International Journal on Computational Modelling Applications
Vol. 03, Iss. 02, S. No. 02, pp. 20-46, 2026
ISSN (Online): 3048-8516

Received: 06 May 2026, Accepted: 28 Aug. 2026, Published: 31 Aug. 2026
Digital Object Identifier: https://doi.org/10.63503/j.ijcma.2026.255

Review Article

Advances in Indian Sign Language Processing: A State-of-the-Art Review and a Real-Time Bidirectional Translation Architecture

Shefali Khera[1], Shreya[1], Meghna Sharma[1], Srishti Sharma[2]

[1]Department of Computer Science & Engineering, The NorthCap University, Gurugram, Haryana, India

[2]Department of Computer Science, University of Southampton (Delhi), India

shefali22csu426@ncuindia.edu, 22csu515@ncuindia.edu, meghnasharma@ncuindia.edu, srishti@ncuindia.edu

*Corresponding author: Shefali Khera, shefali22csu426@ncuindia.edu

ABSTRACT

This paper presents a comprehensive review of existing datasets, methodologies, and experimental frameworks for Indian Sign Language (ISL) recognition and translation, highlighting recent advancements driven by deep learning architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Transformers. Despite this progress, the review finds that current systems remain limited by dataset diversity, lack of standardization, and insufficient support for real-time, user-centric applications. Building on these findings, this study proposes an integrated ISL framework, "Beyond," that combines dataset curation with real-time, bidirectional, and multilingual (Hindi-English) translation, supported by a finger-spelling fallback for out-of-vocabulary words. The framework further incorporates word-level video alignment, background-normalized gesture recognition, predictive text support, and accessibility-oriented controls, along with a proposed multimodal dataset designed for linguistic alignment and temporal coherence. Together, these contributions aim to bridge ISL research and real-world deployment, supporting scalable, real-time, and inclusive communication for the Deaf and Hard-of-Hearing community.

Keywords: *Indian Sign Language (ISL), Deep Learning, Dataset Benchmarking, MediaPipe, Convolutional Neural Network (CNN), Transformer, ISL-CSLTR, INCLUDE Dataset, SignFlow, HWGAT*

1. Introduction

Research on Indian Sign Language recognition and translation is progressing rapidly owing to state-of-the-art deep learning developments coupled with new multimodal sensing technologies [34]. Quality datasets have played an important role in this development, as they have aided in the creation of sensitive computer models that can detect the subtle spatial, temporal, and linguistic patterns embedded in sign language gestures. The increased global interest in accessible and inclusive communication has made this development in ISL research an important step towards bridging the interaction gap between DHH and the hearing community.

Despite the progress in ISL recognition and translation, most existing systems remain limited to experimental environments, lacking real-time usability, multilingual interaction, and user-centric design. Furthermore, there is minimal integration between recognition, translation, and accessibility features within a single unified platform.

To address these challenges, this paper not only reviews existing ISL datasets and methodologies but also proposes an integrated system, *Beyond*, which combines dataset curation with a real-time, interactive, and bilingual ISL translation platform. The system extends beyond traditional frameworks by incorporating speech interfaces, gesture recognition, assistive AI modules, and accessibility-driven features designed for practical deployment.

Contribution of this work beyond a conventional review: while the sections that follow survey existing ISL datasets and recognition techniques in the manner of a conventional review, this paper's principal contribution is the "Beyond" system itself. Specifically, this work contributes:

- A unified architecture that couples dataset curation with real-time, bidirectional ISL recognition and translation, rather than treating them as separate research problems;
- A hybrid, timestamp-annotated dataset construction strategy that merges multiple public ISL resources without requiring storage of redundant video clips;
- Bidirectional translation pipelines for text/speech-to-ISL and ISL-to-text/speech conversion, supporting English and Hindi;
- A finger-spelling fallback mechanism that guarantees output for out-of-vocabulary words;
- Vision-based gesture recognition with adaptive, background-independent preprocessing, combined with accessibility-oriented features such as predictive text and an AI-assisted interaction module aimed at real-world deployment rather than laboratory evaluation alone.

1.1 Overview

Various aspects of developing a system that recognizes and translates Indian Sign Language include computer vision, natural language processing, and human-computer interaction. Indian Sign Language Recognition employs computers that are highly adept at analyzing human actions and the movements of their limbs and fingers. These computers are capable of recognizing and tracking human gestures and modifications to these gestures over time. This has really helped computers understand Indian Sign Language on their own. Indian Sign Language comprises many signs, and the computer must be able to understand them [38]. Yet, the important constraints that exist in this field include the lack of diversity in the database, dependency of the system on the signer, lack of consistency in the annotation schema, and lack of a standardized evaluation protocol. All of these issues need to be sorted out as soon as possible to develop a successful ISL recognition and translation system.

1.2 Indian Sign Language (ISL): Linguistic and Technological Background

Indian Sign Language is the major mode of communication for millions of Deaf and Hard of Hearing people in India. It is a full-fledged, rich visual-gestural language in which information is communicated through spatially coordinated hand movements, hand orientations, facial expressions, and body postures [20]. Unlike spoken or written languages, ISL is organized based on spatial-visual grammatical rules rather than phonetic or textual structures.

Indian Sign Language also exhibits regional variation influenced by the spoken languages of different parts of the country. This variation, combined with the absence of a standardized notation system and the limited availability of labeled data, poses significant challenges for the computational modeling of ISL.

From a technological standpoint, computational models must therefore accommodate the visual and kinematic properties of ISL, including hand shape, orientation, movement trajectory, and signing speed.

ISL recognition and translation not only go beyond identifying gestures but also encompass understanding semantic relationships and dependencies, as well as translating meaningful language. The performance of recognition has improved in recent years, especially with advances in computer vision and the introduction of the CNN, LSTM, and Transformer architectures [32]. Despite this, ISL remains largely underexplored compared with ASL and BSL. The lack of representation highlights the need for standardised datasets, frameworks, and benchmarks to accelerate the development of ISL recognition and translation systems.

1.3 Purpose and Objectives

This paper reviews the datasets, recognition frameworks, and translation systems currently available for Indian Sign Language, with the aim of identifying what has been achieved and what gaps remain and then proposes an integrated system, *Beyond*, which combines dataset curation with a real-time, interactive, and bilingual ISL translation platform. The system extends beyond traditional frameworks by incorporating speech interfaces, gesture recognition, assistive AI modules, and accessibility-driven features designed for practical deployment. The specific objectives of the paper are as follows:

1. Review major publicly available datasets for ISL, such as ISL-CSLTR, INCLUDE, iSign Benchmark, and ISLTranslate, as well as recent multimodal resources like HWGAT and OpenHands.
2. Analyze the modeling architectures used in ISL research, such as CNN, LSTM, Transformer, and hybrid approaches.
3. Compare preprocessing and feature extraction pipelines, including MediaPipe, OpenPose, and related computer vision-based frameworks.
4. Examine key challenges pertaining to signer independence, dataset imbalance, annotation inconsistencies, and a lack of a unified evaluation benchmark.
5. Identify research gaps and propose directions for developing large-scale, transformer-based ISL systems that integrate gesture recognition, translation, and speech interfaces in real time.
6. To design and implement a unified ISL platform, “Beyond” , that integrates dataset curation, real-time gesture recognition, bidirectional translation, and accessibility-driven user interaction within a scalable framework.
7. To incorporate intelligent and user-centric features such as multilingual text and speech input, word-level ISL video mapping, finger-spelling fallback mechanisms, and interactive learning controls to enhance real-time usability and accessibility.

2. Sign Language Recognition and Translation

Persons who are either hard of hearing (HoH) or who have a hearing impairment face a number of challenges, and this affects their social inclusion, leading to emotional problems, social isolation, low self-esteem, as well as educational and professional problems, among others. Even though there are many technological solutions to hearing loss, many problems still persist.

Recent developments in artificial intelligence (AI), computer vision, and machine learning have transformed the potential to understand and translate sign language [37]. With the help of AI-based technologies, it is now possible to interpret complex gestures, thus creating a more natural interaction between the Deaf community and the hearing population. Significantly, India has emerged as a major contributor to research on sign language recognition (SLR) and translation (SLT) globally over the past two decades. The basic concepts of these technologies are discussed in the following sections.

2.1 Review of Sign Language Recognition (SLR) Approaches

However, in recent years, advances in sign language processing techniques have led to a shift from modular or pipelined systems to end-to-end recognition and translation. In the conventional system, sign language recognition (SLR) is performed first on the provided visual inputs, and then translation occurs. Conventional systems, though successful in their initial stages, often face problems with error propagation and intermediate annotations.

Fig. 1 illustrates the conventional pipeline-based architecture for sign language translation, highlighting the separation between recognition and translation components. Although widely adopted in early research, this design has gradually been replaced by end-to-end approaches that directly learn mappings from sign language videos to target language sentences.

Sign Language Recognition (SLR) involves detecting and understanding sign gestures using videos or sensors. It converts both manual and non-manual signs into digital forms that computers can understand. Such sign language recognition systems can be implemented using deep learning techniques, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, or combinations of these, to effectively capture spatial and temporal gesture dynamics.[13].

SLR systems are generally divided into two categories:

- **Isolated Sign Recognition** — This type identifies individual signs on their own.
- **Continuous Sign Recognition** — This type analyzes sequences of signs that form sentences or conversations.

While both tasks are closely related, SLR and SLT differ in scope: SLR performs gesture-level classification, whereas SLT additionally models sequential and grammatical structure to produce coherent target-language output. Because SLT accuracy is bounded by upstream recognition quality, pipeline-based systems propagate SLR errors into translation, which is precisely the limitation that joint end-to-end SLRT architectures, discussed in Section 2.2, attempt to overcome.

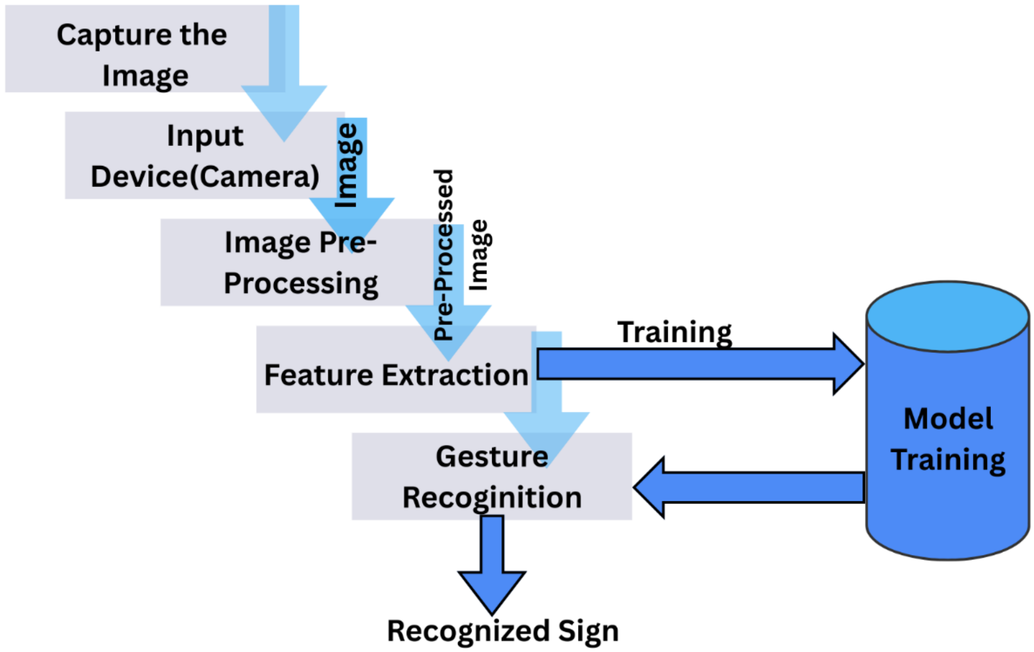

Fig 1. Processed Flow of Sign Language Translation

As shown in Fig. 1, this pipeline processes recognition and translation as sequential stages, so any misclassification at the recognition stage propagates downstream and cannot be corrected during translation - the key motivation for the end-to-end designs discussed in Section 2.2.

2.2 Joint End-to-End Recognition and Translation

A joint end-to-end SLRT system integrates recognition and translation into a single system. In other words, rather than recognizing the signs separately and then translating the text, the joint end-to-end system of SLRT maps the sign videos to text or speech using the Transformer architecture.

A major advantage of this integrated approach is the contextual understanding of the system, as it learns the structure of the sign language. In traditional approaches, an intermediate representation of gloss, or a written shorthand of signs, had to be used. This representation helped to bridge the gap between sign language recognition and translation. For instance, the English sentence "What do you want to become?" will have the gloss sequence in ISL as "You become what." The disadvantage of such gloss-based systems is that they are less fluent in translation

The drawback of joint end-to-end methods is not present, but they require sufficient data to properly train the systems. Joint end-to-end methods have been developed for ASL and BSL, but end-to-end translation systems for ISL are not available, creating a major research gap.

3. Indian Sign Language

Indian Sign Language (ISL) is the officially recognized visual-gestural language utilized by the Deaf and Hard-of-Hearing (HoH) community in India and some neighboring countries. Around 6 million people in India and an additional 6.8 million in nearby areas use ISL as their main form of communication [46].

The sign language situation in India is quite diversified, with the presence of various sign languages such as Tamil Sign Language (TSL), Bangla Sign Language (BdSL), Kannada Sign Language (KSL), and Marathi Sign Language (MSL) [7]. Each of the sign languages has evolved under the influence of local linguistic, interactional, and cultural factors.

The historical development of ISL is closely linked to the development of deaf education in India, which began with the establishment of the first schools for the Deaf. Although sign language was formally introduced into the education system in 2001, its use has increased since then due to advocacy and linguistic studies [39].

As the Deaf community grows in India, ISL plays a significant role in the community's cultural identity. However, there are still challenges ahead, such as low levels of English literacy, limited access to formal interpretation services, and the communication divide between the Deaf and hearing communities. So far, there are fewer than 300 certified interpreters of ISL, which is quite different from the millions of people using the language as the medium of communication in their daily lives [4]

In recent times, with the emergence of Artificial Intelligence (AI) and Machine Learning (ML), researchers are beginning to explore the possibility of using technology to improve communication using Indian Sign Language Recognition and Translation (ISLRT) technology [42].

The objective of this study is to examine and compile the existing ISL data and recognition techniques, their advantages and disadvantages, and possible future developments in the creation of continuous sign language translation.

3.1 Classification of Signs in ISL

Sign languages, including ISL, are expressed through articulators, visible, movable parts of the body such as the hands, head, face, and upper body. These articulators work together to represent words, emotions, and prosody (intonation and rhythm in signing).

Five fundamental parameters define the structure of ISL articulation [21]:

1. Hand and head recognition
2. Hand and head orientation
3. Hand and head movement
4. Location of hand and head
5. Shape of the hand

According to Kumar et al. [35], ISL articulations can be divided into two categories:

- Word representation, where articulators express the complete meanings of words or concepts.
- Fingerspelling, where hand shapes represent individual letters of the alphabet, is often used for names, places, or borrowed terms.

The classification of ISL signs is crucial for organizing the vocabulary and understanding the grammatical structure of the language. Fig2. illustrates a hierarchical representation of ISL signs Table 1 summarizes commonly used sign classifications in ISL, which are relevant for defining annotation categories and labeling strategies in ISL datasets. [22]

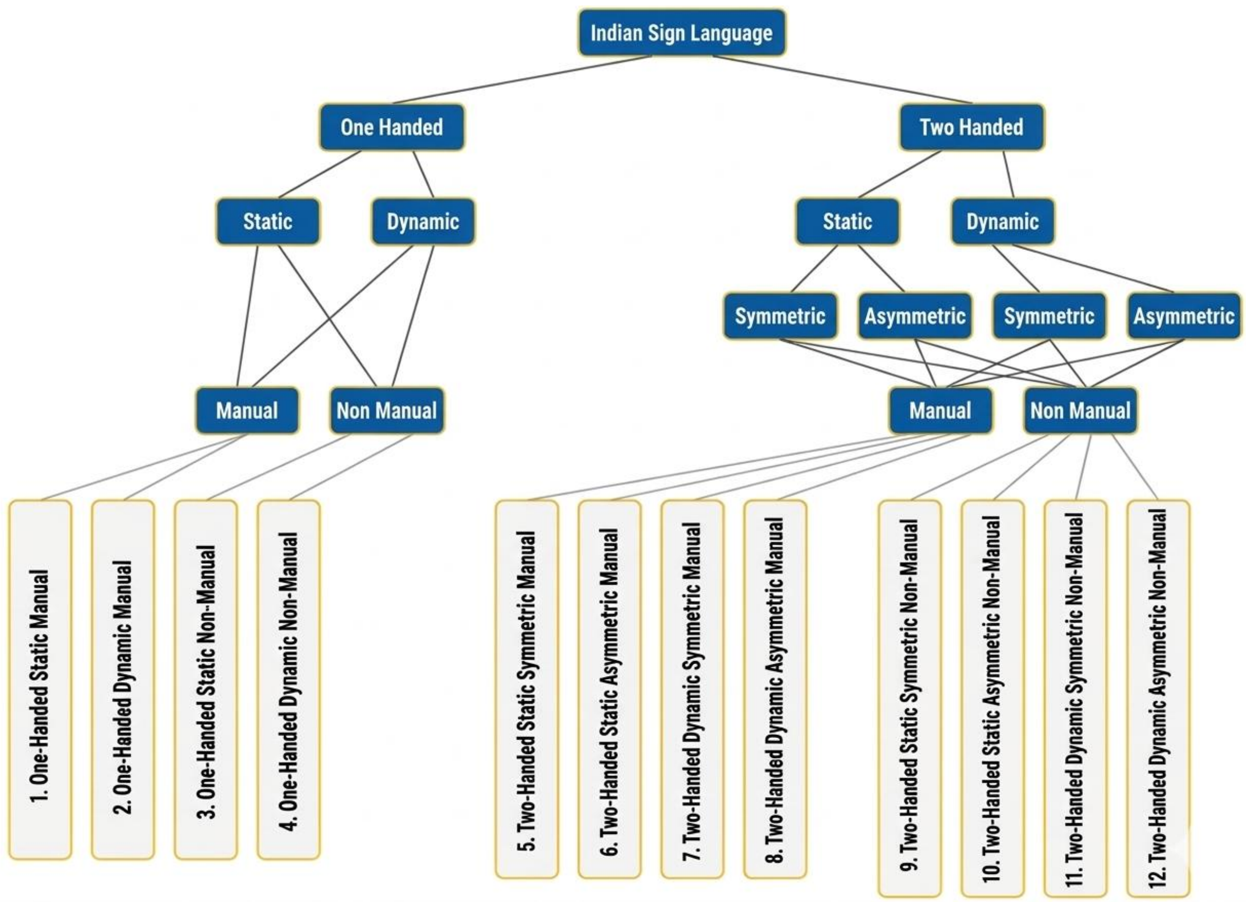

Fig 2. Hierarchical Classification of ISL

The hierarchy in Fig. 2 clarifies why a single flat label set is insufficient for ISL annotation: manual versus non-manual signs and static versus dynamic signs each require distinct annotation conventions and, in practice, distinct recognition strategies, which is why Table 1 explicitly separates these categories.

Table 1. Terminology in ISL Classification

Terminology	**Description**
Manual signs	Formed using one or both hands, involving hand shape, motion, and position. Common examples include ISL numerals and alphabet.
Non-manual signs	Involve facial expressions, head tilts, or body movements, either alone or combined with manual gestures (e.g., "fear," "not," "sleep").
One-handed signs	Produced with a single dominant hand; can be static or dynamic.
Two-handed signs	Use both hands, and it can be static or dynamic, symmetric or asymmetric.
Symmetric signs	Both hands are active simultaneously.

Asymmetric signs	One hand (dominant) performs the main motion while the other assists.
Static signs	Gestures that remain in a fixed position without movement.
Dynamic signs	Gestures involving motion or change in hand configuration to convey meaning.

3.2 Machine learning-based approach for sign language recognition

Over the past decade, various techniques have been proposed for the recognition and translation of Indian Sign Language (ISL). This is due to increased research interest in the field. All the techniques proposed over the past decade can be classified into various categories based on the methodologies used. The techniques include traditional machine learning, deep learning, motion tracking, and graph-based techniques. Although techniques proposed over the past decade show promising results, they remain affected by data size, vocabulary, and the lack of evaluation standards.

From the very beginning, ML methods were among the earliest approaches for ISL recognition. Handcrafted feature extraction and classification algorithms are the typical steps under this paradigm. Some widely used techniques for gesture classification include SVM [44], KNN [49], RF [49], and Naïve Bayes. Feature descriptors like HOG [50], SIFT [50], and SURF [5] are widely used to represent hand shape and motion.

Efficiency was increased by dimensionality-reduction methods such as PCA and LDA, while the use of ensemble classifiers improved robustness. Some works also applied fuzzy logic to accomplish temporal gesture recognition using HMM [53]. Despite notable accuracy on small, controlled datasets, these approaches struggled with scalability, signer variation, and real-time performance.

3.3 Deep Learning Approaches

Deep learning significantly improved ISL recognition performance and eliminated the need to manually extract features. LeNet, AlexNet, VGG16, and ResNet are CNN variants that have demonstrated very high performance in static gesture recognition. [43,3]. Hybrid CNN–LSTM architecture combining spatial and temporal modeling further enhanced dynamic sequence understanding for the same [3,1].

Transfer learning based on pre-trained models like InceptionV3, EfficientNet, and MobileNet provided better generalization performance on small ISL datasets [1,6]. 3D-CNNs and attention-based models captured spatiotemporal patterns in continuous signing [6, 3]. Researchers have also explored light-weight CNNs and MobileNet variants for real-time edge deployment [24,19].

Data augmentation and background subtraction techniques enhanced dataset diversity and the model's robustness [19]. Deep learning outperformed ML-based techniques in terms of accuracy. However, deep learning remains computationally intensive and highly dependent on large labelled datasets.

3.4 Graph-Based Approaches

In graph models, the gesture is represented by the spatial relationships between keypoints. The models used to represent the dependencies between the joints of the hand and body skeletons are

Graph Neural Networks, Graph Convolutional Networks, and Spatial–Temporal Graph Convolutional Networks.[2]. Such approaches can effectively capture spatiotemporal dynamics and outperform conventional CNNs in continuous sign recognition.

The researchers also combined the OpenPose or MediaPipe outputs within the graph models to improve the accuracy of node representation [2]. Improved signer-independent recognition was achieved by the attention-enhanced GCNs and hybrid architectures that combined GCN with LSTM or Transformer layers. Although robust, graph-based models are computationally expensive and sensitive to the precision of keypoint extraction.

3.5 Summary and Insight

Among all the methods, the performance of deep learning and graph methods is the best so far in ISL recognition, especially in continuous and signer-independent settings. As for the lightweight case, both traditional ML methods and geometric estimation methods are valuable. In the future, more attention should be paid to the multimodal fusion methods, the transformer methods, and the data standardization.

4. Datasets

Over the past decade, research in Indian Sign Language (ISL) recognition and translation has advanced rapidly, supported by progress in computer vision, deep learning, and multimodal data processing [55]. The creation of comprehensive and high-quality datasets has been fundamental to these developments, enabling models to capture the complex spatial, temporal, and linguistic characteristics of ISL.

However, when compared to other sign languages, ISL still does not have standardized, large-scale, and linguistically rich resources [8]. The available resources differ significantly in objectives, scope, and data modalities. This section describes the development of ISL resources, some of the important publicly available resources, and the methodologies used as well as the shortcomings of these resources. It highlights some of the shortcomings and research gaps, emphasizing the need to develop new and diverse ISL resources to take the field of ISLR and IT to the next level.

4.1 Overview of ISL Dataset Development

The research on Indian Sign Language (ISL) has transformed from small-scale gesture recognition studies to large-scale multimodal datasets with translation and linguistics goals. The initial studies were limited to static gestures like alphabet and digit recognition, collected under controlled conditions with limited signer variation [29].

With advances in computer vision and deep learning techniques, studies have expanded to dynamic and sentence-level datasets, incorporating modalities such as RGB video, skeletal pose, and textual glosses [25]. Nevertheless, in comparison with sign languages such as American Sign Language (ASL) or British Sign Language (BSL), ISL data are scarce in terms of scale, consistency, and availability.

4.2 Motion Tracking-Based Dataset Modalities

Motion-tracking approaches capture dynamic hand trajectories and body movements. Tools such as Kinect, Leap Motion, and MediaPipe have been used extensively to extract skeletal keypoints and 3D hand motion data [33,47]. Optical flow, Kalman filtering, and contour-tracking methods enabled precise movement detection for both single-handed and two-handed gestures [47,12].

Color segmentation, skin detection, and temporal motion energy images were integrated to further improve the extraction of motion features [12, 30]. These methods, combined with CNN or LSTM models, enabled accurate real-time temporal gesture recognition [30,27]. However, systems like these are sensitive to occlusions, illumination changes, and camera calibration errors [27,2].

4.3 Vision-Based Dataset Modalities

The most popular modality in the collection of sign language datasets, including ISL, is vision-based. These datasets are usually based on RGB video recordings and sometimes depth information. These vision-based modality datasets focus on collecting hand gestures, facial expressions, and upper-body movements. The availability of low-cost cameras and the ability to collect data in natural environments are significant factors in selecting a vision-based modality. [36, 26]

Vision-based datasets have been widely used in research on sign language recognition and translation due to their compatibility with deep learning architectures such as CNNs and RNNs. The effectiveness of vision-based input for learning spatiotemporal representations of continuous signing has been demonstrated by several studies using benchmark datasets of American and European sign languages. [26,16]. Similar approaches have been adopted in ISL research, where most publicly available datasets are collected using RGB video recordings

Despite their scalability, vision-based datasets face several challenges. Variations in illumination, background clutter, camera viewpoints, and occlusions can significantly affect recognition performance. Additionally, the accurate annotation of fine-grained hand movements and non-manual cues such as facial expressions and head movements remains a labor-intensive process [11]. Recent advances in pose estimation and keypoint extraction have partially mitigated these issues; however, their effectiveness is still limited in unconstrained recording environments.

Nevertheless, the most feasible and efficient modalities for creating ISL datasets are vision-based. This is due to these modalities' ability to capture not only manual signs but also non-manual signs, which are of great importance in the development of sign language recognition and translation systems. Therefore, the majority of current ISL datasets, as well as the proposed datasets, are vision-based, either exclusively or in combination with other modalities.

4.4 Existing Datasets

Table 2 below summarises key publicly available ISL datasets, highlighting their scope, scale, and distinctive characteristics.

Table 2. Existing Dataset Study

Dataset	**Scale**	**Type**	**Source**	**Key features**
ISL-CSLTR [6]	~Thousands	Video	Kaggle	Two-handed sentence signs
INCLUDE [4,44] (AI4Bharat)	4K+ videos	RGB + Pose	GitHub	Isolated words; MediaPipe pose data

ISLTranslate [18]	31K pairs	Text + Sign	GitHub	Continuous ISL↔English mapping
iSign Benchmark [23]	118K+ samples	Pose + Text	Hugging Face	Multimodal; gloss & recognition
HWGAT (2024) [40]	40K+ videos	RGB	arXiv	2,002 words; balanced signers
ISLAN [52]	~36 classes	Image + Video	Mendeley	Alphabets A–Z, digits 0–9
ISL – Real-Life Words [34][7]	~1K images	RGB Image	Mendeley	20 healthcare-specific signs

Although the resources exhibit significant progress, they differ in their objectives, data types, and coverage [48]. For example, INCLUDE and HWGAT are centred on pose-based and signer-balanced recognition, but ISLTranslate and iSign are centred on linguistic alignment and translation from text to sign.

Table 3 complements this comparison by summarizing, for each dataset, its public availability (with repository links) and its principal limitation, to make the trade-offs between these resources easier to assess at a glance.

Table 3. Comparative Analysis of ISL Dataset Availability and Limitations

Dataset	**Availability (Repository)**	**Vocabulary/Scope**	**Key Limitation**
ISL-CSLTR	Kaggle: kaggle.com/datasets/drblack00/isl-csltr-indian-sign-language-dataset	100 sentences; 7 signers	Small signer pool limits generalization across signing styles.
INCLUDE (AI4Bharat)	GitHub: github.com/AI4Bharat/INCLUDE	263 signs; isolated words only	Restricted to isolated single-word signs; no continuous or sentence-level data.
ISLTranslate	GitHub: github.com/Exploration-Lab/ISLTranslate	31K sentence/phrase pairs	Drawn mainly from educational videos, limiting topical and stylistic diversity.
iSign Benchmark	Project page: exploration-lab.github.io/iSign	118K+ video-text pairs	Aggregates prior datasets rather than new recordings, inheriting their annotation inconsistencies.

HWGAT (2024)	arXiv: arxiv.org/abs/2407.14224	2,002 words; balanced signers	Isolated-word scope; lacks sentence-level translation labels.
ISLAN	Mendeley Data (see Ref. [52])	36 classes (A–Z, 0–9)	Restricted to static alphabet and digit signs; no dynamic vocabulary.
ISL – Real-Life Words	Mendeley Data (see Refs. [34],[7])	20 healthcare-specific signs	A very small, domain-specific vocabulary limits generalizability.

4.5 Challenges and Limitations

Despite ongoing progress, several limitations persist in ISL dataset research:

- Limited Diversity: Most datasets involve a few signers, restricted backgrounds, and controlled environments, reducing generalization ability.
- Lack of Continuous Signing: Few datasets capture continuous sequences with coarticulation and contextual flow necessary for natural translation.
- Missing Non-Manual Cues: Facial expressions, gaze, and posture—integral components of ISL grammar are often unannotated.
- Inconsistent Annotation: Divergent glossing schemes and segmentation standards hinder cross-dataset comparisons.
- Restricted Accessibility: Some datasets have licensing barriers or incomplete documentation, limiting reproducibility.
- No Standardized Evaluation Protocol: Absence of unified benchmarks and signer-independent splits prevents consistent model assessment.

4.6 Identified Research Gaps

A synthesis of existing literature and datasets highlights key research gaps that motivate the need for a new ISL corpus:

- Absence of a Unified ISL Benchmark: There is no standardized, large-scale dataset supporting both recognition and translation across varied contexts
- Limited Multimodal Integration: Most datasets rely on RGB videos alone, lacking synchronized pose or depth data.
- Low Signer and Environmental Variability: Controlled lab recordings fail to represent natural variations in real-world communication.
- Weak Linguistic Alignment: Few corpora connect ISL glosses to English or Hindi translations, limiting downstream NLP-based tasks.
- Insufficient Real-Time and Temporal Data: Datasets rarely preserve the temporal continuity needed for dynamic gesture interpretation

5. Data Acquisition Methods for Sign Language Recognition

The performance of a sign language recognition (SLR) system heavily relies on the quality, variety, and quantity of the data supplied for training and evaluation purposes [9]. In recent years, with the advent of deep learning and computer vision technologies, data collection for sign languages has shifted from traditional sensor-based systems to modern multimodal, camera-based systems. Sign languages are complex visual-spatial communication systems used to convey meaning through manual gestures such as hand shape, movement, orientation, and placement, as well as non-manual markers such as facial expressions, head positions, and body postures [14]. The correct interpretation of such complex gestures requires effective data collection systems to accurately capture visual cues.

Over the last two decades, various data acquisition techniques have been studied depending on hardware availability, specific sign languages (e.g., ASL, ISL, BSL), and specific goals such as word, sentence, or continuous signing. In terms of ISL, one of the challenges associated with dataset acquisition is diversity, standardization, as well as publicly available datasets.

Five major data acquisition techniques are commonly used in sign language recognition:
● Glove-based systems
● Armband and sensor-based systems
● Leap Motion-based tracking
● Kinect or depth sensor-based systems
● Vision-based systems

Each of these methods has specific advantages in terms of precision, ease of use, and cost, but also specific problems concerning user comfort, environmental dependence, and scalability. Fig. 3 presents some examples of data acquisition setups.

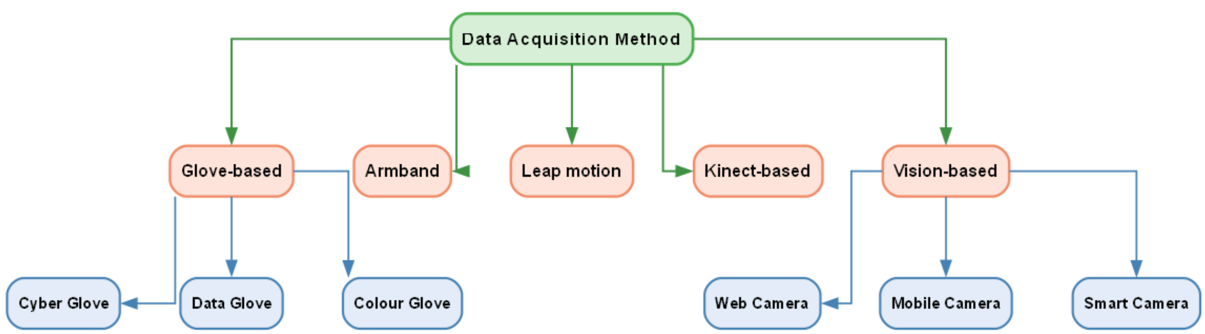

Fig 3. Data Acquisition Method

The setups illustrated in Fig. 3 span the accuracy-versus-practicality trade-off discussed throughout Section 5: sensor- and glove-based rigs at one end offer high measurement precision at the cost of user comfort and wearability, while vision-based recording at the other end is low-cost and non-intrusive but more sensitive to lighting and background variation.

5.1 Glove-Based Systems

Of these, glove-based systems are one of the earliest and most commonly used techniques in SLR research in various forms, as shown in Fig. 4. The basic idea of this technique is that the hands are covered with data gloves, cyber gloves, or even colour gloves, which are used to collect small motion data. These gloves are typically equipped with flex sensors, accelerometers, gyroscopes, or even IMUs to measure finger flexion, angular motion, wrist rotation, and acceleration patterns.

According to Kumar et al. [15], glove-based systems have been shown to efficiently track the articulatory parameters necessary for recognising static and dynamic gestures. In this regard, the cyber glove tracks the finger joint angles, while the colour glove is effective for hand region identification using vision-based tracking.

The major advantage of this method is its high precision, as the data collected by the sensors are unaffected by lighting or background conditions. Another advantage of glove-based systems is their real-time, low-latency capability, which is advantageous in the lab setting.

However, they also present several limitations:

- Users must wear the gloves, reducing naturalness and comfort during signing.
- Calibration is often needed for each user.
- Hardware costs are high for large-scale data collection.
- Gloves can restrict fine hand movements, reducing expressivity.

Despite these drawbacks, glove-based systems remain valuable for creating benchmark datasets and studying isolated signs in constrained settings.

Fig 4. (a) Cyber glove, (b) data glove, (c) colour glove, (d) Leap Motion device and (e) sensors Armband.

5.2 Armband and Motion Sensor Systems

Armband-based systems are an advancement in wearable technology that incorporate electromyography and motion sensing. These are used to detect muscle activities and movement patterns. The armband technology, such as the Myo armband, is used in SLR to detect muscle tension and rotational angles during signing [31].

Such sensors play an important role in SLR because they can record physiological data that would otherwise be difficult to obtain through vision-based sensing techniques. The data include the degree of muscle contraction and the position of the arms relative to the body. Such data is necessary for understanding dynamic and small gestures.

The primary advantages include:

- Robustness against visual occlusion and lighting variations.
- Capability to detect gestures even in non-visual environments.
- Compactness and ease of wireless communication.

However, they also have notable disadvantages:

- Limited comfort due to wearable constraints.
- Reduced sensitivity to fine finger articulations.
- High cost and dependency on device calibration.

Consequently, armband systems are often combined with vision-based methods to achieve hybrid, multimodal recognition.

5.3 Leap Motion-Based Systems

The Leap Motion controller is an infrared optical device designed to track hand movements in 3D space, capturing both position and orientation data for hands and fingers. It operates using two infrared cameras and three LEDs to detect hand motion within a range of about 60 cm. [17]

This technology can track precise hand and finger movements without the need for any markers or physical contact. It has been observed that systems using Leap Motion technology are particularly useful for finger spelling and small gesture detection due to their precise capabilities, which enable 3D data capture. However, it has limitations in its small area of interaction and sensitivity to light, making it unsuitable for large or outdoor environments. It does not detect facial or body language, which is crucial in understanding the expressions used in Indian Sign Language (ISL).

5.4 Kinect-Based and Depth-Sensor Systems

Microsoft's Kinect sensor introduced a significant shift in SLR data collection by combining RGB video with depth perception. The depth sensor projects infrared light and measures the time it takes to reflect off the signer's body, producing a 3D point cloud representation of the signer's body [45].

Kinect technology can detect human skeletal points, such as the hand, elbow, shoulder, and head joints. This technology is quite useful in continuous sign language recognition and segmentation, where the movement flow is an important aspect.

Advantages include:

- High accuracy in detecting 3D motion.
- Real-time tracking of multiple users.
- Useful in varying illumination conditions compared to RGB-only systems.

However, Kinect sensors also have some limitations:

- Limited spatial resolution for small finger gestures.
- Depth sensing errors beyond 4–5 meters.

- Sensitivity to reflective surfaces or sunlight interference.

Despite these issues, Kinect data has contributed substantially to public sign language datasets such as RWTH-PHOENIX-Weather and ASLLVD, and is gradually being adopted for ISL research.

5.5 Vision-Based Systems

Among the various data acquisition schemes, vision-based approaches have been recognized as the most preferred method of ISL recognition in the current era due to their non-intrusive, cost-effective, and realistic aspects. Using RGB or RGB-D cameras, the user's natural signing is recorded without requiring the user to wear any sensors/devices.

The manual components include hand movements, hand orientation, hand location, and hand shape, while the non-manual components include facial expressions, head movements, shoulder movements, and gaze direction, among others. The development of convolutional neural networks (CNNs) and the application of deep learning methods enable the automatic learning of relevant features from images captured by the system, thereby eliminating the need to design features derived from these images [6-9].

Challenges in Vision-Based Acquisition

While natural and scalable, vision-based systems are influenced by several environmental and human factors, such as:

- Variations in lighting and background clutter.
- Occlusion between hands and face.
- Skin color diversity among signers.
- Differences in signing speed and style.

To overcome these challenges, preprocessing steps such as background subtraction, skin-colour segmentation, optical flow estimation, and pose extraction are commonly used before model training.

Applications and Research Examples

Ahmed et al. [54] presented a two-handed dynamic sign language recognition system that achieved 90% accuracy using trajectory-based feature extraction and Dynamic Time Warping (DTW). Rehman et al. [28] presented a real-time Bengali sign language recognition system using Haar-based features with 98.17% accuracy on vowels.

Recent ISL works also integrate color gloves for better hand tracking, hybridizing sensor and vision-based data. This combined approach allows more accurate segmentation in complex lighting and provides complementary spatial information [42–17].

Comprehensive comparative studies of both Sensor-based and Vision-based approaches have been presented by Amrutha et al. [10]. The importance of using vision-based acquisition has been highlighted because it is more naturalistic and inclusive, particularly for large-scale ISL datasets.

5.6 Emerging Approaches in Data Acquisition

In light of the increasing relevance of diversity and generalization in datasets, modern data acquisition techniques have started making use of crowd-sourcing for collecting data using services such as Amazon Mechanical Turk (MTurk), YouTube, and mobile apps.

This ensures that the obtained data includes indications from different subjects with diverse backgrounds, illumination conditions, and camera configurations. On the other hand, there has been increased adoption of multimodal acquisition technologies that capture RGB video, depth data, and IMU signals due to their effectiveness in capturing comprehensive gesture data for continuous and isolated sign language identification tasks.

6. Curation of Dataset and Proposed Integrated Framework

In order to solve the problems associated with the constraints posed by the existing Indian Sign Language (ISL) datasets and recognition systems, which consist of limited vocabulary range, the absence of a standardized dataset, and the inability to implement a standardized framework for real-time processing, this paper presents a novel framework known as Beyond. This new framework combines the procedures for dataset creation, translation, and sign language recognition via visual means into a single framework. The novel technique avoids the conventional approach, in which both procedures are conducted independently, and introduces an interactive pipeline that can run in both directions. Fig. 5 presents the overall workflow of the Beyond framework, showing how dataset curation, vision-based recognition, bidirectional translation, and the accessibility-driven user interface are connected through a continuous feedback loop rather than operating as isolated stages.

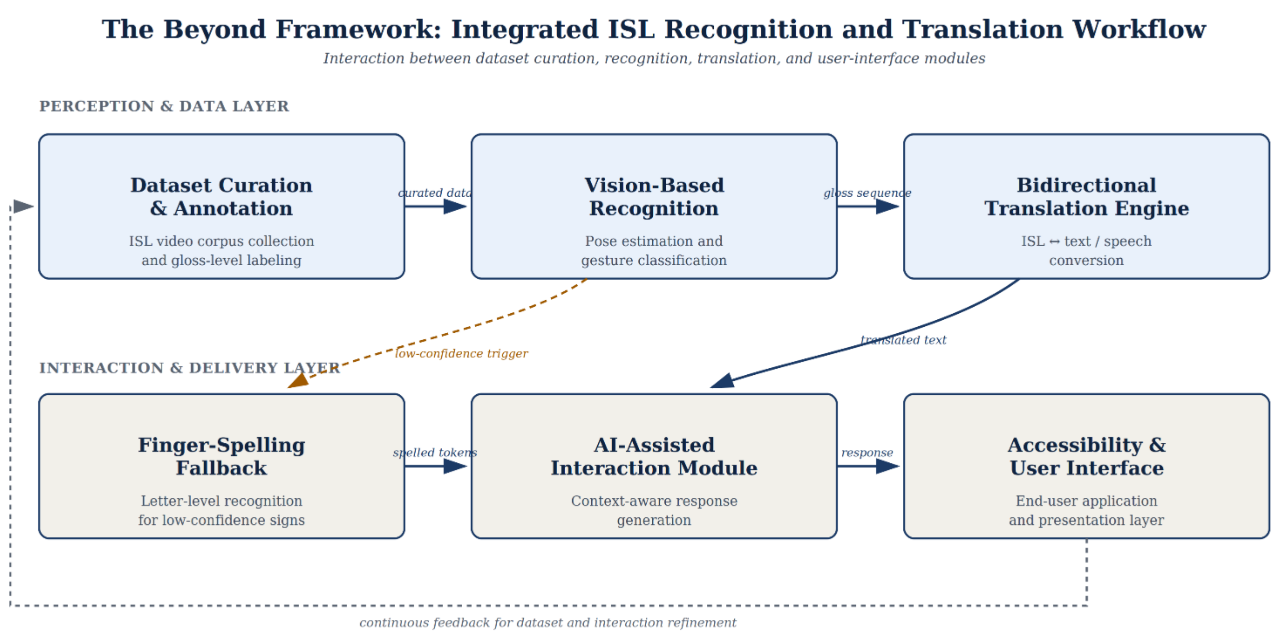

Fig 5. Interaction Between Dataset Curation, Recognition, Translation, and User Interface Modules in the Beyond Framework

6.1 Dataset Curation, Integration, and Annotation Strategy

A key component of the proposed framework is the design of a hybrid dataset that enables both translation and recognition tasks. The hybrid data set is developed by combining publicly available gesture data sets with an annotation framework designed specifically for continuous ISL translation.

In the case of the ISL text-and-speech module, more than one publicly available dataset is used, made available on websites such as Kaggle. The datasets provide data on gestures used for the alphabet, digits, and individual words. Nonetheless, these are always not very diverse or consistent. In an effort to resolve this issue, these datasets are merged into a common corpus after undergoing pre-processing processes such as label normalization, filtering of duplicates and inconsistencies in annotations, and so forth.

More specifically, merging proceeds in three stages. First, gloss and word-level labels from each source dataset are mapped onto a single controlled vocabulary; labels that describe the same gesture under different spellings, casing, or regional naming conventions (for example, synonymous glosses for the same sign) are unified under one canonical tag, while entries that cannot be confidently mapped are flagged and excluded rather than guessed. Second, when two source datasets provide conflicting or overlapping annotations for what is nominally the same sign, the conflict is resolved by preferring the source with the larger signer pool and more detailed timestamp metadata, while the discarded annotation is retained in a supplementary log for future audit rather than being silently dropped. Third, annotation consistency across the merged corpus is checked through spot-verification of a random sample of segments against their source video, together with automated checks for timestamp overlaps, duplicate segments, and label-length mismatches; segments that fail these checks are routed back for manual review before being admitted into the common corpus. This staged process is what allows datasets collected under different protocols to be combined without silently propagating their individual annotation errors into the merged corpus.

For the ISL text and speech module, the dataset is built specifically using carefully selected ISL video clips. Rather than using existing pre-split gestures, timestamp annotation is adopted. The continuous ISL videos are split into appropriate segments by specifying the start and end points of each gesture. Each segment consists of annotated information, including the textual label and the time interval during which the label occurs in the video.

This approach enables dynamic extraction of gesture segments at runtime, eliminating the need for storing large volumes of individual video files. It also allows efficient reuse of source videos and simplifies dataset expansion by appending new annotations without modifying existing data. The combined dataset design ensures coverage for both recognition and translation while maintaining flexibility and scalability.

6.2 Bidirectional Translation Framework

The proposed system employs a bidirectional translation approach, which will enable transliteration between ISL and the spoken language. In this respect, there will be two separate but related processing pipelines: the transliteration of text and speech into ISL, and the transliteration from ISL into text and speech.

6.2.1 Text and Speech to ISL Translation Pipeline

Conversion of input into gestures in the ISL pipeline occurs via text and speech. Inputs can be given in two languages: English and Hindi. If the input is in Hindi, a neural translator is employed to translate it into English, as the labels for the videos in the dataset are in English. Speech recognition software is employed to transcribe speech into text.

The transcribed or input text is tokenized into individual words, which are then mapped to corresponding ISL gestures using the custom annotated dataset. Each word is associated with a specific video segment through timestamp indexing. During execution, the system retrieves the relevant segments from continuous video sources and renders them sequentially to produce the final ISL output.

The implementation of timestamp-based search guarantees temporal consistency among gestures and facilitates efficient reuse of video content. Moreover, the approach allows for the gradual increase of the size of the data set and minimizes the amount of memory used in comparison with conventional clip-based methods.

6.2.2 ISL to Text and Speech Translation Pipeline

ISL to text and speech translation is achieved through the process of inverse translation, where the hand gestures are translated to text and speech forms. This step utilizes the data set collected through different sources. The variety of the data set enhances the generalization capability of the model.

The system collects a live video feed from a camera and uses it to derive spatio-temporal features of hand gestures. The recognition engine then uses these features to learn the class labels of the gestures. The class labels can be viewed as texts and are converted to speech by the text-to-speech converter.

To enhance usability, a contextual prediction mechanism is integrated into the pipeline. Based on the recognized sequence, the system suggests probable subsequent characters or words. This reduces input effort and improves the overall interaction efficiency. The combination of multi-source datasets and real-time processing enables reliable performance in practical environments.

6.3 Robust Handling of Vocabulary Limitations

To address the issue of insufficient vocabulary mapping, the system uses an alternative approach called finger spelling. Words that cannot be found within the database will be broken down into characters, which in turn will be converted into its equivalent in the ISL alphabet.

This approach ensures that the system can produce an output for any input without breaking down. This greatly increases the system's robustness and enables it to cope with unusual or previously unknown words.

6.4 Vision-Based Recognition with Adaptive Preprocessing

The hand gesture recognition system is designed for use under various conditions and uses pre-processing techniques to improve accuracy and robustness. This is achieved by simplifying the background, removing noise, and segmenting the region of interest within the scene. Pre-processing is performed under the assumption that the background is uniform.

Preprocessing ensures that the feature extraction process is independent of changes in lighting and background conditions. After processing the input, spatial and temporal features are generated that are used by the recognition algorithm during classification. The system's output is provided in text form.

The integration of preprocessing with recognition improves overall system reliability and enables consistent performance across different usage scenarios.

6.5 Unified System Architecture and Interaction Model

The implementation of this system will require the adoption of a single integrated architecture in which dataset management, translation, and recognition operations take place. The system will be designed to facilitate real-time communication by supporting both input and output processes.

Both translation directions share common data structures and resources while maintaining independent processing pipelines. This modular yet integrated design allows efficient communication between components and simplifies system maintenance. The architecture supports scalability, enabling future extensions such as additional datasets or improved recognition models.

The interaction model is designed to maintain synchronization between input and output modalities, ensuring that gesture sequences and textual representations remain aligned during operation.

6.6 AI-Assisted Interaction Module

An artificial intelligence-driven interaction module has been added to assist users and improve the application's usability. This module allows for bilingual communication between the user and the application in both English and Hindi. The user queries are processed using a language model, and dynamic responses are provided back to the user.

This module enhances user interaction by providing guidance related to ISL usage and system functionality. It operates as a supplementary component that improves accessibility without interfering with the core translation and recognition pipelines.

6.7 Key Contributions of the Proposed Framework

The Beyond system being developed would have implications on the area of ISL translation and recognition through its use of a hybrid dataset consisting of multi-source public datasets and timestamps from videos, incorporation of a real-time bi-directional translation framework for both languages at once, implementation of a reliable fallback technique for finger spelling in case of unknown vocabulary, use of a gesture recognition system based on vision with adaptive preprocessing, and the creation of a unified architecture that allows for interactive and real-time performance.

7. Discussion and Future Discussion

Deep learning and advances in computer vision have led to tremendous progress in gesture and motion recognition, which is having a major impact on human-computer interaction as well as sign language recognition. A robust real-time SLR system is highly useful in the Deaf and Hard of Hearing community, as it plays an important role in communication with people who hear. Despite tremendous research progress, the goal of real-time ISL recognition with good accuracy remains elusive.[41].

This paper reviewed research on Indian Sign Language related to datasets, recognition, and translation, starting from 2011 onwards. The main conclusions derived are:

- Of these, about 92.59% of studies address isolated signs, while only 7.41% focus on continuous interpretation.
- Static sign recognition dominates with 65.15% as opposed to dynamic gestures with 34.85%.

- DL-based models (CNN–LSTM, ~98% accuracy [31] are ahead of ML-based systems such as HMM (~91% [57]).
- Signer variability, changes in lighting, background, and viewpoint make ISL recognition challenging [56].
- Publicly available ISL datasets are few, so most researchers develop their own data for their systems.
- Sentence-level ISL datasets are small, with limited vocabulary and weak generalization.

Overall, ISL research is still dataset-constrained, especially for continuous signing. There is a need for large-scale annotated corpora, such as RWTH-PHOENIX-Weather-2014T [51]. Datasets could be built with the support of platforms such as ISH News, which provides timestamped ISL videos. ISL dictionary resources can also contribute toward sentence-level dataset creation.

Based on these observations, the Beyond framework attempts to solve a number of recognized problems by designing an integrated system. In contrast to existing models, which mainly concentrate on individual recognition, the model incorporates not only the processes of translation and recognition, but also allows for constant interaction. With a mixed dataset comprising public gesture datasets and time-stamped annotations, the scalability of the dataset-building process is ensured.

The division of pipelines for translating from text/speech to ISL and from ISL to text/speech enables efficient use of task-specific datasets. Using timestamps as a search function in gesture mapping provides more flexibility than the conventional dataset structure. Aggregation of datasets for recognizing gestures increases generalization in the presence of variations in sign language usage. Background normalization is among other preprocessing measures that improve robustness.

A fallback mechanism based on finger-spelling addresses incomplete vocabulary coverage, ensuring continuity in translation. Additionally, integrating bilingual processing with real-time interaction improves the system's practical applicability.

Although these advancements have been made, some issues remain. The existing model is based on the word-level approach, thus failing to incorporate grammar and context in continuous ASL sentences. In the future, efforts could be made to develop a deep learning model for sequence-level translation, such as a transformer.

Some other improvements include scaling up the dataset for handling multiple sample kinds, ranging from varied regions and continuous indicators. Multimodal input mechanisms, such as depth sensing and posture estimation, may aid in recognising the voice more efficiently. It is necessary not to neglect the real-time implementation of resource-limited systems and the enhancement of resistance to speech variation in further scientific research. End-to-end learning algorithms can be employed to accomplish recognition and translation simultaneously.

7.1 Practical Applications

Beyond the research contributions discussed, the framework is intended for deployment in settings where ISL interpretation is currently scarce. In education, bidirectional translation and predictive text features could support Deaf and Hard-of-Hearing students in mainstream classrooms that lack a dedicated ISL interpreter, allowing spoken or written lecture content to be rendered into ISL and student signing to be converted back into text for the instructor. In healthcare, the finger-spelling fallback and the ISL-to-text/speech pipeline could assist patient-provider communication during consultations, particularly for domain-specific vocabulary that may not be covered by general-purpose ISL datasets, echoing the healthcare-specific signs already represented in resources such as the ISL - Real-Life Words dataset. In public service settings such as banks, government offices, and transportation hubs, the real-time recognition and speech-output components could function as an assistive kiosk or mobile companion, reducing dependence on the limited pool of certified ISL interpreters noted in Section 3. These use cases also indicate where future validation should be prioritized, since classroom, clinical, and public-service environments each introduce acoustic and visual conditions that differ from the controlled settings in which most existing ISL datasets were collected.

7.2 Limitations

Alongside the word-level and grammar-related limitations already noted, three further constraints affect the current framework:

- Dataset dependency: recognition accuracy and vocabulary coverage are bounded by the public datasets merged into the hybrid corpus (Section 6.1); regional ISL variation not represented in these source datasets will not be recognized reliably, and the finger-spelling fallback, while functional, does not fully substitute for signer-fluent coverage.
- Computational cost of multimodal processing: combining vision-based recognition, adaptive preprocessing, bidirectional translation, and the AI-assisted interaction module concurrently increases inference latency and memory footprint relative to a single-task model, which may constrain deployment on low-power or offline edge devices.
- Real-time deployment at scale: as the mapped vocabulary grows, timestamp-indexed retrieval and gesture-segment rendering must search a larger annotation space under real-time latency constraints, and maintaining signer-independent accuracy across this larger vocabulary in uncontrolled, real-world recording conditions remains an open challenge that has not yet been evaluated at production scale.

8. Conclusion

This paper provides a review of datasets, techniques, and technologies used for the recognition and translation of Indian Sign Language (ISL). Some major drawbacks of the existing methods have been identified in this analysis, including inadequate vocabulary, non-standardised datasets, and an inability to provide real-time solutions. This clearly indicates the necessity of integration of theoretical concepts with practical application.

To overcome these shortcomings, this paper presents a comprehensive ISL model called "Beyond," which integrates data set generation, two-way translation, and gesture recognition from visual input into a single coherent design. The design presented herein utilizes a mixed data set composed of existing public gesture data sets and a new time-stamped video data labeling method for scalability and flexibility of gesture mapping. The two-way translation module facilitates translations of text and speech to ISL and vice versa.

The use of adaptive preprocessing methods in the recognition process enhances the system's robustness across different environmental conditions. The inclusion of the backup method using finger-spelling makes the system capable of dealing with unknown vocabulary effectively. Moreover, the system is designed to perform bilingual processing and real-time tasks.

To conclude, it is evident that by combining approaches and models across dataset construction, the translation process, and the recognition model, substantial progress can be made in the development of ISL technologies. As such, the presented system is designed to bridge the gap between the research-based approach and its application, facilitating efficient communication within the Deaf and Hard-of-Hearing community. Beyond its research contribution, the framework is intended for practical deployment in education, healthcare, and public service settings, where it can support classroom interpretation, patient-provider communication, and interpreter-independent access to public services, as discussed in Section 7.1.

9. References

[1] V. Adithya, P. R. Vinod, and U. Gopalakrishnan, "Artificial neural network based method for Indian sign language recognition," in *Proc. IEEE Conf. Information and Communication Technologies (ICT)*, Apr. 2013, pp. 1080–1085.

[2] V. Adithya and R. Rajesh, "A deep convolutional neural network approach for static hand gesture recognition," *Procedia Computer Science*, vol. 171, pp. 2353–2361, 2020.

[3] S. C. Agrawal, A. S. Jalal, and C. Bhatnagar, "Recognition of Indian Sign Language using feature fusion," in *Proc. 4th Int. Conf. Intelligent Human Computer Interaction (IHCI)*, Dec. 2012, pp. 1–5.

[4] W. Ahmed, K. Chanda, and S. Mitra, "Vision based hand gesture recognition using dynamic time warping for Indian sign language," in *Proc. Int. Conf. Information Science (ICIS)*, 2016, pp. 120–125.

[5] K. Amrutha and P. Prabu, "A comparative study on Indian sign language representation," in *Proc. Int. Conf. Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, Oct. 2021, pp. 1–6.

[6] G. Anantha Rao, P. V. V. Kishore, A. S. C. S. Sastry, D. Anil Kumar, and E. Kiran Kumar, "Selfie continuous sign language recognition with neural network classifier," in *Proc. 2nd Int. Conf. Micro-Electronics, Electromagnetics and Telecommunications (ICMEET 2016)*, Singapore: Springer, 2017, pp. 31–40.

[7] G. Arun Prasath and K. Annapurani, "A review on deaf and dumb communication system based on various recognitions aspect," in *Proc. Int. Conf. Deep Learning, Computing and Intelligence (ICDCI 2021)*, Singapore: Springer Nature, 2022, pp. 191–203.

[8] G. Arun Prasath and K. Annapurani, "Prediction of sign language recognition based on multi layered CNN," *Multimedia Tools and Applications*, vol. 82, no. 19, pp. 29649–29669, 2023.

[9] M. K. Bhuyan, "FSM-based recognition of dynamic hand gestures via gesture summarization using key video object planes," *International Journal of Computer and Communication Engineering*, vol. 6, pp. 248–259, 2012.

[10] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10023–10033.

[11] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7291–7299.

[12] A. Dumbre, S. Jangada, S. Gosavi, and J. Gupta, "Classification of Indian sign language characters utilizing convolutional neural networks and transfer learning models with different image processing techniques," in *Proc. IEEE World Conf. Applied Intelligence and Computing (AIC)*, Jun. 2022, pp. 423–430.

[13] E.-S. M. El-Alfy and H. Luqman, "A comprehensive survey and taxonomy of sign language research," *Engineering Applications of Artificial Intelligence*, vol. 114, p. 105198, 2022.

[14] R. Elakkiya and K. Selvamani, "Enhanced dynamic programming approach for subunit modelling to handle segmentation and recognition ambiguities in sign language," *Journal of Parallel and Distributed Computing*, vol. 117, pp. 246–255, 2018.

[15] R. Elakkiya and K. Selvamani, "Subunit sign modeling framework for continuous sign language recognition," *Computers and Electrical Engineering*, vol. 74, pp. 379–390, 2019.

[16] J. Forster, C. Schmidt, O. Koller, M. Bellgardt, and H. Ney, "Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-Weather," in *Proc. Language Resources and Evaluation Conf. (LREC)*, May 2014, pp. 1911–1916.

[17] A. S. Ghotkar and G. K. Kharate, "Dynamic hand gesture recognition and novel sentence interpretation algorithm for Indian sign language using Microsoft Kinect sensor," *Journal of Pattern Recognition Research*, vol. 1, pp. 24–38, 2015.

[18] R. Gogoi, P. Kumar, and R. Damdoo, "End to end simple Indian sign language sentence translation using sign transformer network," in *Proc. 15th Int. Conf. Computing Communication and Networking Technologies (ICCCNT)*, Jun. 2024, pp. 1–13.

[19] V. Gomathi, "Indian sign language recognition through hybrid ConvNet-LSTM networks," *EMITTER International Journal of Engineering Technology*, vol. 9, no. 1, pp. 182–203, 2021.

[20] R. Gupta and S. Rajan, "Comparative analysis of convolution neural network models for continuous Indian sign language classification," *Procedia Computer Science*, vol. 171, pp. 1542–1550, 2020.

[21] M. Hasan, T. H. Sajib, and M. Dey, "A machine learning based approach for the detection and recognition of Bangla sign language," in *Proc. Int. Conf. Medical Engineering, Health Informatics and Technology (MediTec)*, Dec. 2016, pp. 1–5.

[22] P. Jadav, "Indian sign language recognition system," *International Journal of Engineering and Technology*.

[23] A. Joshi et al., "iSign: A benchmark for Indian sign language processing," in *Findings of the Assoc. Computational Linguistics (ACL 2024)*, Aug. 2024, pp. 10827–10844.

[24] G. Joshi, A. Gaur, and Sheenu, "Interpretation of Indian sign language using optimal HOG feature vector," in *Proc. Int. Conf. Advances in Computing and Data Sciences*, Singapore: Springer, Apr. 2018, pp. 65–73.

[25] P. V. V. Kishore, M. V. D. Prasad, D. A. Kumar, and A. S. C. S. Sastry, "Optical flow hand tracking and active contour hand shape features for continuous sign language recognition with artificial neural networks," in *Proc. IEEE 6th Int. Conf. Advanced Computing (IACC)*, Feb. 2016, pp. 346–351.

[26] O. Koller, J. Forster, and H. Ney, "Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers," *Computer Vision and Image Understanding*, vol. 141, pp. 108–125, 2015.

[27] D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman, and S. A. Bahaj, "SIGNFORMER: deepvision transformer for sign language recognition," *IEEE Access*, vol. 11, pp. 4730–4739, 2023.

[28] L. Krishnan, "A systematic literature review on the robustness of sign language recognition methods in low-light environments," *Authorea Preprints*, 2024.

[29] E. K. Kumar, P. V. V. Kishore, A. S. C. S. Sastry, M. T. K. Kumar, and D. A. Kumar, "Training CNNs for 3-D sign language recognition with color texture coded joint angular displacement maps," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 645–649, 2018.

[30] E. K. Kumar, P. V. V. Kishore, M. T. K. Kumar, and D. A. Kumar, "3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2-stream CNN," *Neurocomputing*, vol. 372, pp. 40–54, 2020.

[31] E. K. Kumar, P. V. V. Kishore, D. A. Kumar, and M. T. K. Kumar, "Early estimation model for 3D-discrete Indian sign language recognition using graph matching," *Journal of King Saud University – Computer and Information Sciences*, vol. 33, no. 7, pp. 852–864, 2021.

[32] P. Kumar, R. Saini, S. K. Behera, D. P. Dogra, and P. P. Roy, "Real-time recognition of sign language gestures and air-writing using leap motion," in *Proc. 15th IAPR Int. Conf. Machine Vision Applications (MVA)*, Nagoya, Japan, May 2017, pp. 157–160.

[33] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A modified LSTM model for continuous sign language recognition using leap motion," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7056–7063, 2019.

[34] M. Nakrani, "A review paper on sign language recognition system for deaf and dumb people using image processing," *International Journal of Engineering Research and Applications*, 2016.

[35] P. A. Nanivadekar and V. Kulkarni, "Indian sign language recognition: database creation, hand tracking and segmentation," in *Proc. Int. Conf. Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, Apr. 2014, pp. 358–363.

[36] S. C. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873–891, Jun. 2005.

[37] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: a review of techniques," *Journal of Imaging*, vol. 6, no. 8, p. 73, 2020.

[38] C. Oz and M. C. Leu, "Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker," *Neurocomputing*, vol. 70, no. 16–18, pp. 2891–2901, 2007.

[39] D. U. Patel and J. M. Joshi, "Review of Indian dynamic sign language recognition system," in *Proc. Int. Conf. Communication and Computational Technologies (ICCCT 2022)*, Singapore: Springer Nature, 2022, pp. 17–31.

[40] S. Patra, A. Maitra, M. Tiwari, K. Kumaran, S. Prabhu, S. Punyeshwarananda, and S. Samanta, "Hierarchical windowed graph attention network and a large scale dataset for isolated Indian sign language recognition," *arXiv preprint arXiv:2407.14224*, 2024.

[41] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," *Expert Systems with Applications*, vol. 164, p. 113794, 2021.

[42] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, "Real-time computer vision-based Bengali sign language recognition," in *Proc. 17th Int. Conf. Computer and Information Technology (ICCIT)*, Dec. 2014, pp. 192–197.

[43] J. Rekha, J. Bhattacharya, and S. Majumder, "Shape, texture and local movement hand gesture features for Indian sign language recognition," in *Proc. 3rd Int. Conf. Trendz in Information Sciences and Computing (TISC)*, Dec. 2011, pp. 30–35.

[44] S. Reshna and M. Jayaraju, "Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and SVM," in *Proc. Int. Conf. Wireless Communications, Signal Processing and Networking (WiSPNET)*, Mar. 2017, pp. 386–390.

[45] E. J. Robert and H. J. Duraisamy, "A review on computational methods based automated sign language recognition system for hearing and speech impaired community," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 9, p. e7653, 2023.

[46] B. Samal and M. Panda, "Integrative review on vision-based dynamic Indian sign language recognition systems," in *Proc. 1st Odisha Int. Conf. Electrical Power Engineering, Communication and Computing Technology (ODICON)*, 2021, pp. 1–6.

[47] A. Sharma, N. Sharma, Y. Saxena, A. Singh, and D. Sadhya, "Benchmarking deep neural network approaches for Indian Sign Language recognition," *Neural Computing and Applications*, vol. 33, no. 12, pp. 6685–6696, 2021.

[48] K. Shenoy, T. Dastane, V. Rao, and D. Vyavaharkar, "Real-time Indian sign language (ISL) recognition," in *Proc. 9th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2018, pp. 1–9.

[49] A. Shinde and R. Kagalkar, "Sign language to text and vice versa recognition using computer vision in Marathi," *International Journal of Computer Applications*, vol. 975, p. 8887, 2015.

[50] A. Shinde and R. Kagalkar, "Advanced Marathi sign language recognition using computer vision," *International Journal of Computer Applications*, vol. 118, no. 13, 2015.

[51] O. M. Sincan and H. Y. Keles, "AUTSL: A large scale multi-modal Turkish sign language dataset and baseline methods," *IEEE Access*, vol. 8, pp. 181340–181355, 2020.

[52] A. Sivan and C. Eswaran, "Indian sign language alphabet recognition and speech synthesis using a hybrid deep learning approach," *Archives of Acoustics*, 2026.

[53] N. V. Tavari, A. V. Deorankar, and P. N. Chatur, "Hand gesture recognition of Indian sign language to aid physically impaired people," *International Journal of Engineering Research and Applications*, vol. 5, p. 60, 2014.

[54] K. Tripathi, N. Baranwal, and G. C. Nandi, "Continuous dynamic Indian Sign Language gesture recognition with invariant backgrounds," in *Proc. Int. Conf. Advances in Computing, Communications and Informatics (ICACCI)*, Aug. 2015, pp. 2211–2216.

[55] M. Varsha and C. S. Nair, "Indian sign language gesture recognition using deep convolutional neural network," in *Proc. 8th Int. Conf. Smart Computing and Communications (ICSCC)*, Jul. 2021, pp. 193–197.

[56] M. Vázquez-Enríquez, J. L. Alba-Castro, L. Docío-Fernández, and E. Rodríguez-Banga, "Isolated sign language recognition with multi-scale spatial-temporal graph convolutional networks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 3462–3471.

[57] A. Wadhawan and P. Kumar, "Deep learning-based sign language recognition system for static signs," *Neural Computing and Applications*, vol. 32, no. 12, pp. 7957–7968, 2020.