gathering data. ○ It offers better recognition accuracy. ○ Sensor-based feature extraction is comparatively simpler. ○ Can be used for a wide range of applications as compared to data gloves, including gaming, education, training, and medical simulations.	○ Take some time to learn and get used to. ○ It reduces the naturalness of interaction. Unfriendly in terms of use, but a cyber glove is significantly less inconvenient as compared to a data glove. ○ Very expensive. ○ PC is required to check if fingers are spread widely enough for whatever the need to represent the signs. ○ Does not track non-manual features.
Data glove	○ These devices are not impacted by the environment while gathering data. ○ It offers better recognition accuracy. ○ Sensor-based feature extraction is comparatively simpler.	○ It reduces the naturalness of interaction. The data glove could likewise be less agreeable to be worn by the user. ○ Very expensive ○ Errors in the sensors are attributed to the failures, especially while altering hand movements as significantly more exposed to noise [44]. ○ Does not track non-manual features
Colour glove	○ Improved accuracy than skin colour segmentation. ○ A colour glove can be trained to recognize a wide range of SL gestures and can be customized for specific users and dialects. ○ Cheaper.	○ Wearing a glove may not be comfortable for all users, particularly for extended periods. ○ The accuracy of the colour glove relies on the effectiveness of the algorithms used to analyse the data. ○ Does not track non-manual features.
Sensors armband	○ Generally lightweight and comfortable to wear. ○ Can often be synced with smartphone apps or other devices to provide real-time data and insights.	○ May not be able to accurately capture the intricate movements and nuances of SL. ○ Number of signs they can recognize may be limited. ○ Can be sensitive to movement, which means they may not be able to capture signs accurately if the user is moving their arms quickly or if there is a lot of background movement. ○ Can be expensive. ○ Does not track non-manual features.
Leap motion controller	○ It processes information at a quicker rate of about 200 fps and has high recognition accuracy. ○ It can locate and track hands and fingers.	○ Due to its high sensitivity, slight changes in the sign's position can impact accuracy. ○ Does not track non-manual features.
Kinect sensor	○ Useful data such as the depth and the colour of images can be easily obtained. ○ Suitable for varied applications involving human-computer interaction.	○ Distance detection has a finite range. ○ The illumination, hand and facial segmentation, a complex background, and noise can have an impact on Kinect. ○ Due to sensitivity to sunlight, Kinect is not appropriate for outdoor use.
Camera/Webcam	○ Users are not required to wear any uncomfortable external devices and need to use their hands inside the camera collection range. ○ Economical. ○ Convenient and simple to use. ○ Facial expressions can be included.	○ Environmental elements like light, skin tone, background conditions, and occlusion have a significant impact on it. ○ Several image processing techniques are needed that might reduce recognition accuracy. ○ Computers are used for processing, and the user must always have a camera.

In [49], they continued their work on the continuously signed ISL sentence classification. The sentences consist of three to four words.

- Leap motion control: As seen in Figure 6d, the leap motion control is a small USB-connected gadget that can be set up on a desktop. It can acquire 3D fingertip coordinates that transform signals into computer commands. The sensor typically picks up a 1-cubic-meter hemispherical irregular area. It can differentiate the fingers' joints and track their movements (https://www.ultraleap.com/). The monitor displays the 3D model of the subject's current gesture. With the leap motion controller, users can perform a range of gestures, such as pointing, grabbing, and swiping, to control applications and interact with virtual environments. The technology is particularly well-suited for virtual reality applications. Using natural hand and finger movements, it allows users to interact with virtual objects and environments. Anshul et al. [17] used a leap motion controller method for data acquisition of sign language. The authors acquired 12 dynamic 3D data points/features for sign recognition of 35 words and trained the system on 3150 images. They achieved an accuracy of 72.3% for sentences and 89.5% for isolated sign words. Kumar et al. [50] suggested real-time SLR using a leap motion sensor. They manually collected 2240 static signs. They carried out experiments with SVM-BLSTM, and achieved 63.57 accuracy%. Naglot and Kulkarni [41] demonstrated an ANN-based ISL recognition system for numbers using leap motion controller (LMC). They classified single-handed dynamic signs using multilayer perceptron (MLP) and attained 100% accuracy.

- Kinect sensor: Microsoft created the Kinect sensor as a motion-sensing device for Windows PCs and Xbox game consoles. It works by using a combination of sensors and software to detect and track the user's movements in real-time. The sensor includes a depth camera and an RGB camera. The depth camera uses an infrared projector to emit a pattern of dots onto the user and surrounding environment. The camera then measures the time it takes for the dots to bounce back to the sensor, which allows it to create a 3D map of the user and their surroundings. The data from these sensors is processed by specialized software that can recognize and track the user's movements in real-time. The system can perceive individual body parts, such as hands, feet, and head, and track their movements as the user interacts with the system. The RGB camera captures traditional video footage of users to provide additional data for tracking and recognition. Raghuveera et al. [89] segmented hand region and extracted hand features using Kinect Xbox 360. Using the histogram of oriented gradients, and local binary patterns it ensembles three feature classifiers trained using SVM to improve the average recognition accuracy up to 71.85%. The authors then translate the sequence of recognized hand gestures into English sentences(text and speech). Mehrotra et al. [74] found a technique to identify Indian double-handed signals. In this, 37 sign words were captured using a Microsoft Kinect, and skeleton joint features were collected. With multi-class SVM, the system can classify with an accuracy of 86.16%. Kumar et al. [79, 117] built a framework for sensor-based SLR. They collected 7500 samples for 50 sign words using Kinect and leap motion devices API to retrieve the fingertip's location and direction features. HMM-BLSTM is used to classify the sign words, and 95.60% and 84.57% accuracy were obtained respectively. In 2017 a similar approach was taken by Kumar et al. [81]. They proposed a coupled HMM-based SLR system for 16 single-handed and 14 double-handed dynamic sign words. They collected 2700 sign videos using Kinect. A position and rotation invariant system achieved 83.77% accuracy.