



A Review of Deep Learning for Real-Time Sign Language Recognition

Minal V. Patil¹, Pravin I. Patil², Dr. D. M. Patel³

Research Scholar, M.E. Computer Sci. and Engg, Shri Jaykumar Rawal Institute of Technology, Dondaicha, Maharashtra, India.¹

HOD, M.E. Computer Sci. and Engg, Shri Jaykumar Rawal Institute of Technology, Dondaicha, Maharashtra, India²

Principal, M.E. Computer Sci. and Engg, Shri Jaykumar Rawal Institute of Technology, Dondaicha, Maharashtra, India³

Abstract: Communication is the fundamental exchange of thoughts and messages through speech, signals, visuals, and behavior. For deaf and hard-of-hearing individuals, sign language serves as the primary medium for interaction. However, because most non-disabled individuals do not understand sign language, a significant communication barrier exists. This paper presents a user-friendly Human-Computer Interface (HCI) software system designed to translate American Sign Language (ASL) fingerspelling hand gestures into text and speech in real time. The proposed system captures gesture images via a webcam, extracts hand landmarks using the MediaPipe library, and projects them onto a clean white canvas to overcome real-world lighting and background variations. A Convolutional Neural Network (CNN) model is trained on these skeletal representations to recognize individual alphabets (A–Z) and form complete words. The recognized text is then converted into spoken audio using the pyttsx3 library. The proposed approach achieves 97% recognition accuracy under standard conditions and over 99% accuracy in controlled environments.

Keywords: American Sign Language, Convolutional Neural Network, MediaPipe, Human-Computer Interface, Gesture Recognition, Text-to-Speech Conversion.

I. INTRODUCTION

American Sign Language (ASL) is a predominant sign language used by millions of individuals worldwide. Since the primary disability experienced by deaf and non-verbal individuals is communication-related when using spoken languages, sign language serves as their sole medium to exchange ideas, learn, work, access public services, and participate fully in society. Communication is defined as the process of exchanging thoughts, ideas, and messages through speech, signals, behavior, and visuals. Deaf and non-verbal people make use of hand gestures to visually convey nonverbal messages.

In our project, we basically focus on producing a model which can recognize fingerspelling-based hand gestures in order to form a complete word by combining each gesture.

A. Objective

More than 70 million deaf people around the world use sign languages to communicate. Sign language allows them to learn, work, access services, and be included in their communities. It is hard to make everybody learn the use of sign language with the goal of ensuring that people with disabilities can enjoy their rights on an equal basis with others. So, the aim is to develop a user-friendly human computer interface (HCI) where the computer understands American Sign Language. This project will help dumb and deaf people by making their life easy. To create computer software and train a model using CNN which takes an image of a hand gesture of American Sign Language, shows the output of the particular sign language in text format, and converts it into audio format.

B. Scope

This system will be beneficial for both dumb/deaf people and people who do not understand sign language. They just need to communicate with sign language gestures and this system will identify what he/she is trying to say. After identification, it gives the output in the form of text as well as speech format.

II. LITERATURE REVIEW

A. Mahesh Kumar N. B. (2018)

This paper shows the sign language recognition of 26 hand gestures in Indian Sign Language using MATLAB. The proposed system contains four modules: pre-processing and hand segmentation, feature extraction, sign recognition, and sign to text. By using image processing, segmentation can be done using the Otsu algorithm. Features such as Eigenvalues



and Eigenvectors are extracted for recognition. Linear Discriminant Analysis (LDA) was used for gesture recognition, and recognized gestures are converted into text and voice format.

B. Krishna Modi (2013)

In this proposed system, they intend to recognize basic elements of sign language and translate them to text. Video is captured frame-by-frame, processed, and appropriate images are extracted using BLOB analysis. The retrieved images are sent to a statistical database where captured images are compared with stored samples to determine performed alphabet signs. Implementing American Sign Language fingerspelling, they construct words and sentences with a 93% accuracy probability.

C. Bikash K. Yadav et al. (2020)

Sign language is one of the oldest and most natural forms of communication. They propose a real-time method using Convolutional Neural Networks (CNN) for fingerspelling-based ASL. The hand image is passed through a filter and then through a classifier that predicts the class of hand gestures, reaching a model accuracy of 95.8%.

D. Ankit Ojha et al. (2020)

Created a desktop application using a webcam to capture ASL gestures and translate them into corresponding text and speech in real time. The architecture uses a Convolutional Neural Network (CNN) for gesture detection. Modules include image acquisition, hand region segmentation, hand detection and tracking, posture recognition, and text/speech display, achieving an accuracy of 95%.

E. Vectorial Adebimpe Akano (2018)

Converts ASL hand gestures into text and speech using unsupervised feature learning. Sample images were collected using a Kinect sensor via MATLAB. Five hundred data samples were collected for training. The combination of FAST and SURF with KNN (k=10) achieved 92% accuracy with supervised feature learning and 78% with unsupervised feature learning.

F. Rakesh Kumar (2021)

Proposed a hand gesture recognition algorithm based on image contours to identify standard ASL. The proposal optimizes performance overhead through identification of 17 characters and 6 symbols using image contours and convexity measurement without complex algorithms or specialized hardware, obtaining an average accuracy of 86%.

III. SYSTEM ANALYSIS

A. Literature Comparison

TABLE I
COMPARISON OF EXISTING LITERATURE

Author Name	Algorithm / Classifier	Accuracy	Year
Mahesh Kumar N. B.	LDA	79%	2018
Krishna Modi	Blob Analysis	91%	2013
Bikash K. Yadav	CNN	95.5%	2020
Aayush Pandey / Ankit Ojha	CNN	90%	2020
Vectorial Adebimpe Akano	KNN	91%	2018
Rakesh Kumar	Contour Measurement	88%	2021

B. Research Gap

In the first research paper [1], they used the LDA algorithm and converted RGB images to binary images, but the image processing is not good enough to extract detailed features of signs.

In the second research paper [2], signs were recognized using direct image pixel comparison against a database after converting RGB to binary, which removed essential spatial features.

In the third and fourth research papers [3], [4], they applied CNN algorithms for sign recognition effectively, but lacked comprehensive pre-processing before feeding data to the model.

In the fifth research paper [5], they used KNN, a simple algorithm, without sufficient image pre-processing, resulting in moderate accuracy.

In the sixth research paper [6], contour and convexity measurements were used, but the algorithm failed to yield high accuracy.



C. Project Feasibility Study

1) Operational Feasibility: The purpose of this system is to handle tasks accurately and efficiently with minimal time consumption. The application is user-friendly, requiring only basic knowledge of American Sign Language and standard desktop operating procedures.

2) Technical Feasibility: Front-end and back-end selection were based on extensive platform research. Python Tkinter was chosen for the front-end GUI to assist non-IT users with a simple interface. Python was selected for the back-end due to its broad scientific library support (OpenCV, MediaPipe).

3) Economic Feasibility: Since the system is developed as part of project work and uses accessible webcam hardware, there are no manual operational costs, making it economically practical.

D. Detailed Module Description

1) Data Acquisition: Data acquisition approaches include electromechanical gloves and vision-based methods. While gloves provide exact positioning, they are expensive and non-ergonomic. Vision-based methods use a standard webcam, creating natural interaction at low cost despite challenges from lighting, skin color variations, and background noise.

2) Pre-Processing & Feature Extraction: To eliminate background interference, hand landmarks are extracted using MediaPipe. Landmark points are rendered and connected on a plain white canvas using OpenCV. This skeletal representation removes background clutter and lighting dependencies. A total of 180 skeleton images were collected for each alphabet sign (A–Z).

E. Text and Speech Translation

The model translates recognized gestures into text words. The pyttsx3 library converts predicted text into speech in real time, simulating real-life conversation.

IV. CONCLUSION AND FUTURE WORK

Using our approach, we can estimate any character within the [a–z] set with 97% accuracy under standard conditions, and above 99% accuracy in clean, controlled environments. Future work includes developing a native Android application for mobile gesture prediction.

REFERENCES

- [1]. Alyami S., Luqman H., Hammoudeh M., "Reviewing 25 years of continuous sign language recognition research: Advances, challenges, and prospects," *Information Processing & Management*, vol. 61, no. 5, 2024, Art. 103774. DOI: 10.1016/j.ipm.2024.103774.
- [2]. Obi Y., Claudio K. S., Budiman V. M., Achmad S., Kurniawan A., "Sign language recognition system for communicating to people with disabilities," *Procedia Computer Science*, vol. 216, 2023, pp. 13–20. DOI: 10.1016/j.procs.2022.12.106.
- [3]. Alsolai H., Alsolai L., Al-Wesabi F. N., Othman M., Rizwanullah M., Abdelmageed A. A., "Automated sign language detection and classification using reptile search algorithm with hybrid deep learning," *Heliyon*, vol. 10, no. 1, 2024, e23252. DOI: 10.1016/j.heliyon.2023.e23252.
- [4]. Arooj S., Altaf S., Ahmad S., Mahmoud H., Mohamed A. S. N., "Enhancing sign language recognition using CNN and SIFT: A case study on Pakistan sign language," *Journal of King Saud University – Computer and Information Sciences*, vol. 36, no. 2, 2024, Art. 101934. DOI: 10.1016/j.jksuci.2024.101934.
- [5]. Liu Y., Jiang X., Yu X., Ye H., Ma C., Wang W., Hu Y., "A wearable system for sign language recognition enabled by a convolutional neural network," *Nano Energy*, vol. 116, 2023, Art. 108767. DOI: 10.1016/j.nanoen.2023.108767.
- [6]. Koyineni S. N., Sai G. K., Anvesh K., Anjali T., "Silent expressions unveiled: Deep learning for British and American sign language detection," *Procedia Computer Science*, vol. 233, 2024, pp. 269–278. DOI: 10.1016/j.procs.2024.03.216.
- [7]. Bora J., Dehingia S., Boruah A., Chetia A. A., Gogoi D., "Real-time Assamese sign language recognition using MediaPipe and deep learning," *Procedia Computer Science*, vol. 218, 2023, pp. 1384–1393. DOI: 10.1016/j.procs.2023.01.117.
- [8]. Liu T., Tao T., Zhao Y., Li M., Zhu J., "A signer-independent sign language recognition method for the single-frequency dataset," *Neurocomputing*, vol. 582, 2024, Art. 127479. DOI: 10.1016/j.neucom.2024.127479.
- [9]. Lu X., Wang H., Zhang J., Zhang Y., Zhong J., Zhuang G., "Research on J-wave detection based on transfer learning and VGG16," *Biomedical Signal Processing & Control*, vol. 95, 2024, Art. 106420. DOI: 10.1016/j.bspc.2024.106420.
- [10]. Chen Y. C., Fu S., Yin W., Liu K., Qian S., "VGG16-based intelligent image analysis in the pathological diagnosis of IgA nephropathy," *Journal of Radiation Research and Applied Sciences*, vol. 16, no. 3, 2023, Art. 100626. DOI: 10.1016/j.jrras.2023.100626.



- [11]. Singh D. K., "3D-CNN based dynamic gesture recognition for Indian sign language modelling," *Procedia Computer Science*, vol. 189, 2021, pp. 76–83. DOI: 10.1016/j.procs.2021.05.071.
- [12]. Rahaman M. A., Oyshe K. U., Chowdhury P. K., Debnath T., Rahman A., Khan M. S. I., "Computer vision-based six layered convolutional neural network to recognize sign language for both numeral and alphabet signs," *Biomimetic Intelligence and Robotics*, vol. 4, no. 1, 2024, Art. 100141. DOI: 10.1016/j.birob.2023.100141.
- [13]. Suharjito R. A., Wiryana F., Ariesta M. C., Kusuma G. P., "Sign language recognition application systems for deaf-mute people: A review based on input–process–output," *Procedia Computer Science*, vol. 116, 2017, pp. 441–448. DOI: 10.1016/j.procs.2017.10.028.
- [14]. Bilge Y. C., Ikizler-Cinbis N., Cinbis R. G., "Cross-lingual few-shot sign language recognition," *Pattern Recognition*, vol. 151, 2024, Art. 110374. DOI: 10.1016/j.patcog.2024.110374.
- [15]. Mohan P., Sabarwal T., Preethiya T., "Indian sign language character recognition system," in *Proc. 4th Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, 2023, pp. 1029–1034. DOI: 10.1109/ICESC57686.2023.10193309.
- [16]. Diba B. S., Plabon J. D., Mahmudur Rahman M., Mistry D., Saha A. K., Mridha M., "Explainable federated learning for privacy-preserving Bangla sign language detection," *Engineering Applications of Artificial Intelligence*, vol. 134, 2024, Art. 108657. DOI: 10.1016/j.engappai.2024.108657.
- [17]. Shenoy K., Dastane T., Rao V., Vyavaharkar D., "Real-time Indian sign language (ISL) recognition," in *Proc. 9th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, 2018, pp. 1–9. DOI: 10.1109/ICCCNT.2018.8493808.
- [18]. Liu T., Tao T., Zhao Y., Zhu J., "A two-stream sign language recognition network based on key-frame extraction method," *Expert Systems with Applications*, vol. 253, 2024, Art. 124268. DOI: 10.1016/j.eswa.2024.124268.
- [19]. Vashisth H. K., Tarafder T., Aziz R., Arora M., Alpana, "Hand gesture recognition in Indian sign language using deep learning," *Engineering Procedia*, vol. 59, no. 1, 2023.
- [20]. Sreemathy R., Turuk M., Chaudhary S., Lavate K., Ushire A., Khurana S., "Continuous word level sign language recognition using an expert system based on machine learning," *International Journal of Cognitive Computing in Engineering*, vol. 4, 2023, pp. 170–178. DOI: 10.1016/j.ijcce.2023.04.002.
- [21]. Sarma N., Talukdar A. K., Sarma K. K., "Real-time Indian sign language recognition system using YOLOv3 model," in *Proc. 6th Int. Conf. Image Inf. Process. (ICIIP)*, vol. 6, 2021, pp. 445–449. DOI: 10.1109/ICIIP53038.2021.9702611.
- [22]. Saini B., Venkatesh D., Chaudhari N., Shelake T., Gite S., Pradhan B., "A comparative analysis of Indian sign language recognition using deep learning models," *Forum Linguisticum Studiorum*, vol. 5, 2023, pp. 197–222.