Use Case 8: AI-Powered Indian Sign Language (ISL) Detection and Translation

Organization: International Institute of Information Technology, Hyderabad.

Country: India

Contact Person:

Pandey, Aishani, aishani.pandey@research.iiit.ac.in
Sachdeva, Arush, arush.sachdeva@research.iiit.ac.in
Mathur, Vivek-vivekofficialwork1@gmail.com

1 Use Case Summary Table

Item	Details
Category	Accessibility
Problem Addressed	There is a big communication gap between people who speak ISL and those who don't. For the deaf community, social integration and information access are hampered by the absence of real-time translation tools.
Key Aspects of Solution	Data Acquisition: Gather and preprocess Indian Sign Language(ISL) datasets from a variety of pertinent and helpful open-source sources, such as expert-curated databases, publicly accessible sign language resources, and YouTube tutorials. To ensure accurate representation, specific context, vocabulary, phrases, and videography protocols related to the ISL context will be applied during data curation. Tech Stack: Python (Pandas, Open Computer Vision (OpenCV)), YouTube Application Programming Interface (API), Web Scraping (BeautifulSoup), Video Annotation Tools (e.g., Visual Geometry Group (VGG) Image Annotator) Data Enhancement: Augment data for low-resource variations using pose estimation, synthetic data generation, and adversarial augmentation techniques to improve generalization. Tech Stack: TensorFlow, PyTorch, OpenPose, Generative Adversarial Networks (GANs) for augmentation

(continued)

Item	Details
	Model Development: Use either the ensemble of current Convolutional Neural Network (CNN)/Transformer models or by identifying key characteristics from each and creating a new model to implement a lightweight, real-time ISL recognition model that is optimised for efficiency on edge devices, guaranteeing a seamless user experience and low latency. We can use language models to help us preserve the emotion and context. Recent advancements in continuous sign language recognition, like the use of motor attention mechanisms [1] and multi-scale feature enhancement [2], suggest potential improvements in real-time detection. A recent study by Hirooka et al. [3] demonstrates the use of Stack Transformer models with spatial-temporal attention for dynamic multi-culture sign language recognition, which can offer significant improvements in dynamic ISL detection by preserving temporal context and motion features. Tech Stack: TensorFlow, PyTorch, OpenCV, CNNs (ResNet, VGG), Transformer Models (Bidirectional Encoder Representations from Transformers(BERT)), Generative Pre-trained Transformer(GPT)), Keras, MobileNet for real-time inference on edge devices Intermediate Representation: Extract intermediate representations in latent and embedding space, such as: – Skeletal keypoint sequences for tracking of user. – Encoding for mapping signs to linguistic structures. – Temporal feature representation to capture the timing and transitions between signs using detection of hand movements. – Facial feature embedding to track and preserve emotion and expression (with de-personification to ensure privacy). Tech Stack: OpenPose, Dlib (for facial feature extraction), TensorFlow/Keras (for embedding), PyTorch (for dynamic embeddings) Privacy-Preserving Mechanisms: Ensure anonymization of data using pose estimation for skeletal key points and avoid using facial or personal identifiers. Encrypt data during transmission to prevent unauthorized access. Tech Stack: Cryptography Libraries (PyCryptodome, Open Secure Sockets Layer(SSL)), Secure Data Transmission (Hyper Text Transfer Protocol Secure (HTTPS), Transport Layer Security (TLS)), General Data Protection Regulation(GDPR)-compliant Data Storage Systems Model Evaluation and Fine-Tuning: Compare performance between existing Transformer and CNN models and our novel model for ISL recognition using various interpretability techniques and feature importance analysis. Tech Stack: Shapley Additive Explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME)
Technology Keywords	AI, Natural Language Processing(NLP), Sign Language Recognition, Pose Estimation, Real-time Translation, Embedding Space
Data Availability	Public
Metadata (Type of Data)	Video, Text (ISL videos annotated with linguistic mapping)
Model Training and Fine-Tuning	CNN, Transformer-based NLP, Pose Estimation, Adversarial Augmentation
Testbeds or Pilot Deployments	P2SLR:Privacy-Preserving Sign Language Recognition[4]

2 Use Case Description

2.1 Description

Context and Background

Millions of people in India who are deaf or hard of hearing rely on ISL as their main form of communication. Nonetheless, there is still a big communication gap between people who speak ISL and those who don't. For the deaf community, social integration and information access are hampered by the absence of real-time translation tools. The language barrier that separates ISL from commonly spoken languages like Hindi and English also makes communication more difficult and restricts the participation of deaf people in public services, education, and mainstream conversations. While the core focus of this work is on developing an accurate ISL recognition model, we recognize that effective communication solutions must reflect the rich sociocultural diversity of India's deaf and hard-of-hearing communities. Incorporating detailed user archetypes and narratives that capture lived experiences and regional linguistic diversity will be essential for real-world adoption and accessibility. These qualitative insights will guide future phases of this project to ensure the technology is inclusive and responsive to community-specific communication barriers.

Objective and Aims

This research aims to develop an AI-powered system that can effectively detect and translate ISL into English, and vice versa. The core objectives include:

- Real-time ISL Recognition and Translation: Enabling real-time translation between ISL and English, allowing smooth interaction between deaf individuals and others.
- Bidirectional Translation: The system will not only translate from ISL to English but also facilitate translation from English to ISL, making it versatile.
- Multi-language Support: Since translation from other languages (e.g., Hindi) to ISL will be enabled by first converting the text to English, this approach will support indirect translation to ISL from a variety of languages.
- Proof-of-Concept (PoC) Development: The project will create a PoC system to demonstrate the feasibility of ISL recognition and translation in real-time using open-domain ISL datasets and lightweight machine learning models. Use models like the CLIP-SLA method [5], which offers parameter efficient adaptations for continuous ISL recognition.

Use Case Status: In Development (PoC Stage) and this are currently developing a PoC system.

Partner: N/A

2.2 Benefits of the use case

Quality Education: Integrating ISL translation tools into educational settings enhances accessibility for deaf and hard-of-hearing learners. This ensures that individuals who rely on sign language can fully engage with classroom activities, online learning platforms, and academic programs, contributing to a more inclusive and equitable education system.

Industry, Innovation, and Infrastructure: The project showcases a forward-thinking application of artificial intelligence and machine learning to drive social good. By developing technologies such as real-time sign language recognition, pose estimation, and natural language processing, the initiative strengthens digital infrastructure and encourages innovation that supports accessibility for all.

Reduced Inequality: The solution directly empowers deaf and hard-of-hearing communities by enabling greater participation in education, employment, and everyday social interactions. By removing communication barriers, it promotes equal access to opportunities, resources, and services, thereby helping to reduce disparities and foster a more inclusive society.

2.3 Future Work

- Pilot Study Completion: Conduct a comprehensive pilot study evaluating initial ISL detection and translation models.
- Prototype Development: Deploy a functional PoC system demonstrating realtime ISL recognition and translation.
- Research Paper Submission: Prepare and submit a research paper detailing the methodology, findings, and potential improvements based on the pilot study.
- Dataset Release: Publish an initial dataset of annotated ISL signs, including pose estimation and space embeddings representations.
- Technical Report: Document findings from early testing phases and propose refinements for future work.
- AI for Good Summit Presentation: Demonstrate prototype performance and preliminary results at the AI for Good Summit 2025.
- Avatar Generation: Investigate avatar generation for ISL translation, creating dynamic sign language avatars for enhanced visual translation.
- Ethical Safeguards and Data Privacy Measures: Implement and document robust privacy protections for sensitive video recordings, including anonymization techniques (e.g., skeletal abstraction and facial de-identification), secure storage using GDPR-compliant frameworks, and user-informed consent procedures. A Data Protection Impact Assessment (DPIA) will be conducted prior to any public data release to ensure ethical standards are met.

3 Use Case Requirements

- REQ-01: Development of an organic dataset with following:
    - Data collection from native signers, YouTube videos, and expert collaboration.
    - Annotation strategies for regional sign language variations.
    - Quality control measures to ensure real-world usability and linguistic diversity.

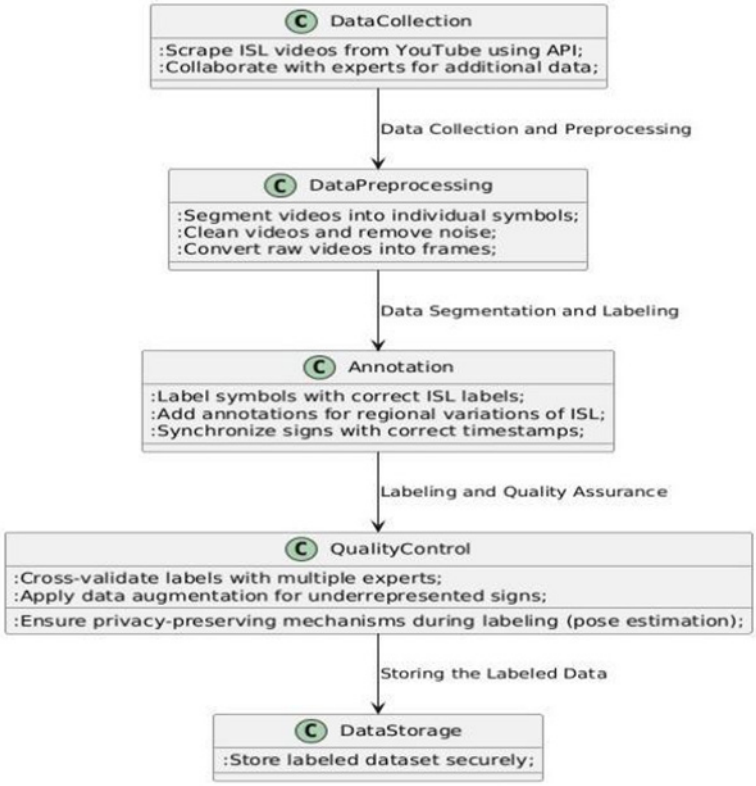
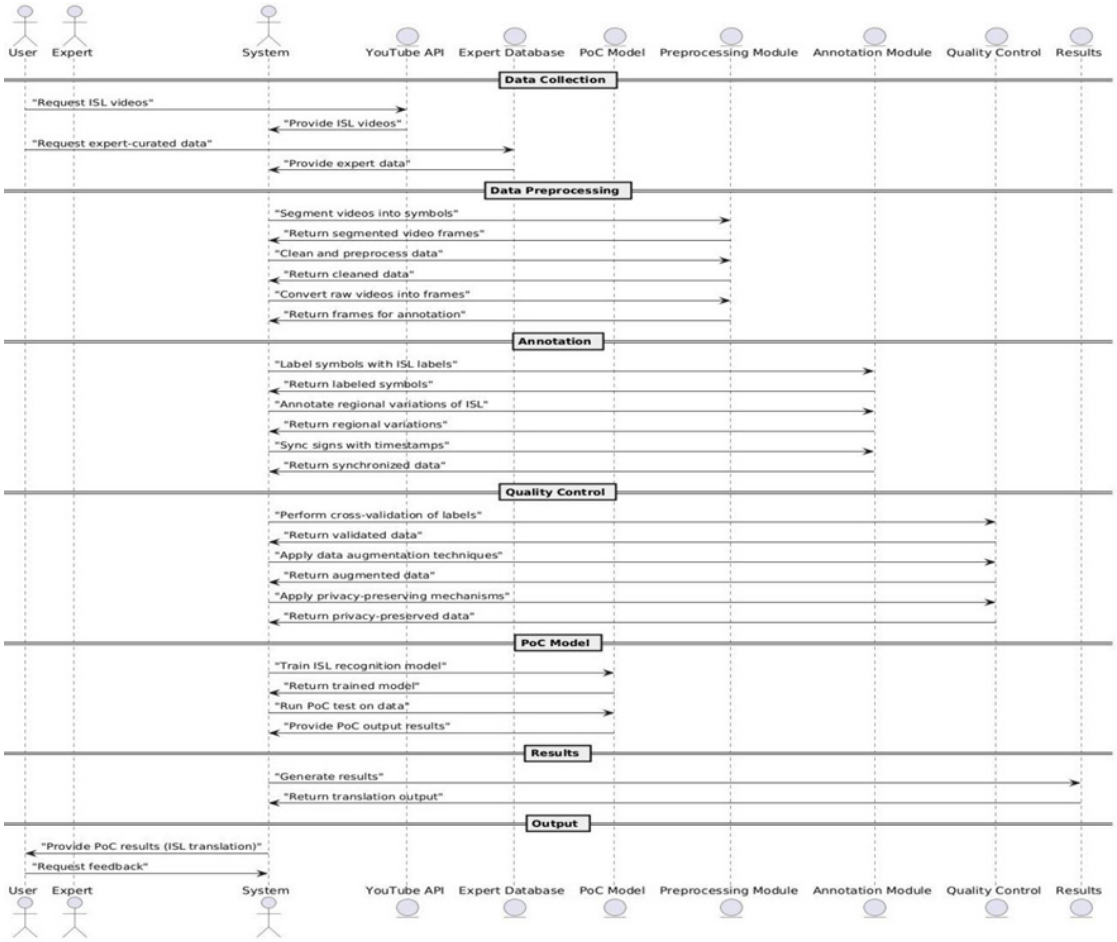


Figure 1: Proposed Data Collection Pipeline

- REQ-02: It must handle temporal differences between consecutive signs, including when to update the predicted letter on the screen.
- REQ-03: The model should join letters to form words and verify if the word belongs to a dictionary.
- REQ-04: The system should understand basic key statement expressions in sign language.
- REQ-05: The system should use pose estimation and embedding representations for better accuracy.
- REQ-06: The AI model should support translation in local languages such as Hindi and English.
- REQ-07: Comparing outputs of existing Transformer and CNN models with our model to understand their performance differences using interpretability techniques, finding feature importance in each of the models.
- REQ-08: Develop a pipeline by addressing regional variations, linguistic nuances, and real-world usage scenarios.
- REQ-09: Preserve the context and emotion of the message to be conveyed throughout the translation and identification process.

4 Sequence Diagram

The following Diagram (High-Level System Architecture for ISL Detection and Translation) Illustrates the proposed System Pipeline:



5 References

[1] Q. Zhu, J. Li, F. Yuan, and Q. Gan, "Continuous sign language recognition based on motor attention mechanism and frame-level self-distillation," Machine Vision and Applications, vol. 36, no. 1, pp. 1–12, 2025. doi: 10.1007/s00138-024-01371-5.

[2] Z. Wang, D. Li, R. Jiang, and M. Okumura, "Continuous sign language recognition with multi-scale spatial-temporal feature enhancement," IEEE Access, 2025. doi: 10.1109/ACCESS.2025.1234567.

[3] K. Hirooka, A. S. M. Miah, T. Murakami, Y. Akiba, Y. S. Hwang, and J. Shin, "Stack transformer based spatial-temporal attention model for dynamic multi-culture sign language recognition," arXiv preprint, 2025. eprint: arXiv:2503.16855. [Online]. Available: https://arxiv.org/abs/2503.16855.

[4] V. K. Tanwar, g. sharma gaurav, B. Raman, and R. Bhargava, "P2slr: A privacy-preserving sign language recognition as-a-cloud service using deep learning for encrypted gestures," Feb. 2022. doi: 10.36227/ techrxiv.19064063.v1. [Online]. Available: http://dx.doi.org/10.36227/techrxiv.19064063.v1.

[5] S. Alyami and H. Luqman, "Clip-sla: Parameter-efficient clip adaptation for continuous sign language recognition," arXiv preprint, 2025. eprint: arXiv:2504.01666. [Online]. Available: https://arxiv. org/abs/2504.01666.

[6]