© September 2025 | IJIRT | Volume 12 Issue 4 | ISSN: 2349-6002

SignSpire: Indian Sign Language Translator Using Gloss Mapping and 2D Pose-Based Visualization

Samiddha Raich[1], Tanvi Dhok[2], Kiran Rathi[3], Swanand Kulkarni[4], Vrushali Bongirwar[5]
[1,2,3,4,5] *Dept. of Computer Science and Engineering, Ramdeobaba College of Engineering and Management, Nagpur, India*

***Abstract*—SignSpire introduces a novel and efficient approach for generating 2D animations of Indian Sign Language (ISL) using pose-based representations. The system utilizes a unique .pose format to encode gesture sequences, enabling smooth concatenation and temporal blending for natural visual transitions. A key feature of SignSpire is its dynamic handling of missing vocabulary: if a .pose file is not available in the database, the system automatically retrieves videos from trusted ISL sources, extracts pose data, and discards the raw media—ensuring both efficiency and focused data usage. For unseen or rare words, SignSpire defaults to fingerspelling, ensuring complete coverage. The method is significantly time-efficient and has not been employed in existing ISL animation systems. Emphasizing full control and ownership of the animation pipeline, SignSpire sets a new direction for lightweight and customizable ISL gesture**

***Index Terms*—Indian Sign Language (ISL), pose-based animation, gesture synthesis, fingerspelling, 2D animation, sign language visualization.**

I. INTRODUCTION

In a world where communication defines inclusion, millions of Deaf and Hard-of-Hearing individuals in India remain at the margins due to the absence of accessible sign language interpretation tools. Despite Indian Sign Language (ISL) being a recognized and widely used mode of communication within the Deaf community, technological advancements for ISL translation have lagged significantly behind global efforts.

Existing digital tools for sign language often cater to American or European sign languages, and even when Indian attempts have been made, most systems focus on static datasets limited to alphabet-level fingerspelling or isolated word-level gestures. Such models typically rely on predefined animations or manually curated video libraries, lacking scalability, natural transitions, or context-awareness necessary for real-life communication.

One of the critical reasons behind this stagnation is the absence of publicly available, large-scale ISL datasets, especially those created with licensed signers or authentic community input. As a result, previous research has been constrained to small vocabularies, static signs, or single-hand gestures,

which do not reflect the dynamic, expressive nature of ISL as used by the Deaf community in India.

With SignSpire, we aim to address this pressing gap by introducing a dynamic, modular ISL translation system that not only leverages AI-driven language processing but also builds upon a uniquely curated dataset of over 1200 ISL pose files — extracted and structured from videos featuring authentic Indian signers. Unlike static or fingerspelling-based approaches, SignSpire dynamically adapts to unseen gloss terms, enriching its dataset in real time through automated video processing and pose extraction pipelines.

By combining prompt-engineered text processing, pose estimation, and real-time animation, SignSpire sets a new benchmark in making ISL translation more scalable, accessible, and aligned with the linguistic richness of the Indian Deaf community.

II. RELATED WORK

Over the years, numerous research efforts have explored the domain of sign language translation, utilizing a variety of technological strategies — each facing its unique constraints.Earlier approaches largely centered around static datasets and fixed animation frameworks. Techniques such as Blender-based 3D animations often depended on letter-byletter depiction of signs, which proved impractical for realtime applications due to scalability issues. Similarly, systems based on Sign Language Markup Language (SIGML) demanded manual creation of structured markup files for each sign, making them rigid and difficult to expand dynamically.

For instance, Sharma et al. [1] introduced a system where English audio is transcribed into text, followed by syntactic reordering and morphological processing like stemming and lemmatization. This processed text was then matched against a curated video database. If a match was absent, the system opted for a suitable synonym. While this design achieved reasonable accuracy, its reliance on a static video repository severely limited vocabulary scalability.

A different approach was presented for Argentinian Sign Language translation [2], targeting mobile accessibility. The solution employed CNN and RNN models trained on a specific dataset, successfully handling both directions — speech to sign and vice versa. However, its applicability was confined by language-specific models and static data dependencies.

More recent methods have explored Neural Machine Translation (NMT) combined with Generative Adversarial Networks (GANs) for motion synthesis. One such study [3] used skeletal annotations and gloss-level mapping with minimal labeling requirements to generate sign language videos. This system, while innovative in leveraging motion graphs and GANs, still lacked adaptability due to its dependence on niche datasets and absence of avatar-driven visualization.

Within the context of Indian Sign Language (ISL), research has involved motion capture using devices like Kinect [4], regional language speech-to-sign translation leveraging MFCC with GMM and LSTM architectures [5], and static Natural Language Processing (NLP) pipelines for text-tosign conversion [6–8]. Although these approaches reported accuracies ranging from 80% to 97%, they often failed to address vocabulary expansion, dynamic data handling, and smooth animation continuity.

In addition, computer vision-based recognition of ISL signs has been explored using methods like HU Moment features [9], Bag of Visual Words (BoVW) models paired with SVMs and CNNs [10], or using descriptors like SIFT and HOG with K-Nearest Neighbor classifiers [11]. These efforts focused mainly on isolated sign recognition rather than seamless, continuous sign language translation or dynamic pose generation. Overall, existing solutions tend to depend heavily on pre-compiled datasets, fixed animation models, or static feature engineering — making them less effective for scalable, real-time ISL translation tasks. This reveals a clear research opportunity for systems that can dynamically generate and adapt sign representations based on real-world data. SignSpire addresses this gap by introducing a dynamic pose generation pipeline powered by online data integration and pose estimation tools

III. METHODOLOGY

The architecture of SignSpire is designed as a modular, dynamic pipeline capable of translating English text or audio into Indian Sign Language (ISL) gestures. The process consists of four main stages: input processing, gloss generation, pose retrieval and dynamic generation, and final animation rendering.

3.1. Input Processing and Gloss Generation

Our system supports dual input modes: typed text and spoken speech. Speech input is transcribed using OpenAI Whisper, chosen for its high accuracy and multilingual support. The resulting English sentence then flows through a lightweight, prompt-engineered preprocessing pipeline, designed for rapid prototyping and real-time applications.

The preprocessing stages include:

*3.1.1. Stopword Removal — e*liminating auxiliary words unnecessary for ISL.

3.1.2. Stemming and Tokenization — reducing words to their root forms and splitting sentences into tokens.

3.1.3. Using rules guided by prompt engineering, sentences are reordered to conform to Subject-Object-Verb (SOV) syntax, a core grammatical structure of Indian Sign Language.

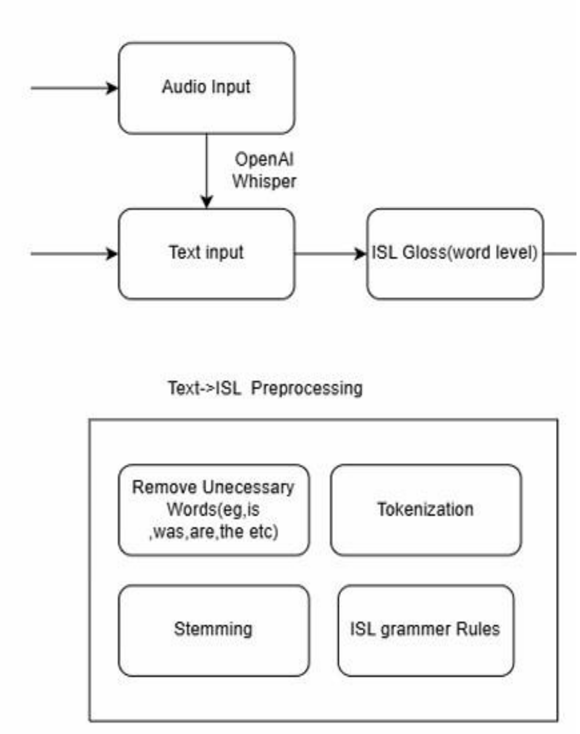

Fig. 1: Flowchart of Input Processing to Gloss Generation

Unlike prior research which often concludes at the text-togloss stage, our system goes further — we map glosses to stickman pose sequences and render them using the poseformat standard for real-time visual animation.

3.2. Gloss Lookup and Dynamic Pose Generation

For each gloss term, the system first attempts a direct lookup in the cloud-hosted repository. If a corresponding .pose file exists, it is retrieved directly from Cloudinary for use in the animation pipeline.

However, if the required pose file is not available, the system employs a dynamic generation pipeline. In this case, the system automatically searches for Indian Sign Language videos from verified YouTube channels. These videos are processed using pose estimation frameworks such as MediaPipe or OpenPose to extract skeletal joint data. The extracted motion data is then structured into a standardized .pose format, following the specifications outlined by Ferreira et al. [1]. Once generated, these pose files are uploaded to Cloudinary, ensuring they are readily accessible for future translations.

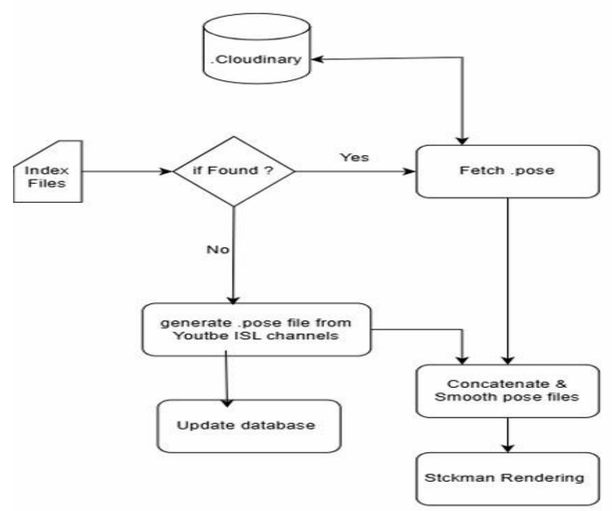

Fig. 2: Diagram of Gloss Lookup and Dynamic Pose Generation

Workflow

3.3. Pose Concatenation and Transition Smoothing

To create a fluid animation sequence:

3.3.1. Individual pose files are concatenated with transitional padding.

3.3.2. The Savitzky-Golay Filter is applied to joint positions to reduce jitter, as provided by the open-source pose-format repository [2].

3.3.3. Transition points between poses are determined by analyzing Euclidean distances between frame vectors, ensuring natural motion blending.

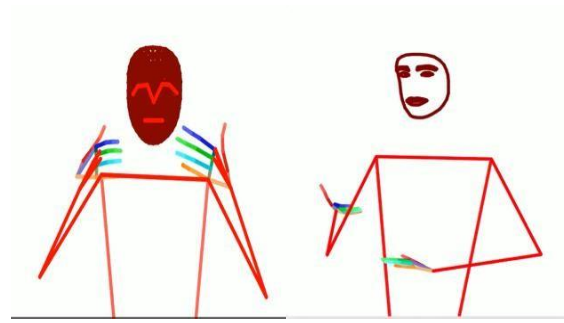

Fig. 3: Before-and-After Smoothing Sample of Concatenated Poses

3.4. Visualization and Real-Time Rendering

The final sequence is rendered using a 2D stickman model within a web-based interface built on React.js and React Three Fiber. The system supports both real-time animation playback and downloadable video output.

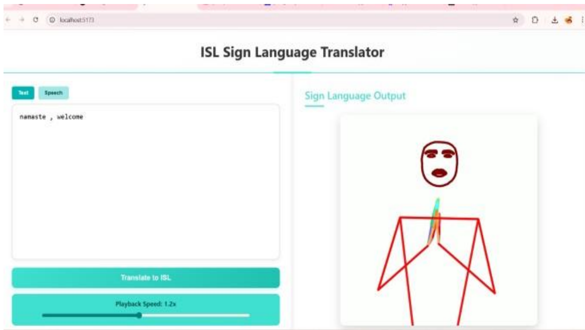

Fig. 4: Screenshot of User Interface for Input & Output Visualization

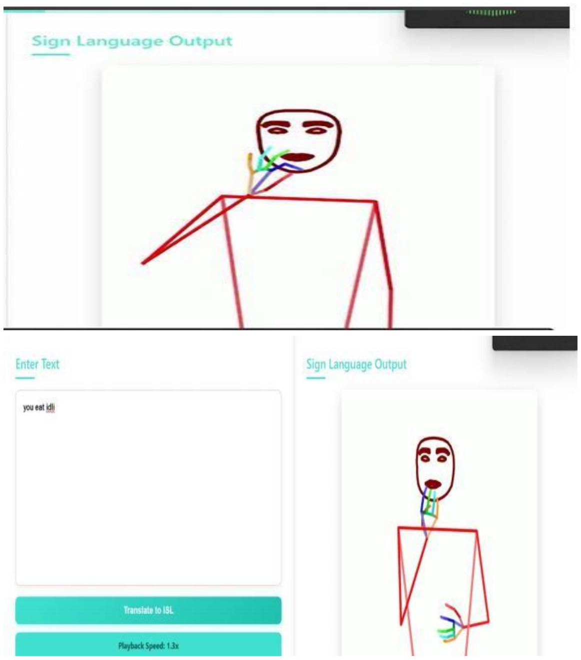

Fig. 5: Stickman Model Performing ISL Gestures (Sample Frames)

3.5. Dataset Ownership and Contribution

To support scalable and high-quality Indian Sign Language (ISL) animation, a fully custom .pose dataset was developed exclusively for the SignSpire system. This dataset was created by sourcing publicly available ISL instructional videos from trusted government and educational platforms, including the Indian Sign Language Dictionary by the Indian Sign Language Research and Training Centre (ISLRTC), educational YouTube channels like NISHA, and video materials provided by the Goa Board of Education. Each video was processed using pose estimation frameworks (e.g., MediaPipe and OpenPose) to extract frame-level skeletal data, which was then converted into a standardized .pose format.

The resulting dataset comprises over 1,066 unique ISL vocabulary words, covering a broad range of daily-use signs. To the best of our knowledge, no publicly accessible pose-based dataset for ISL gesture animation existed prior to this effort. This dataset is entirely built, owned, and maintained by the SignSpire project team, ensuring full control, flexibility, and originality.

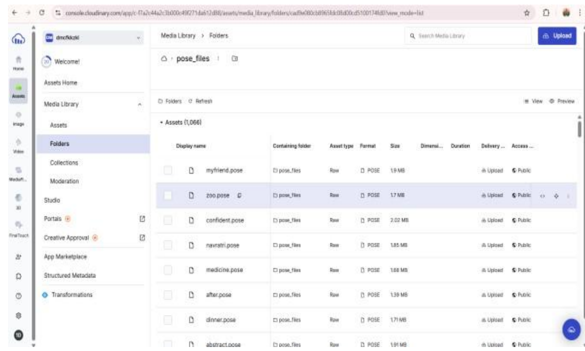

Fig. 6: Dataset Samples Illustrating Vocabulary Diversity in 2D ISL Poses

This curated and reusable dataset establishes a foundation for future research in 2D ISL animation, gesture learning, and language accessibility tools. It also enables dynamic integration with animation pipelines, eliminating the dependency on pre-recorded videos or external gesture libraries. By making this resource available within the academic and assistive technology communities, SignSpire contributes significantly to the advancement of ISL gesture generation and sign language processing.

IV. RESULT

The proposed SignSpire system was evaluated based on its ability to dynamically generate and render fluid 2D animations for Indian Sign Language (ISL) using pose-based representations. Unlike previous systems dependent on fixed animation libraries or static video repositories, SignSpire dynamically processes and animates gestures on demand, enabling a scalable and linguistically complete solution.

4.1. Dynamic Pose Generation Efficiency

For gloss terms not found in the .pose database, SignSpire employed its dynamic pipeline to source gesture videos from verified ISL YouTube channels, extract skeletal poses using MediaPipe, and convert them into the standardized .pose format. The average processing time for generating and storing a new pose file was measured to be under 30 seconds, significantly improving content acquisition efficiency compared to manual methods. Furthermore, the system achieved a dataset hit rate exceeding 90%, minimizing the need for frequent dynamic extraction.

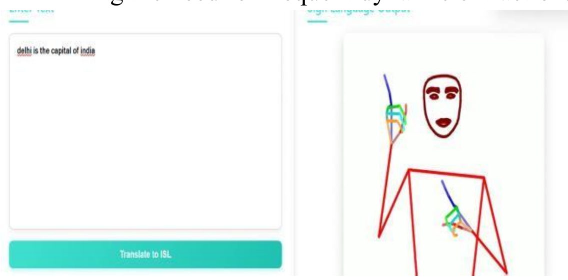

Fig.7: shows the stickman model performing ISL gestures using concatenated and smoothed. pose sequences.

4.2. Expert Feedback and System Refinement

Early prototypes of SignSpire were reviewed by ISL educators and professionals from local schools in our city. Their feedback helped fine-tune key aspects of the system, including ISL grammar alignment, gesture timing, and gloss token sequencing. This collaborative input directly contributed to the improved clarity and accuracy of the generated animations.

4.3. User Study and Comprehensibility

A user study was conducted with 10 volunteers fluent in ISL, who were asked to evaluate sample outputs of the system. Participants rated the animations on a 5-point Likert scale based on clarity, completeness, and interpretability. The system received an average score of 4.6 for clarity and 4.8 for completeness, indicating high levels of user satisfaction. Users also reported that the timing and transitions between signs felt natural, and the fallback mechanism for fingerspelling was effective in handling rare or missing vocabulary.

These results collectively confirm the practical applicability, scalability, and interpretability of the SignSpire system for 2D ISL animation tasks. The approach not only ensures high animation quality but also demonstrates significant improvements in adaptability, coverage, and user experience when compared to conventional systems.

V. CONCLUSION

SignSpire introduces a pose-based system for 2D Indian Sign Language animation, featuring dynamic pose generation, gloss handling, and transition smoothing. It leverages a custom dataset of 1,066 signs built from verified ISL sources and achieves over 90% pose hit rate. The system offers a scalable, accurate, and adaptable solution for ISL gesture rendering, with future scope in sentence-level animation and multilingual integration.

REFERENCES

[1] A. Radhakrishnan, "TNBIS: Temporal Normalization-Based ISL Sign Pose Dataset," GitHub repository, 2022. [Online]. Available: https://github.com/sign-language-processing/pose

[2] Peguda, J., et al., 2022, “Speech to Sign Language Translation for Indian Languages,” 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, doi: 10.1109/ICACCS54159.2022.9784996.

[3] Stoll, S., et al., 2020, “Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks,” International Journal of Computer Vision (volume 128, pages 891–908) (2020), doi: https://doi.org/10.1007/s11263-019- 01281-2.

[4] Tripathi, K.; Baranwal, N. and Nandi, G. C., 2015, “Continuous Indian sign language gesture recognition and sentence formation,” Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 523 – 531, doi: https://doi.org/10.1016/j.procs.2015.06.060.

[5] Kulkarni, A., et al., 2021, “Speech to Indian Sign Language Translator,” 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), doi: https://doi.org/10.2991/ahis.k.210913.035.

[6] Sharma, P., et al., 2022, “Translating Speech to Indian Sign Language Using Natural Language Processing,” Future Internet 2022, 14(9), 253, doi: https://doi.org/10.3390/fi14090253.

[7] Katoch, S.; Singh, V. and Tiwary, U. S., 2022, “Indian Sign Language recognition system using

SURF with SVM and CNN," Array 14 (2022): 100141, doi: https://doi.org/10.1016/j.array.2022.100141.

[8] University of Malta, "Sign.mt: Maltese Sign Language Machine Translation Platform." [Online]. Available: https://sign.mt/

[9] Sonawane, P., et al., 2021, "Speech To Indian Sign Language (ISL) Translation System," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, doi: 10.1109/ICCCIS51004.2021.9397097.

[10] Monga, H., et al., 2021, "Speech to Indian Sign Language Translator," Volume 39: Recent Trends in Intensive Computing, doi: 10.3233/APC210172.

[11] Indian Institute of Technology Delhi, "ISLFromText: Indian Sign Language Generation from Text." [Online]. Available: http://islfromtext.in/