IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151105

TWO-WAY INDIAN SIGN LANGUAGE COMMUNICATION ASSISTANT USING MACHINE LEARNING

Rashmi R[1], Bhavana L[2], Sarah Ninan[3], Adithya Pratheep[4], Dr. Beenarani Manoj[5]

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, Chennai, India[1-4]

Guide, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, Chennai, India[5]

Abstract: It can be sometimes a tough challenge for the Indian Sign Language (ISL) user to communicate with the non-ISL learner. In this regard, we decided to address the issue by creating a Two Way Indian Sign Language Communication Assistant using Machine Learning. Instead of simply converting the sign to speech, our system focuses on both sides of communication. The system also provides the extra functionality for ISL learners. A webcam is used for capturing the sign gestures. The points of interest for hands, face and body are identified using MediaPipe Holistic. All these points are taken under consideration for recognizing the gestures through LSTM model analysis of multiple video frames. Signs recognized are then converted to text and spoken. In case of communication by speech, it takes input from the microphone, converts to text, and shows corresponding signs in the form of images/videos/animations. Apart from this, there is a learning mode, where one can practice and get familiar with the sign gestures of ISL. All these features combine sign recognition, speech conversion and learning of sign language into a single application. The goal of this project is to ease the communication between ISL users and non-signers and also to create more awareness about the language.

Keywords: Indian Sign Language , Sign Language Recognition, Two-Way Communication, Speech-to-Sign Translation, LSTM, MediaPipe Holistic, Gesture Recognition, Human-Computer Interaction, Assistive Technology, Sign Language Learning

1. INTRODUCTION

Most hearing people in India have never learned Indian Sign Language (ISL). For Deaf and hard-of-hearing individuals, this means that something as routine as a bank visit or a doctor's appointment often depends on whether an interpreter happens to be available. Often, none is. Computer vision can now track the hands and make a reasonable guess at what a gesture means, and a number of projects already use it to turn signs into text. Most of them stop there. The hearing person has no easy way to reply, and a curious beginner has no way to learn. A single system is required that covers both directions and teaches as well. The assistant has three modules. Sign-to-Speech reads gestures through a webcam and produces text along with a spoken voice. Speech-to-Sign works in reverse: it takes what a person says into the microphone and displays the matching ISL signs on screen. Learning Mode is meant for beginners. A learner attempts a sign in front of the camera and is told straight away how close it was.Everything runs on an ordinary laptop with a webcam and a microphone. Gloves and sensors were ruled out early, mainly because they cost money that many users cannot spare. The aim is modest. A short conversation between a signer and a non-signer should feel less like a struggle than it does today, while ISL should feel a little less out of reach for anyone willing to try learning it.

2. RELATED WORKS

The project began by reading what others had already done on sign language recognition. It was a lot, and not all of it was relevant to us. The first thing we used were datasets. INCLUDE, by Sridhar et al., is a large collection of ISL videos [1]. Joshi et al. built CISLR for word-level recognition [2]. A model can’t learn a sign it has barely seen. That’s why these datasets matter. CISLR also reminded us that signs are not still pictures. Then there is the question of speed. Nandy et al. tested real-time recognition of isolated ISL gestures [3]. That is useful, but it only gets you halfway. The system says "this was the sign for water" and then what? The other person still has to read or hear it. For the model itself we considered CNNs and LSTMs. CNNs are good at finding patterns in images, and LeCun et al. is the work usually cited for that [4]. One image can't hold a sign that takes a second or two to perform. LSTMs are built for that, since they look at a run of frames together. It marks landmarks on the hands and body in real time, so we can track a set of points instead

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2026.151105

of a whole video frame. Facial expression and posture can change the meaning of a sign too, and those are easier to follow this way. The research included Whisper by Radford et al. [6], a model trained on large amounts of audio to convert speech into text. The resulting text is then matched with corresponding signs. The aim was to develop a single tool that recognises signs, converts speech into signs, and provides a learning platform for beginners to practise ISL.

3. SYSTEM ARCHITECTURE

The project includes mainly three parts. That is sign to speech, speech to sign and learning mode. Each of these parts have their own functionality. The first two parts assist people to communicate with each other. The third part helps people to learn sign language. All these features combined into one app.

3.1. Sign to Speech Module

Sign to speech module is used when someone who knows ISL, tries to communicate with someone who doesn't understand ISL. A webcam is used, where the user shows the sign and the movement is captured. Here OpenCV is used to process the video and MediaPipe Holistic is used to find important points of the user's hand, face and body. This points provides the information about the gesture done by the user. After that, the LSTM model is used. The LSTM model observes those points in a series of frames and tries to recognize the sign. This is very important because gesture is the movement, not just the one particular position of the hand. When the sign is recognized. The system will display the text and also a text to speech tool is used to speak that text.

3.2. Speech to Sign Module

This module is intended to be on the other end of the conversation. In this case, you can just talk normally and the system will try to interpret the message in sign language. It records the speech using a microphone and converts it to text using Whisper. Now the application will try to find signs in its collection of words. And if it finds a match, then it will display this sign on the screen in the form of an image or video or animation. The problem is, maybe it doesn't have all the words that you use. If it doesn't find a match, it may not be able to show that part of your message. So there is room to improve this by adding more signs to its collection.

3.3 Learning Mode

A Learning Model has also been added for those who are new to ISL. The user can choose a sign and it will show them how it is done. Then the user can try to make the same gesture in front of their webcam. If the model supports that sign, then it can compare the user's gesture and the expected sign and can give feedback based on that. So the user can practice and learn at their own pace.

3.4 System Components

Various tools are used for different purposes.

- For the main program and connecting the features we use Python.
- For video of a webcam and its frames we use OpenCV.
- Important points for the hand, face and body using MediaPipe Holistic.
- For training and running machine learning models, we use TensorFlow/Keras.
- For basic models (just for comparison) and evaluation of the results, we use Scikit-learn.
- We convert speech to text with Whisper.
- We transform recognized text to speech with pyttsx3
- And for interaction of users, we use an interface that is built with Streamlit.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151105

4. RESULTS

The goal of the system is to do sign recognition, speech recognition and ISL learning all in one app. The system needs to be tested to ensure that it works throughout an entire conversation. For the Sign to speech, the webcam captures a gesture. Landmarks are extracted and used by the model to predict the sign, and then text and speech output is generated. In any of the steps, there could be a weak point. It needs to be tested by using some examples that the model has not seen. The model is supposed to differentiate between the signs. For the speech to sign there are two possibilities of the weak point. The first could be the misunderstanding of speech using whispers. In another case, even if the text is correct it may predict the wrong sign. If there is no entry or wrong entry in the sign map, then either user will see nothing or wrong. So, the sign collection is as important as the speech recognition. For learning mode. Can the learner open the demo? Try the sign and get proper feedback. Can a person without technical background navigate through it?

There are many things that can go wrong. The bad lighting, strange camera angle, cluttered background, signing too quickly, or just your unique style can affect recognition. The signs may look similar so it'll get confused. For the speech, the noise or accent can affect the errors. Also the problem is, ISL isn't just the hand movement. It has to do with facial expression, posture, context etc. to change meaning. And it isn't like speech where words can be replaced with gestures. It will be incorrect for a system to treat it that way. The system doesn’t have end-to-end measured accuracy yet. A complete set of experimental results is not yet available for reporting. How good it can do still has to be shown. The system needs to be tested using a proper dataset, with separate training and testing data. Measure the accuracy, precision, recall, response time. Feedback from user.

5. CONCLUSION

This is the project to help communication between native sign language users (ISL) and non-signer users. It contains Sign-to-Speech, Speech-to-Sign and Learning modules which helps in daily communication and learning of ISL. This is a sign-to-speech, speech-to-sign, learning based system that uses computer vision, landmarks, LSTM for gestures, speech recognition, and text-to-speech. The project only requires a laptop, webcam, mic, no need for gloves or sensors. So it's easy to test for people.

Other systems are focused on sign reading. But this system is both ways. Plus it has a learning mode which helps novices to practice. But there are many issues,sign recognition has to be made reliable. More mapping, sign difference, face reading and body language reading. These need careful designing and testing. By doing all this if the system can be made successfully. It will help in inclusive communication and also increase the interest in learning Indian sign language.

REFERENCES

[1] A. Sridhar, R. G. Ganesan, P. Kumar, and M. Khapra, "INCLUDE: A Large Scale Dataset for Indian Sign Language Recognition," in Proc. 28th ACM Int. Conf. Multimedia (MM '20), 2020, pp. 1366–1375, doi: 10.1145/3394171.3413528.
[2] A. Joshi, S. Agrawal, and A. Modi, "CISLR: Corpus for Indian Sign Language Recognition," in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2022, pp. 10357–10366.
[3] A. Nandy, J. S. Prasad, S. Mondal, P. Chakraborty, and G. C. Nandi, "Recognition of Isolated Indian Sign Language Gesture in Real Time," in Information Processing and Management (BAIP 2010), Communications in Computer and Information Science, vol. 70, Springer, 2010, pp. 102–107.
[4] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural Sign Language Translation," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7784–7793.
[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[7] F. Zhang et al., "MediaPipe Hands: On-device Real-time Hand Tracking," arXiv:2006.10214, 2020.
[8] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," arXiv:1906.08172, 2019.
[9] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust Speech Recognition via Large-Scale Weak Supervision," arXiv:2212.04356, 2022.

 This work is licensed under a Creative Commons Attribution 4.0 International License