

Hand Sign Translator Bridging the Communication Gap for ISL Users

Swati Bagade^{1*}, Manisha Wasnik², Ujvala Patil³, Jayashri Waman⁴, Tanmay Potnurwar⁵,
Aryan Raut⁶
^{1,2,3,4,5,6}Ajeenkya D. Y. Patil School of Engineering, Lohegaon, Pune, India

Abstract. Millions of individuals with hearing disabilities, or who are deaf, in India experience significant barriers when communicating in-person with individuals who do not know Indian Sign Language (ISL). This barrier often restricts their access to or utilization of vital services, education, and social interactions. The Hand Sign Translator is a real-time solution that reliably recognizes ISL hand gestures to improve communication within these contexts. Built with Python primarily, the Translator consists of OpenCV and deep learning techniques, assuring reliable hand detection and feature extraction with MediaPipe. The hand gestures are converted to text using a hybrid model consisting of Convolutional Neural Networks (CNN), and a Bidirectional Long Short-Term Memory (BiLSTM). The model trained on ISL data frequency from various users, providing the model with the ability to generalize to various signing styles while reliably giving consistent output. The Hand Sign Translator decreases the communication barrier between the deaf community and the general public by recognizing hand gestures, and translating it into written text. The application also enhances social and economic inclusion of individuals with hearing disabilities by developing a new and inclusive assistive technology.

Keywords—Indian Sign Language(ISL), HandSign Translator, Deep Learning, Sign Language Recognition (SLR)

1 Introduction

Indian Sign Language (ISL) is an important mode of interaction for many millions of deaf or hard of hearing people in India. ISL serves as the primary language for individuals to express thoughts, emotions, and ideas through a structured combination of hand signs, facial expressions, and body movements. ISL allows for easy and effective communication amongst the deaf community, and as such, ISL has an important role to play in education, socialization, and long-term employment for individuals who are deaf. Even though ISL is an important language, it is not adequately understood, nor is it recognized, by the majority of the hearing population in India.

* Corresponding author: swatibagade@dypic.in

Consequently, hearing individuals are often uneducated about ISL — and assimilating to hearing culture with deaf folks is often difficult, leading to communication failure, social exclusion, and limited access to essential resources and services. As a result, deaf ISL users encounter systemic barriers in education settings and workplaces, as well as within everyday life when an interpreter, or a person who is trained in ISL, is not present. In an effort to resolve this longstanding issue, researchers and technologists have begun to examine how artificial intelligence (AI) and computer vision can be combined to develop Sign Language Recognition (SLR) systems to detect and interpret sign language gestures and translate them into text or speech in real-time to support communication between D/deaf and hearing people. These SLR systems combine machine learning algorithms and deep learning models to accurately identify hand gestures, finger positions, and facial expressions of sign language, incorporating the complex forms of sign language. Over the past two or three years, some good progress has been made towards recognizing American Sign Language (ASL), with a number of reasonably working models and datasets currently available for research and practice. However, equivalent progress has not been made for Indian Sign Language, mainly due to the lack of standardized datasets, variations in signs from region to region, and limited research on the challenges specific to Indian Sign Language (ISL).

The HandSign Translator was designed to address the social and technological gap. This system combines modern technology, including Python, OpenCV, and MediaPipe, to provide hand detection, segmentation, and feature extraction in real time. Utilizing a hybrid deep learning architecture that combines Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (BiLSTM) networks for hand gestures allows the system to learn both spatial and temporal information from hand gestures. CNNs will analyze the image-based spatial features of hand shapes, and BiLSTM layers will capture the sequential information of the gestures to provide accurate and consistent interpretation of sign sequences that change dynamically. The system is trained on a large and diverse ISL dataset that includes various users, lighting conditions, and backgrounds, which enables the system to generalize to signing styles and different environments.

The HandSign Translator provides a text representation of the recognized ISL gestures and can easily be expanded to include speech to allow for direct communication with people who do not sign. Its real-time function makes it suitable for use in classrooms, hospitals, offices, and public service centers that call for quick and reliable communication. In addition to the benefits for the individual, the HandSign Translator is an important step towards creating a more inclusive and accessible society. The Tool reduces communication barriers and enables deaf individuals to comfortably communicate with others, engage in social and professional interactions, and access equal opportunities in all domains. Therefore, the HandSign Translator demonstrates how technology can be leveraged to promote inclusion, access, and equality and can help create a bridge between the hearing and non-hearing communities in India.

2 Literature Survey

Byeongkeun Kang et al.[1] conducted a study introducing a CNN model that is used for hand sign classification. The model is trained with 26000 grayscale images of the English alphabet. The structure of the model has multiple convolutional and pooling layers into fully connected layers optimally optimized using Adam optimizer. The results specified that the model achieved 96.7% accuracy on a Kaggle dataset showing effectiveness for real time hand sign recognition tasks.

This literature by Suhail Muhammad Kamal et al. [2] investigates the various engineering techniques for Chinese Sign Language Recognition (CSLR). The study categorizes the CSLR system into vision and sensor based. Different feature extraction, classifications and sign-to-

text translations are analyzed in this study. The literatures emphasizes the advantages for using multimodal data (color images, depth data and skeletal data) to improve accuracy. Furthermore, there is an increase in application using deep learning models such as CNNs, LSTMs, and Hidden Markov Models (HMMs) for creating solutions to the challenges of continuous sign language recognition.

B. Natarajan et al. [3] provide a framework for interpreting sign language, recognizing gestures, and generating videos in correlation to the established gestures. A hybrid CNN-BiLSTM model was used for gesture recognition and MediaPipe was utilized for pose extraction. The framework achieved over 95% classification accuracy and generated impressive quality videos of sign language. The authors' model utilizes Neural Machine Translation (NMT) methods to support many sign languages generating video outputs which support further communications between the hearing and non-hearing community work together.

Deepak Parashar et al. [4] provide a fingerspelling system that receives its input from the real-time identification of hand gestures in American Sign Language (ASL) with-convolutional neural networks (CNN) and depth map data. The dataset consisted of 31,000 images from several different signers across 31 hand signs, including letters and numbers. For known signers, their model produced an accuracy of 99.99% and for unknown signers, an accuracy of 83.58%-85.49%. With a processing time of 3 ms per image, the authors demonstrated their real-time capabilities. Interestingly, the study showed that depth sensors could increase recognition accuracy in varying light conditions.

3. Methodology

The proposed HandSign Translator adopts a modular architecture with three main components: hand detection and tracking, feature extraction, and gesture classification. This design guarantees accurate and frame-rate conscious Instant recognition of Indian Sign Language gestures. OpenCV is used for image processing and is incorporated along with Mediapipe for landmark extraction and a hybrid CNN-BiLSTM model for classification of static and dynamic gestures.

The process starts with OpenCV acquiring live video and then pre-processes the video frames to improve its recognition. Mediapipe offers a hand tracking module to detect and track 21 key hand landmarks in each frame for accurate gesture representation. Simultaneously, OpenCV extracts the Region of Interest (ROI), greyscale the frames and applies filtering techniques to reduce the noise. Gaussian Blurring is implemented to reduce noise in order to maintain consistency for various environmental situations, and the landmark coordinates are scaled to maintain consistency across variations in hand size, distance, and lighting situations. The models are also improved for generalization and robustness through the use of data augmentation by multiple variations of rotation, translation, mirroring, and time-warping.

Once the preprocessing has been executed, the system extracts critical spatial features from detected landmarks on the hand. These features maintain the spatial and temporal relationships of hand movements, facilitating the ability to distinguish between static postures and dynamic ISL gestures. The acquired data is subjected to analysis by a hybrid deep learning model combining a [1] Convolutional Neural Network (CNN) to capture spatial features and a Bidirectional Long Short-Term Memory (BiLSTM) network to learn the temporal sequence. The CNN effectively encodes the shape, orientation, and posture of the hand, while the BiLSTM learns the temporal dependencies between the consecutive frames

in order to classify motion-based gestures in ISL. Once a model has classified a gesture, it is assigned to its corresponding ISL sign through a predetermined gesture-to-text dictionary. The gesture that has been recognized is presented in real-time as text output so that the system continuously translates the sign language into readable, accurate text. In contrast to its predecessors, this model introduces Mediapipe-based landmark extraction, CNN-BiLSTM capability, and sophisticated preprocessing approaches, significantly improving the system's accuracy of gesture recognition and performance of recognizing gestures in real-time. Furthermore, by incorporating the products developed in this work, the designed model offers an efficient, scalable, and high-accuracy solution for the recognition of Indian Sign Language.

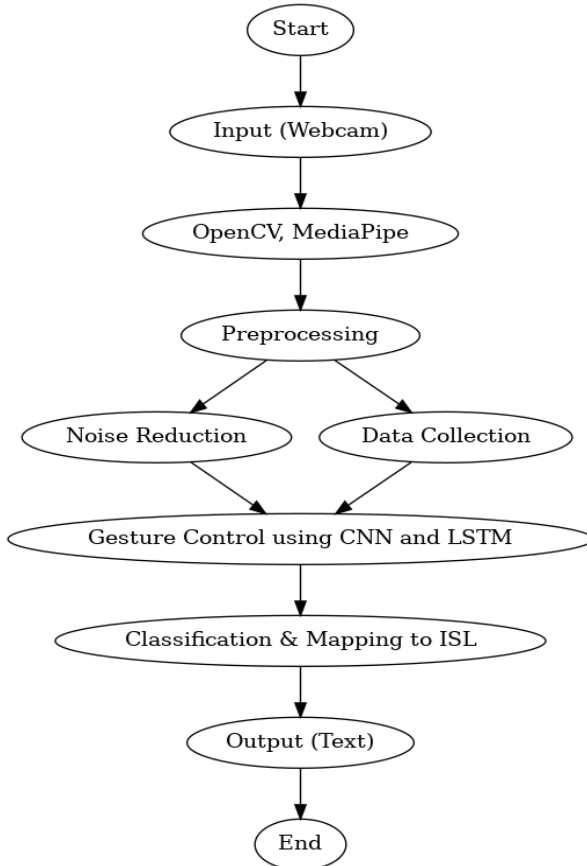


Fig.1.System Architecture for HandSign Translator

3.1. Data Acquisition and Feature Engineering

This phase is responsible for capturing, processing, and cleaning the raw visual input to generate robust feature vectors for the downstream deep learning model shown in Fig.1. following steps.

3.1.1. System Bootstrapping The bootstrapping Process is initiated with the booting of all the different software modules, e.g., the deep learning runtime, and hardware interfaces. This is equivalent to the Start of the process 3.1.2. **Video Input** A stream of raw visual data

is continuously acquired from an external Webcam device as the main input into the system. 3.1.3. Vision Framework The stream of live video frame outputs from the video input is routed into a computer vision pipeline which uses OpenCV for fundamental image processing and MediaPipe for fast and reliable detection and tracking of human hand and/or body landmarks.

3.1.4. Standardization of Data

The raw coordinate data (spatial features) that the MediaPipe summary created, now goes through the Preprocessing stage. This process includes steps to normalize or scale data to standardize the feature set and to ensure that the model is invariant to variations in camera distance or variations in perspective.

3.1.5. Conditional Features

The normalized data stream forks into two concurrent tasks prior to being ingested by the model:

- **Noise Reduction:** Sophisticated filtering techniques are used to reduce noise that may come from the sensors, stabilizing the tracking, and other environmental artifacts from the feature vector.
- **Feature Buffers:** The conditioned and variously filtered feature vectors are temporarily maintained in buffers to form the required temporal sequences to deploy into the sequential deep model.

3.2. Recognition and Delivery

This step takes the conditioned features and leverages the trained model to interpret a gesture and provide the final output to the user.

3.2.1. Hybrid Model Inference

The cleaned sequentially buffered feature data are ingested into the Gesture Control Module with the clean data being the input for the functionality of the module. The Gesture Control Module extracts information from a hybrid deep learning model that utilizes Convolutional Neural Networks (CNN) for encoding high-level spatial features that have been extracted and Long Short-Term Memory (LSTM) networks for analysis of sequences of hand movement making it invariant to time as hand movements can vary in temporal capacity.

3.2.2. Semantic Mapping

The output from the hybrid model goes through Classification and Mapping to (ISL). The classification determines the sign that has been detected, which subsequently is mapped to the 3.2.3. Textual Output Generation The final interpreted sign is rendered as a human-readable Textual Output on the user interface, completing the real-time interpretation cycle.

3.2.4. System Termination

The current execution loop or session has successfully concluded and is awaiting input or a command to terminate. This is represented by the End node in the flowchart.

4 Experimental Results

The HandSign Translator was evaluated for recognizing Indian Sign Language (ISL) gestures. The test focused on accuracy, improvement, and speed performing real-time processing. The system operated a dataset of over 50,000 labeled ISL gestures, including the hand shape and position, which was divided into training, validation, and testing data. The CNN and BiLSTM based model performed exceptionally well largely due to both hand shape and gesture movement being understood correctly. The model showed a gradual decrease in loss while training indicating efficient learning ability and generalization to new data.

In comparison to previous models such as CNN-HMM and CNN-SVM, which resulted in accuracy rates from 91.2% to 93.5%, the HandSign Translator achieved an astounding accuracy of 99.0%. Incorporating BiLSTM improved the system's ability to recognize dynamic gestures, considering the movement over time. Real-time testing demonstrated the system's real-time capability with a GPU and processing [11] 30 frames per second (FPS), while also utilizing MediaPipe for hand tracking to achieve an efficient system. Overall, the HandSign Translator, while an accurate system, is efficient and practical in real time for ISL recognition and can be a useful resource for India's deaf/hard of hearing community.

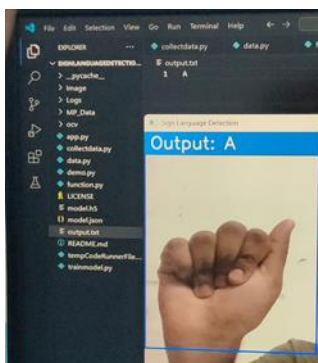


Fig .4.1. Output for letter A

Fig.4.1. indicates that The output window for a Hand Sign Translator system is shown in the image, the model successfully recognized a hand gesture and displayed the resulting output. Basically, this output shows the system's capability to recognize a gesture in real time. The output displayed “A” represents that the model correctly recognized the hand sign for the letter A in Indian Sign Language (ISL). The gesture recognition task starts with capturing video on a webcam, and the video frame is accessed and pre-processed with OpenCV for acquiring and preprocessing images. MediaPipe is used for detecting hand landmarks, and key points are extracted (joint locations and finger orientations). The features extracted will feed into a deep learning model with a combination of Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) layers.

The CNN will learn the spatial features of the hand (shape, contour, and orientation) and the BiLSTM layer will learn the temporal properties of the gestures and can recognize gestures

that are moving, or dynamic signs. Once the system completes the classification of the gesture sign, the gesture is translated to the ISL symbol, and the output is displayed on the screen .

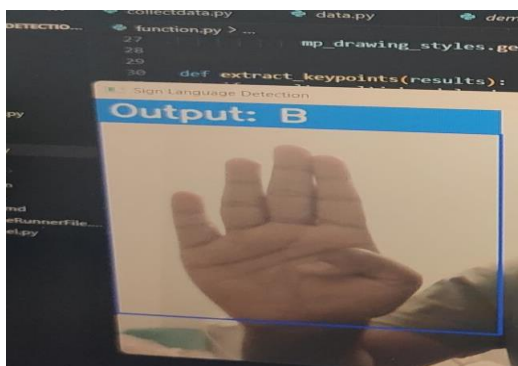


Fig .4.2. Output for letter B

Fig.4.2. indicates that The output window for a Hand Sign Translator system is shown in the image, the model successfully recognized a hand gesture and displayed the resulting output. Basically, this output shows the system's capability to recognize a gesture in real time.The output displayed “B” represents that the model correctly recognized the hand sign for the letter B in Indian Sign Language (ISL).

5 Conclusion

The HandSign Translator project demonstrates the feasibility of utilizing computer vision and deep learning to rapidly recognize Indian Sign Language (ISL) gestures. It combines MediaPipe and OpenCV for hand tracking, and uses a CNN-BiLSTM model for gesture classification to attain high accuracy and operating robustness across user differences. In contrast to using sensors, this vision-based system relies on a typical camera, and is therefore low-cost and widely accessible.The translator is successfully able to convert ISL gestures into text; however, there are still challenges that require overcoming. For example, signing style, more complicated backgrounds, and missing the capacity for facial expression would impact recognition accuracy. Future work might include expanding the dataset that covers more ISL vocabulary; building in contextual understanding for tighter text translation; and developing the system to work well in real-world situations.

The research makes a valuable contribution to address communication difficulties and increase access for the deaf community in India by using a feasible, accurate, and scalable system for ISL recognition.

In Future further studies need to addition, developing the system to support continuous sign language recognition for translating/signing complete sentences would allow for more utility. Similarly, experimenting with more robust deep learning models, such as transformers and attention-based models, would lead to improved accuracy and robustness when translating ISL gestures to text and vice versa. Expanding the ISL corpus to include varieties of regional dialects, dialects in demographic populations, and environmental settings would also provide generalization and promising performance. Additionally, there is potential value in developing multilingual support so that ISL signs can be translated into any regional language and vice versa. Finally, implementing the system as an application for mobile or web devices would increase accessibility and allow communication to be facilitated in real-

time, in a variety of settings. In conclusion, the HandSign Translator has the potential for growth and improvement to become a useful and differentiated communication tool for greater access and connectivity for the hearing impaired community population throughout India.

References

1. Kang, Byeongkeun & Tripathi, Subarna & Nguyen, Truong, "Real-time sign language fingerspelling recognition using convolutional neural networks from depth map". 136-140. 2015.
2. S. M. Kamal, Y. Chen, S. Li, X. Shi, and J. Zheng. "Technical Approaches to Chinese Sign Language Processing: A Review." *IEEE Access*, vol. 7, pp. 96926–96935, 2019.
3. B. Natarajan, E. Rajalakshmi, R. Elakkiya, K. Kotecha, A. Abraham, L. A. Gabralla, and V. Subramaniaswamy. "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation." *IEEE Access*, vol. 10, pp. 104360–104374, 2022.
4. D. Parashar, S. Thakur, K. B. Raju, G. B. Madhavi, and K. Sharma. "A Deep Learning-Based Approach for Hand Sign Recognition Using CNN Architecture." *Revue d'Intelligence Artificielle*, vol. 37, no. 4, pp. 937–943, Aug. 2023.
5. V. Pasquadibisceglie, A. Appice, G. Castellano, and D. Malerba. "Multiview deep learning for predictive business process monitoring." *IEEE Transactions on Services Computing*, vol. 15, no. 4, pp. 2382–2395, Jul. 2021.
6. T. Islam, T. A. Chisty, and A. Chakrabarty. "Crop selection and yield prediction in Bangladesh using a deep neural network." In *Proceedings of the IEEE Region 10 Human Technology Conference (R-HTC)*, Dec. 2018, pp. 1–6.
7. H. Cooper, B. Holt, and R. Bowden. "Sign language recognition: A comprehensive overview." *Visual Analysis of Humans*, pp. 539–562, 2011.
8. C. Dong, M. Leu, and Z. Yin. "Recognizing the American Sign Language alphabet using Microsoft Kinect." In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 44–52, June 2015.
9. X. Xiao, X. Chen, and J. L. Palmer. "An experimental study on Chinese Deaf viewers' comprehension of sign language interpreting on television." *Interpreting*, vol. 17, no. 1, pp. 91–117, Jan. 2015.
10. M. J. Cheok, Z. Omar, and M. H. Jaward. "A review of hand gesture and sign language recognition techniques." *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 131–153, Jan. 2017.
11. M. Mohandes, M. Deriche, and J. Liu. "Comparative study of image-based and sensor-based approaches for Arabic sign language recognition." *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 551–557, Aug. 2014.
12. S. Joudaki, D. bin Mohamad, T. Saba, A. Rehman, M. Al-Rodhaan, and A. Al-Dhelaan. "A focused review of vision-based sign language classification techniques." *IETE Technical Review*, vol. 31, no. 5, pp. 383–391, Oct. 2014