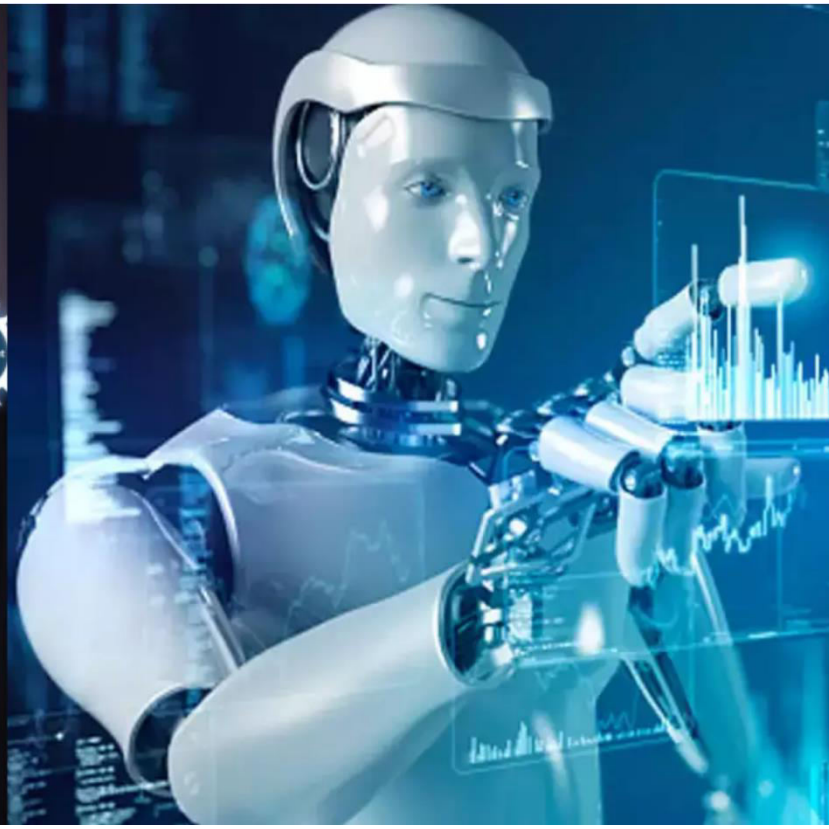
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 3, March 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

AI Tool for Indian Sign Language (ISL) Generator from Audio - Visual Content in English to ISL Content and Vice-Versa

Dr. T. Kameswara Rao[1], Kolati Praisy Joy[2], K Pavan Kumar[3], M Prasanth Raju[4] ,M Dileep Kumar[5]

Professor, Department of Computer Science (Artificial Intelligence and Machine Learning), Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[1]

StudentS, Department of Computer Science (Artificial Intelligence and Machine Learning), Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[2345]

Abstract: Sign language is an important communication medium for the deaf and mute population, particularly in India, where Indian Sign Language (ISL) serves as a primary mode of interaction. However, a significant communication gap persists between ISL users and non-sign language users. This project aims to bridge this gap by creating a Multilingual Sign Language Avatar System that converts English and Hindi (text/audio) into ISL-based sign language avatars, and vice versa, translating ISL gestures into text/audio using Convolutional Neural Networks (CNNs) and Natural Language Processing (NLP). The system features two core functionalities: Text/Audio to ISL Avatar Conversion (supporting English and Hindi inputs) and ISL to Text/Audio Translation. By integrating deep learning, computer vision, NLP, and speech processing—trained on ISL datasets — the system enhances accessibility for India's hearing-impaired community and fosters inclusive communication.

Keywords: Indian Sign Language(ISL), Bidirectional Translation, Convolutional neural networks (CNN), Natural Language Processing (NLP), Gesture Recognition, Accessibility Technology.

I. INTRODUCTION

A sign language (SL) is a natural visual-spatial language that employs three-dimensional space to express linguistic utterances rather than sound to communicate meaning. It combines at the same time hand shapes, orientation, and movement of the hands, arms, upper body, and facial expressions to convey a linguistic message. The language emerged as a result of the demands of the deaf, mute, and hard-of-hearing communities in India. Various communities of deaf and mute people across the globe have various sign languages, just like spoken languages including English, French, and Urdu. For example, American Sign Language (ASL) is employed in America, British Sign Language (BSL) in Britain, and Indian Sign Language (ISL) in India to enable communication. Though interactive systems for ASL and BSL have been created, such developments for ISL are in the process of evolving. Sign language is an important communication medium for the dead and mute population, but there exists a strong communication gap between sign language users and non-sign language users. This project hopes to fill this gap with the creation of a Multilingual Sign Language Avatar System that converts English and Hindi (text/audio) to sign language avatars and sign language to text/audio based on Convolutional Neural Networks (CNNs) and Natural Language Processing (NLP).

The system has two main functionalities: Text/Audio to Sign Language Avatar (English & Hindi) and Sign Language to Text/Audio. By combining deep learning, computer vision, and speech processing, this system improves access for the hearing-impaired community and encourages more accessible communication. Improving real-time performance and supporting more languages will be areas of future work.

In India, close to 63 million individuals are severely impaired for hearing, with their share about 6.3% in the population. More than 30% among them are younger than 20 years, with almost 50% between ages 20 to 60 years. Most such people use signs for communication purposes since they will find it challenging to communicate by verbal means. However, sign languages rarely have an apparent structure or grammar, confining their acceptability to beyond

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

the deaf population. Studies of ASL have developed it into a full-fledged language with inherent grammar, syntax, and linguistic features. In 1978, research began on ISL, and these studies also established that ISL is a full- fledged structured natural language[1].

Communication barriers are a principal challenge for hearing-impaired people, especially in public areas such as railway stations, bus stands, banks, and hospitals. A hearing individual might not comprehend sign language, and a deaf individual might find it difficult to decipher spoken or written speech. To overcome this communication gap, translation systems are needed.

As per the 2011 Indian Census : 6.3% of the overall population (around 63 million) suffer from severe hearing loss. Of these people, 76-89% know no languages ,signed, spoken, or written.

There are a number of reasons for this low literacy rate, including:

- A lack of Sign Language interpreters.
- A lack of available ISL tools.
- Limited studies on ISL.

Because of these issues, public communication in places like railways, banks, and hospitals continues to be hard for the deaf community. To rectify this, a system to translate text into Indian Sign Language and vice versa is essential, greatly enhancing their lives. Contrary to spoken languages, sign languages have not been researched deeply, and much remains to be known and understood.

An audio-to-sign language translator is a web application that is intended to help deaf and hard-of- hearing people. It converts English audio into Indian Sign Language by processing basic English sentences, creating ISL-gloss, and converting it into the Hamburg Notation System (HamNoSys). The HamNoSys representation is used to give signing instructions to a sign synthesis module, which produces an animated ISL representation for the user based on dependency trees to organize ISL syntax.

Further, the Sign Language to Text/Audio system provides an option to input Indian Sign Language alphabets and convert them to text, which in turn can be converted into audio. This feature is implemented using Convolutional Neural Networks (CNNs) and other deep learning methods in order to perform sign recognition accurately as well as smooth translation.

II. EXISTING SYSTEMS

Current attempts at automating sign language translation have mostly been directed towards Western sign languages, such as American Sign Language (ASL) and British Sign Language (BSL), with little concern for Indian Sign Language (ISL). These approaches typically experience limited domain applicability, limited lexicons, or poor non-manual handling. Following is an overview of three typical systems, pointing out their shortcomings in tackling ISL's linguistic and technological issues.

TESSA (Cox et al., 2002) was the first to introduce domain-specific translation for BSL in the context of post-office transactions. The system employs direct translation from a fixed vocabulary of 370 pre-defined English template phrases (e.g., "bill payments," "passport renewals") to their corresponding BSL animations. Although effective in the case of scripted interactions, TESSA's use of template-based phrases makes it unable to handle novel sentences. It maintains English's Subject-Verb-Object (SVO) syntax, ignoring BSL's spatial grammar [1], and has no support for real- time audio inputs. For example, the question "How much does this parcel cost?" needs to be explicitly chosen from a menu instead of being spoken, which restricts its usability in dynamic situations such as hospitals or public transport.

INGIT (Dasgupta et al., 2008), designed at IIT Kanpur, aims at ISL translation for railway bookings. It analyzes Hindi text with Fluid Construction Grammar (FCG) and produces ISL glosses by hand-coded rules. INGIT's vocabulary, though, is limited to 90 domain-specific words (e.g., "train," "ticket") and thus fails for general communication. The system does not support real-time speech processing and takes transcribed Hindi text as input, and uses simple HamNoSys animations without 3D avatars[2,3]. For instance, the question "When is the next train to Delhi?" is

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

rendered as "Train Delhi next when?" but without non-manual signs such as eyebrow raises that are essential for question sentences in ISL[4]. This creates stiff animations that cannot capture linguistic detail.

SignSynth, an open-source ASL synthesizer, enables users to create signs by manually choosing handshape, location, and movement parameters through ASCII- Stokoe notation. Though innovative, the system requires technical knowledge to enter phonological specifics (e.g., finger extensions, palm orientation) via dropdown menus.SignSynth does not support facial expressions and body language, critical for ASL and ISL grammar, and does not provide integration with speech-to-text technologies.Generation of a sign such as "happy" involves individual selections for handshape (flat palm), location (chest), and movement (circular), a procedure not feasible for non-specialists. Additionally, its vocabulary is restricted to elementary nouns and verbs, without the use of adaptation to ISL's Subject-Object-Verb (SOV) syntax or local dialects.

Comparative Analysis of Existing System Limitations:

Criteria	TESSA	INGIT	SignSynth
Vocabulary Size	370 words	90 words	Basic noun/verbs
Input Method	Manual Text	Transcribed Hindi	Manual Parameters
ISL Grammar Support	No	Partial	No
Animation Quality	2D Animations	Static HamNoSys	2D Web3D

III. PROPOSED WORK

The designed system is a two-way translation platform aimed at filling the communication gap between hearing and deaf populations through multilingual input (English/Hindi) and conversion of sign language to text/audio. The system takes audio input in Hindi or English, which is converted into text through automatic speech recognition (ASR). In the case of Hindi input, the converted text is machine-translated into English via machine translation (MT) to follow the pre-defined ISL grammar rules. The syntactically rearranged English translation is mapped into the Subject-Object-Verb (SOV) structure of ISL through parse tree transformations. Bi-linguistic rules are enforced by a rule-based translation engine to translate the transformed parse tree into ISL glosses, which are translated into Hamburg Notation System (HamNoSys) and output as animations through 3D avatars.

To facilitate sign language-to-text/audio synthesis, the system utilizes live webcam-based gesture recognition to recognize ISL alphabets and gestures. By applying computer vision algorithms (e.g., OpenCV, MediaPipe), hand positions, motions, and facial expressions are tracked in real-time. For example, ISL alphabet handshapes (e.g., "A" as a closed fist, "B" as an open flat palm) are identified frame by frame and translated to HamNoSys notation. These signs are synthesized into words by temporal analysis, where successive alphabets (e.g., "H-E-L-L-O") are joined together using Long Short-Term Memory (LSTM) networks to preserve temporal coherence. The identified signs are translated into text, the text is then converted to speech on users input. To enhance usability, a predictive text engine leverages n-gram language models to suggest probable words (e.g., "HELLO," "HELP") as the user signs, reducing input effort.

The proposed architecture consists of two key components:

- Text/Audio to Sign Language – Translates spoken or written language into Indian Sign Language representations.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

- Sign Language to Text/Audio – Converts Indian Sign Language gestures into corresponding text or audio output.

3.1 Text/Audio to Sign Language

The system is architected to convert spoken or written words into Indian Sign Language (ISL), as illustrated in Fig. 1, using the following major modules:

➢ Audio to Text Conversion:

This module records audio input from an external or internal microphone on a device like a Personal Digital Assistant (PDA). Using Google Cloud Speech, the system decodes speech into text, with support for a variety of languages, and the output being English standardized. The implementation is created using Python so that correct speech recognition is achieved and punctuation is automatically inserted.

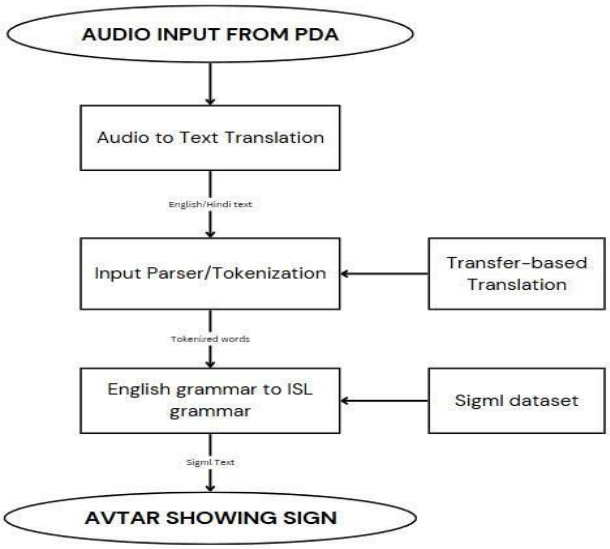


Fig. 1 : System Architecture

➢ Text Processing & Tokenization :

After the text is created, it is divided into separate sentences and then into words. This is done through a process called tokenization, which is done with the help of Natural Language Processing (NLP) methods. The tokenizer assists in pulling out meaningful words while removing unnecessary characters such as punctuation marks.

➢ ISL Translation & Grammar Adaptation :

As ISL has a dissimilar grammatical pattern compared to that of English, the process of translation is converting sentences to fit the syntax of ISL. This is done using a Transfer-Based Translation Model, which translates the original text and runs it through linguistic rules to transform its structure into ISL-supporting grammar[1,5].

➢ Reordering of Sentences :

English adopts a Subject-Verb-Object (SVO) pattern, while ISL generally adopts a Subject-Object-Verb (SOV) pattern. The system reorders the sentence structure.

➢ Dealing with Interrogative Sentences :

Question words like who, what, when, where, and why are moved to the end of the sentence in ISL.
Example:
English: "When is your birthday?"
ISL: "YOUR BIRTHDAY TIME + QUESTION"

➢ Stop Word Removal :

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

Frequently occurring words like is, am, are, was, were, a, an, and the are removed since they do not add to ISL meaning.

- Lemmatization & Stemming:

Words are reduced to their base form in order to correspond to ISL representation. For instance: "Running" → "Run" "Organizes" → "Organize"

- Synonym Matching for Constrained Vocabulary:

ISL having a limited vocabulary, there could be English words that lack exact ISL equivalents. Under these circumstances, the system tries to find synonyms based on WordNet and substitutes words with the closest ISL-compatible equivalent [5].

- Sign Language Animation Generation:

The last step is to translate the processed text into ISL signs. Every word is translated to a corresponding SIGML (Signing Gesture Markup Language) file, which holds pre-defined hand gesture animations based on HamNoSys notation [2,3]. If a word lacks a corresponding file, the system spells out the word using separate alphabet signs to generate the sign. These gestures are then displayed as an animated avatar, building on frameworks like ViSiCAST [6,7] executing ISL signs, thereby passing on the translated message.

3.2 Sign Language to Text/Audio

This module supports two-way communication by translating ISL gestures into text and synthesized speech. The pipeline, as illustrated in Fig. 2, is composed of data acquisition, preprocessing, gesture classification, and translation, as explained below:

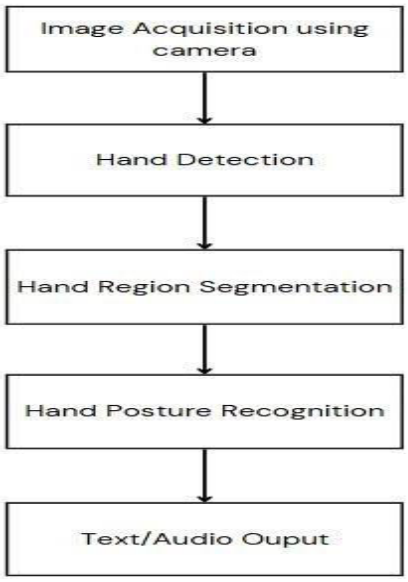


Fig. 2 : System Architecture

- Data Acquisition:

The system uses vision-based approaches to detect ISL gestures through a regular webcam, without requiring expensive wearable units such as sensor gloves. Drawbacks are variation in hand appearance based on illumination, skin color, and cluttered background. To overcome these, the architecture focuses on ensuring robustness to environmental variations to achieve consistent gesture detection under a wide range of real- world conditions.

- Data Preprocessing and Feature Extraction:

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

Hand areas are found through the MediaPipe library, which detects 21 landmarks on the hand (such as fingertip locations, knuckles). These landmarks are then marked on a white canvas to separate gestures from intricate backgrounds. Major steps are:
Region of Interest (ROI) Extraction: Extract hand areas from webcam frames.
Noise Reduction: Use Gaussian blur to reduce sensor noise.
Thresholding: Process images into binary form for increased contrast.

- Gesture Classification

A Convolutional Neural Network (CNN) classifies ISL alphabets based on landmark-augmented images. The CNN structure includes:
Convolutional Layers: Identify spatial features (e.g., edges, shapes) through 5×5 filters.
Max-Pooling Layers: Minimize dimensionality while maintaining key features.
Fully Connected Layers: Project extracted features to 8 pre-defined gesture classes (e.g., [A, E, M, N, S, T], [G, H]) to accommodate inter-class similarities.
For uncertain cases (e.g., distinguishing "A" from "E"), rules from post-processing based on landmark distances narrow down predictions.

- Text and Speech Synthesis

Detected gestures are combined into words via temporal analysis. For example, consecutive detection of landmarks for "H," "E," "L," "L," "O" spells out "HELLO." The text is converted to speech through the pyttsx3 library, mimicking real-time conversation. This capability enables hearing users to hear audible.

IV. RESULTS

This section describes the evaluation approach and overall results of the suggested bidirectional translation system. The assessment is centered around functional performance over various use cases, with particular focus on technical issues and qualitative findings. Representative screenshots of the system's user interface are provided in Figures 3, illustrating key functionalities such as real-time gesture recognition and bidirectional translation.

4.1 Text/Audio to ISL Translation

The system was tested utilizing sentences from health, transport, and education scenarios, as shown in Fig.3 (providing screenshot of web interface).

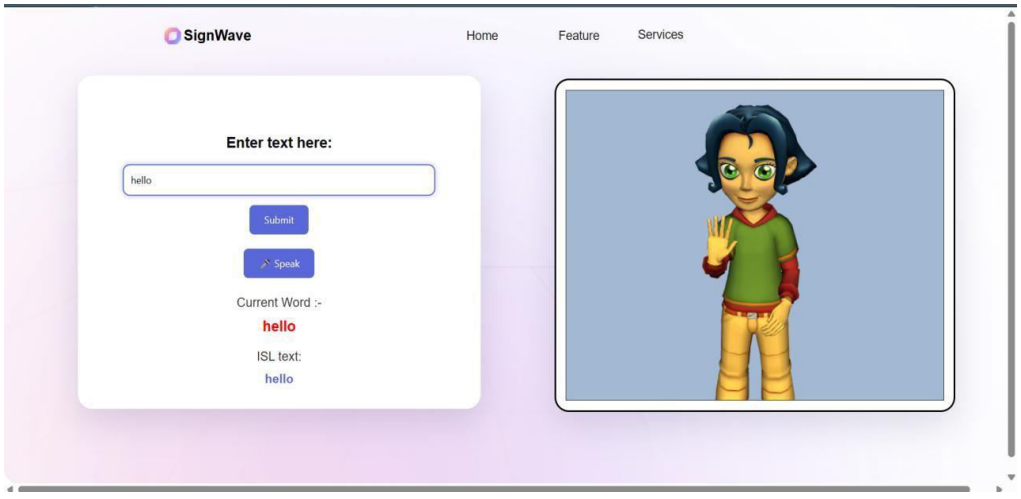


Fig.3 : Screenshot of web interface

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

Major observations are:

- Syntactic Adaptation:The system effectively transformed English SVO syntax into ISL's SOV arrangement in easy and moderately complex sentences. For instance, "Submit the application by Tuesday" was translated as "Application Tuesday by submit."

- Lexical Gaps:Uncommon or specialized words(e.g., technical vocabulary) sometimes necessitated synonym resolution, resulting in approximations.

4.2 ISL to Text/Audio Translation

The gesture recognition module was evaluated in changing lighting and background conditions:

Hand Landmark Robustness:MediaPipe's landmark detection was robust across environments but sometimes lost hand tracking with extreme lighting changes (e.g., shadows, glare).
Gesture Classification: The CNN model performed steadily for standard ISL alphabets but struggled with similar-looking signs (e.g., "G" vs. "H").

V. CONCLUSION AND FUTURE SCOPE

This AI-based bidirectional translation system is a landmark development in the process of connecting hearing and hearing-impaired populations in India through communication. The system, with its integration of audio-visual input processing along with state-of-the-art NLP and computer vision, facilitates real-time English/Hindi text/audio to ISL and ISL to English/Hindi text/audio translation. The system effectively tackles elementary ISL grammar principles, including Subject-Object-Verb (SOV) word order and interrogative repositioning, along with facial expression incorporation to complement linguistic expression. Addition of Hindi language support expands accessibility among India's language-diverse population, and bidirectional capability—transforming ISL gestures into text/audio utilizing CNNs as well as webcam-based gesture detection—provides a platform for inclusive two-way communication. Yet, difficulties are still encountered with processing non-manual elements (e.g., subtle facial displays for negation), regional differences in dialects, and high-level syntactic structures, precluding the full imitation of ISL natural discourse by the system. Follow-up research will address the extension of the robustness and the flexibility of the system.

Pivotal priority is the synthesis of non-manual gestures for coordinating facial movements, head shifts, and posture with manual signing to provide complete grammatical translation. Enlarging the ISL lexicon to domain-specific phrases (e.g., medical terminology, educational jargon) and typical expressions (e.g., salutations) will enhance contextual usability in public places such as schools and hospitals. Moreover, optimization of the vision-based gesture recognition module to manage low-light and background clutter situations will enhance real-world usability. Creating a mobile app will provide democratization and make on-the-move translation possible for users within resource-poor environments. In addition, integrating regional ISL dialects and enhancing real-time performance with lightweight neural structures will facilitate wider adoption. Through the filling of these gaps, the system hopes to develop into a holistic device for fair communication, enabling India's deaf community to engage actively in socio-economic life.

REFERENCES

[1] Johnston, T., 1991. Transcription and glossing of sign language texts: Examples from AUSLAN (Australian Sign Language). International Journal of Sign Linguistics 2(1): 3–28.
[2] Prillwitz, S. et al., 1987. HamNoSys. Hamburg Notation System for Sign Languages. An introduction. Hamburg: Zentrum für Deutsche Gebärdensprache.
[3] Hanke, T., 2002b. HamNoSys in a sign language generation context. In R. Schulmeister and H. Reinitzer (eds.), Progress in sign language research: in honor of Siegmund Prillwitz. Seedorf: Signum. 249–266.
[4] Domain Bounded English to Indian Sign Language Translation Model, CSE, Sharda University, Noida.
[5] Translation of Hindi Text to ISL and extension of ISL Dictionary with WordNet, Thapar University, Patiala.
[6] Elliott, R., J. Glauert, R. Kennaway and I. Marshall, 2000. The development of language processing support for the ViSiCAST project. In The fourth international ACM conference on Assistive technologies (ASSETS). New York: ACM. 101–108.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403154|

[7] Schulmeister, R., 2001. The ViSiCAST Project: Translation into sign language generation of sign language by virtual humans (avatars) in television, WWW and face-to-face transactions. In C. Stephanidis (ed.), Universal access in HCI : towards an information society for all. Hillsdale, NJ: Erlbaum. 431-435.

[8] Sharma, S., & Kumar, A. (2021). Sign Language Recognition Using Convolutional Neural Networks and Computer Vision. IEEE Transactions on Human- Machine Systems.

[9] Lugaresi, C., et al. (2019). Real-Time Hand Tracking Using MediaPipe for Sign Language Applications. CVPR Workshops.

[10] Patel, R., & Singh, D. (2020). Natural Language Processing Techniques for Sign Language Translation. ACM Transactions on Asian Language Processing.

[11] Verma, S., & Joshi, N. (2022). Deep Learning Architectures for Indian Sign Language Recognition. Springer Journal of Ambient Intelligence and Humanized Computing.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details