STM JOURNALS

Journal of
Image Processing & Pattern Recognition Progress

ISSN: 2394-1995
Volume 12, Issue 2, 2025
May–August
DOI (Journal): 10.37591/JoIPPRP

https://journals.stmjournals.com/joipprp

Review JoIPPRP

Sign Talk for Blind and Deaf Translating Hand Gestures into Audible and Textual Communication

S. Saraswat[1], Aditya Gulhane[2,]*, Sujal Sahu[2], Jyotiraditya More[2], Shraddha Wakale[2]

Abstract

Sign language serves as a vital communication method for individuals who are deaf or have hearing impairments. It relies on hand gestures, facial cues, and body language to convey messages and express meaning. However, many people who do not know sign language find it hard to communicate with sign language users. The latest progress in artificial intelligence and computing has enabled the creation of systems that can automatically recognize sign language. These systems can help recognize hand gestures in real-time, making it easier for individuals with hearing difficulties to communicate. Traditionally, recognizing gestures required special equipment like sensors or cameras, which could be expensive and not widely available. Now, using deep learning and computer vision techniques, we can recognize gestures directly from regular webcam feeds, which is more affordable and accessible. In our project, we created a real-time hand gesture recognition system that uses a webcam to capture hand signs and processes them with a deep learning model. It then converts these gestures into text and voice outputs. This system aims to help people with hearing and speech challenges by providing a simple communication tool. Our study explains how the system works, how we built it, and how well it performs. The findings indicate that deep learning is capable of accurately identifying hand gestures, thereby enhancing communication for individuals who are deaf or mute.

Keywords: Hand gesture recognition, sign language translation, deep learning, computer vision, CNN

INTRODUCTION

Sign language is a crucial form of communication for those who are deaf or mute. However, since most people are not familiar with it, mutual understanding can be challenging. To solve this problem, we suggest a system that can recognize hand gestures in real-time and turn them into text and speech using advanced technology. Our system uses a webcam to take pictures of hand gestures. It then processes these images with a trained model to identify the gestures and convert them into written words and spoken language. We have created a simple interface using Tkinter that makes it easy for users to interact in real-time. The model is trained using an extensive dataset of American Sign Language (ASL) gestures to guarantee precise recognition. This system does not require any special devices or expensive equipment, making it an affordable and accessible way to help with sign language communication. Tests show that our approach works well and accurately identifies gestures. This study demonstrates how artificial intelligence can assist in overcoming communication obstacles for individuals with hearing and speech challenges, opening the door for more advancements in real-time sign language translation.

***Author for Correspondence**
Aditya Gulhane
E-mail: adityagulhane20@gmail.com

[1]Professor, Department of Information Technology, Pune District Education Association's College of Engineering, Manjari (Bk), Pune, Maharashtra, India
[2]Student, Department of Information Technology, Pune District Education Association's College of Engineering, Manjari (Bk), Pune, Maharashtra, India

Received Date: March 25, 2025
Accepted Date: April 16, 2025
Published Date: April 28, 2025

Citation: S. Saraswat, Aditya Gulhane, Sujal Sahu, Jyotiraditya More, Shraddha Wakale. Sign Talk for Blind and Deaf Translating Hand Gestures into Audible and Textual Communication. Journal of Image Processing & Pattern Recognition Progress. 2025; 12(2): 20–26p.

LITERATURE SURVEY

Many studies have been done on recognizing sign language using different types of sensors and machine learning techniques. Here is a simple overview of some recent research in this area.

In the study *"Signformer: Deep Vision Transformer for Sign Language Recognition"*, researchers developed a new method using a type of AI called a Vision Transformer (ViT) to recognize static Indian Sign Language. The goal is to create better recognition systems to help hearing-impaired and non-hearing-impaired people communicate. Their method achieves an impressive accuracy of 99.29% with very few training hours, and it works well even with different techniques for improving recognition [1].

The paper titled *"Hand Gesture Recognition and Voice, Text Conversion using CNN and ANN"* discusses a system developed to enhance communication for individuals who are blind and deaf. Since sign language varies, it can be hard for people with hearing and visual impairments to express themselves, especially in emergencies. The researchers utilized a camera to detect hand gestures and employed Convolutional Neural Networks (CNN) along with Artificial Neural Networks (ANN) to translate those gestures into text and speech, thereby enhancing communication. They also looked at the pros and cons of recognizing hand motions in real-life situations [2].

In the study titled *"Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects"*, researchers analyzed 58 papers on Sign Language Translation Systems (SLTS) and explored the role of artificial intelligence in enhancing sign language recognition. Although sign language is important, it has not been studied as much due to its unique features. This review highlights major advancements and challenges in the field, showing that deep learning techniques like CNN and Recurrent Neural Networks (RNN) can accurately recognize gestures. They also found that considering facial expressions and body movements can improve accuracy. Future work should focus on making these models better and incorporating more non-manual features to enhance performance [3].

The paper titled *"Real-Time Hand Sign Language Translation: Text and Speech Conversion"* introduces a system that uses a webcam to instantly convert sign language into text and speech. The researchers used a library called Media Pipe to identify hand positions from video footage. They saved these coordinates in a file to train a Random Forest algorithm that identifies different sign language gestures. In real-time, the model predicts the gesture being signed and shows the results on the live video feed. This study integrates computer vision with machine learning to help recognize nonverbal communication [4].

METHODOLOGY

The system we propose is built to identify hand gestures and convert them into both text and spoken output (Figure 1). The process has three main parts: preparing the dataset, training the model, and detecting gestures in real time. Here is a breakdown of each part.

Dataset Preparation

The first step in creating a hand gesture recognition system is to prepare a good dataset. Since we want to recognize American Sign Language (ASL) gestures, we used a collection of images that show different hand gestures for letters, words, or commands. We either took images from an existing dataset (like the ASL Alphabet Dataset) or captured them ourselves using OpenCV. The dataset has many images for each gesture taken in different lighting, angles, and hand positions to help the model learn better [5]. We applied data augmentation methods to expand the image dataset and enhance the robustness of the model. For image preprocessing, we converted the images to grayscale to make them simpler to process and resized them to a uniform size (like 64×64 pixels) to keep everything consistent.

Model Selection for Gesture Recognition

For gesture classification, we used a Convolutional Neural Network (CNN) because it seems good at picking up details from images. The model has several layers that use filters to find edges, shapes, and

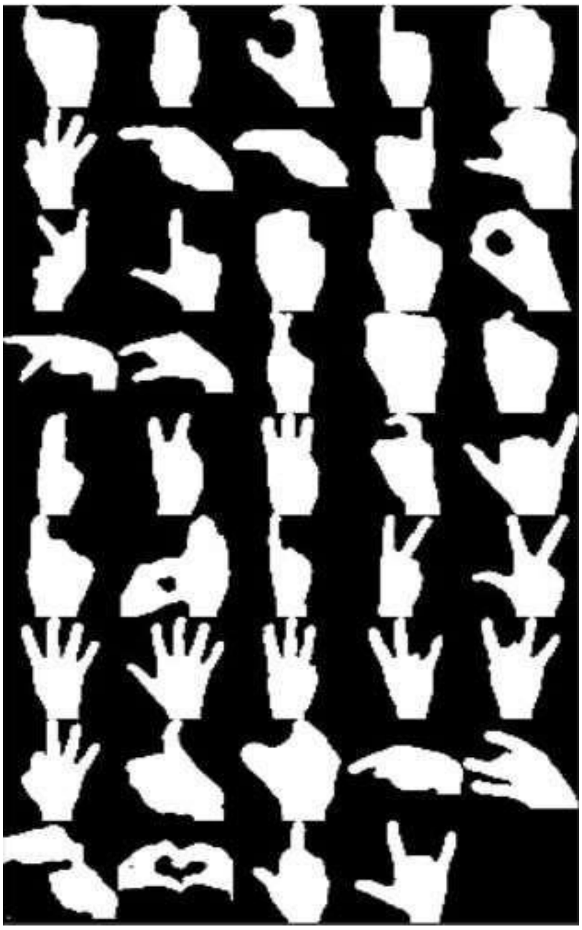

Figure 1. Hand gestures.

patterns in hand gestures. ReLU activation functions are added to help introduce some nonlinearity [6]. There are also max pooling layers that help reduce the size of the feature maps, keeping important details while making things more efficient. The extracted features are subsequently fed into fully connected layers, which analyze the patterns and classify the corresponding gestures. To help avoid overfitting, we included dropout layers that randomly turn off some neurons during training. Finally, a SoftMax layer gives probability scores for each class, helping to helping determine the most likely gesture. Overall, this CNN approach aims to provide accurate and real-time gesture recognition [7].

Real-Time Gesture Recognition

The real-time gesture recognition function utilizes a webcam to record hand movements and converts them into both text and audio output. It uses OpenCV to access the webcam and show live video in a Tkinter-based interface. The program tries to make the video run smoothly. Each frame it captures is changed a bit, like turning it into grayscale, resizing, and normalizing it, to try and make it easier to recognize gestures. Some background removal techniques are applied to help focus on the hand, which might lead to better results [8].

Then, the processed image goes to a trained CNN model using TensorFlow and Keras, which looks for features and tries to identify the gesture with a score. The identified gesture shows up as text in the Tkinter interface. SQLite3 is used to keep track of the gestures recognized, which could be useful for later analysis. To assist users with vision issues, it also uses pyttsx3 to read the recognized text aloud [9]. An alert message box is included, and Pickle is utilized to efficiently save and load the trained model. The Tkinter interface allows users to start or stop the camera, see real-time results, hear the spoken gesture, and reset the system. This aims to provide a friendly and helpful way to translate sign language accurately [10].

ALGORITHM

CNN stands for Convolutional Neural Networks, which are mainly used for looking at images and videos. They help with things like recognizing images, detecting objects, and segmenting parts of images. CNNs consist of four primary types of layers:

1. *Convolutional layer:* In a normal neural network, every input neuron connects to the next layer. But in a CNN, only a small part of the input neurons connects to a neuron in the hidden layer.
2. *Pooling layer:* This layer helps to make the feature maps smaller and less complex. There are usually several pooling layers in the hidden layer of a CNN.
3. *Flatten:* Flattening means turning the data into a single line (1-dimensional array) so it can be used in the next layer. We do this by taking the output from the convolutional layers and turning it into a long feature vector.
4. *Fully connected layer:* These layers are some of the last ones in the network. The input to these layers comes from the final pooling or convolutional layer, which has already been flattened before being sent to the fully connected layer.

SYSTEM ARCHITECTURE

The workflow diagram of system architecture is shown in Figure 2.

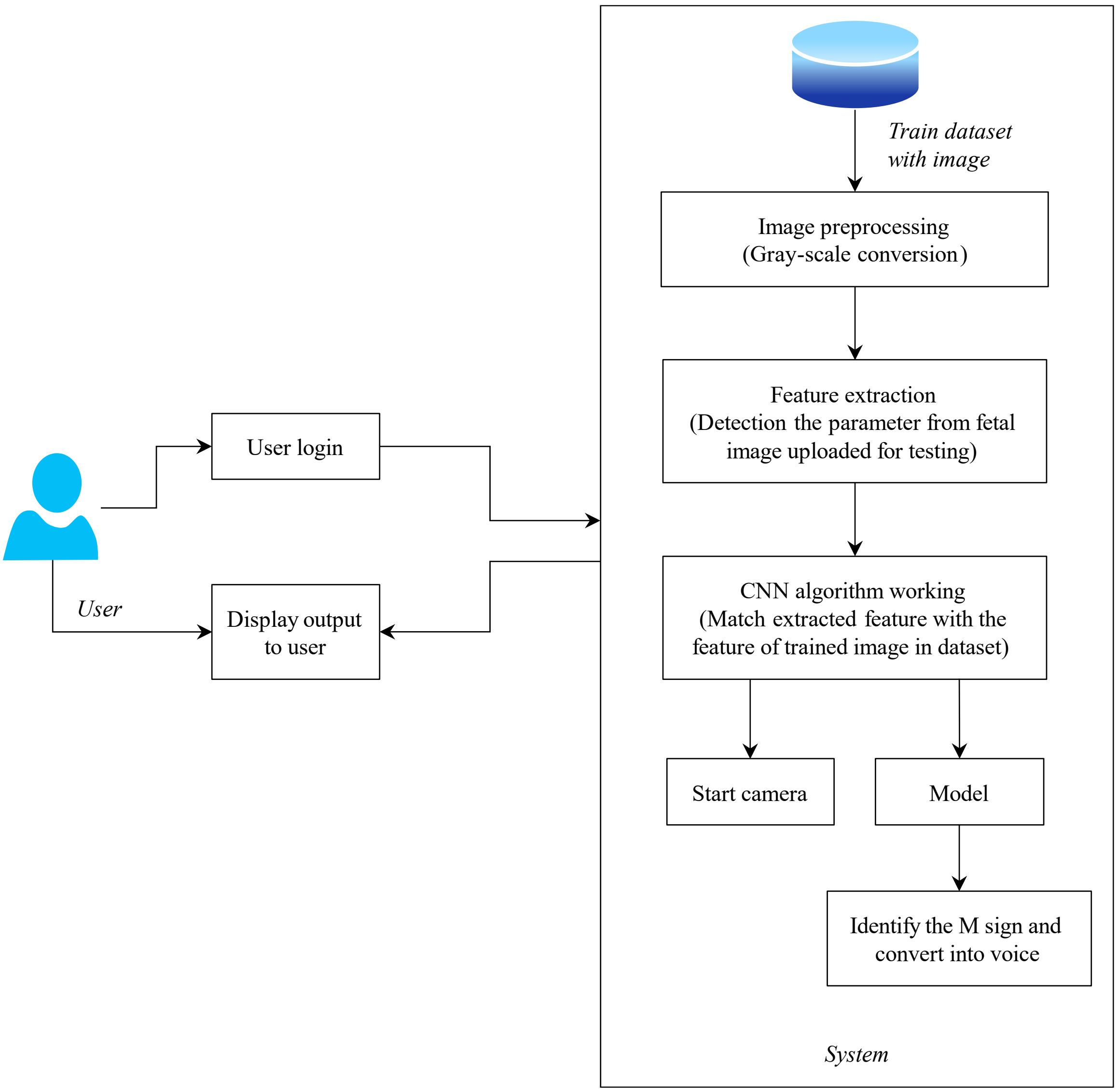

Figure 2. System architecture.

RESULT

Figures 3–6 show the interface of the application.

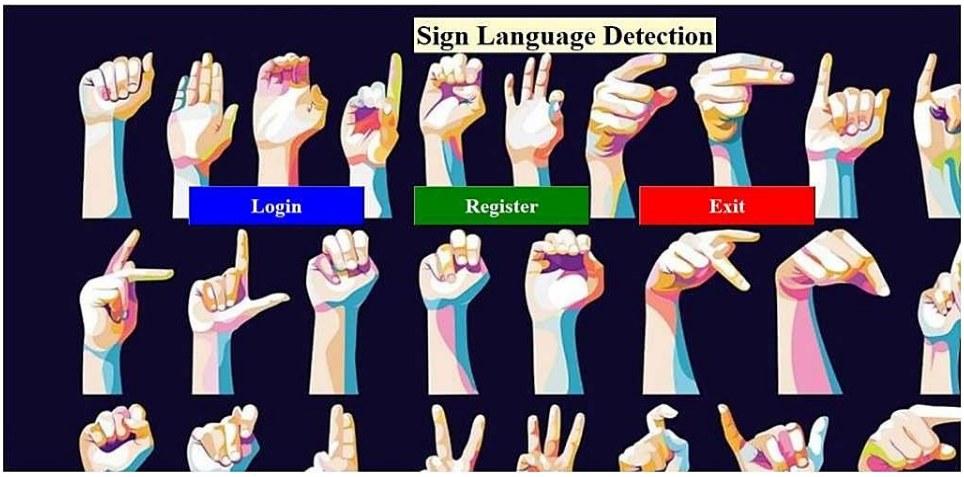

Figure 3. Sign language detection.

Figure 4. Login page.

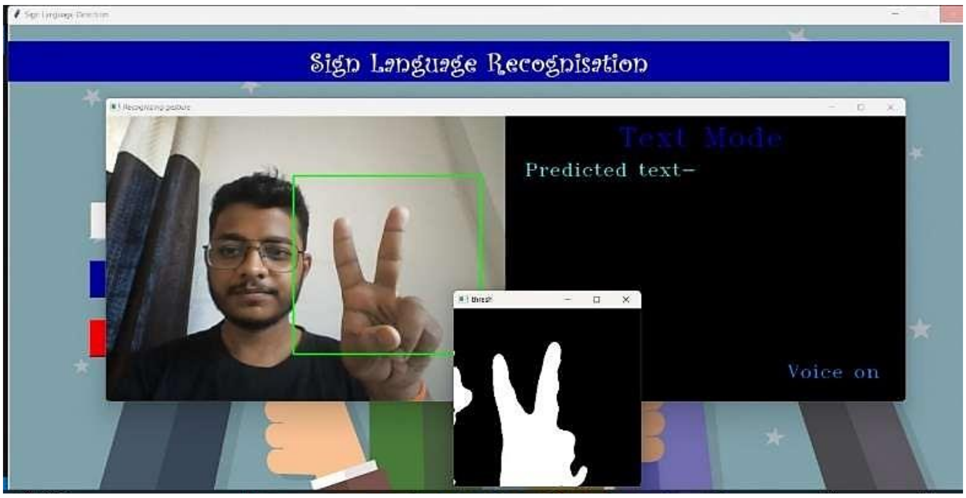

Figure 5. Sign language recognition.

Figure 6. After login interface.

CONCLUSION

This study explores a real-time hand gesture recognition system that translates sign language into both text and speech. It leverages deep learning and computer vision to detect hand gestures via a webcam. The captured gestures are processed using OpenCV and classified through a CNN model developed with TensorFlow and Keras. There is also a user-friendly interface made with Tkinter, which shows the recognized gestures in text and uses pyttsx3 to talk them out loud. This could help people with hearing and speech difficulties communicate more easily. The system uses SQLite3 for data management and Pickle to handle the model, which helps it run smoothly and store recognized gestures. The use of messagebox (ms) in the GUI adds helpful alerts and feedback for users. The results show decent accuracy in recognizing gestures, suggesting that deep learning can be useful for translating sign language.

In the future, it might be good to expand the gesture dataset for better accuracy, include more sign languages, or create a mobile or web version for easier access. This project shows that AI could help reduce communication barriers and support inclusivity for the deaf and mute community.

Acknowledgment

We would like to express our heartfelt gratitude to everyone who contributed to this research work. First, we are thankful for the guidance and support from our supervisor, whose expertise was invaluable throughout the research process. We also appreciate our research team for their collaboration and insightful feedback, which greatly enhanced the quality of this work. Thank you to the funding organization and all participants who made this study possible. Finally, we want to acknowledge our family and friends for their encouragement and support during this journey. Thank you all!

REFERENCES

1. Kothadiya DR, Bhatt CM, Saba T, Rehman A, Bahaj SA. Signformer: deep vision transformer for sign language recognition. IEEE Access. 2023 Jan 9; 11: 4730–9.
2. Surekha P, Vitta N, Duggirala P, Ambadipudi VS. Hand Gesture Recognition and voice, text conversion using CNN and ANN. In 2022 IEEE Second International Conference on Artificial Intelligence and Smart Energy (ICAIS). 2022 Feb 23; 167–171.
3. Al Abdullah B, Amoudi G, Alghamdi H. Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects. IEEE Access. 2024 Sep 10; 12: 128871–128895.
4. Jeevanandham P, Hariharan A, Keerthana G. Real-Time Hand Sign Language Translation: Text and Speech Conversion. In 2024 IEEE 7th International Conference on Circuit Power and Computing Technologies (ICCPCT). 2024 Aug 8; 1: 488–493.

5. Fareed AI, Ramanathan M, Yeswanth R, Devi S. Translation Tool for Alternative Communicators using Natural Language Processing. In 2024 IEEE 5th International Conference on Electronics and Sustainable Communication Systems (ICESC). 2024 Aug 7; 842–848.
6. Katoch S, Singh V, Tiwary US. Indian Sign Language recognition system using SURF with SVM and CNN. Array. 2022 Jul 1; 14: 100141.
7. Rathi P, Kuwar Gupta R, Agarwal S, Shukla A. Sign language recognition using resnet50 deep neural network architecture. In 5th International Conference on Next Generation Computing Technologies (NGCT-2019). 2020 Feb 27.
8. Puente A, Alvarado JM, Herrera V. Fingerspelling and sign language as alternative codes for reading and writing words for Chilean deaf signers. Am Ann Deaf. 2006; 151(3): 299–310.
9. Chansri C, Srinonchat J. Hand gesture recognition for Thai sign language in complex background using fusion of depth and color video. Procedia Comput Sci. 2016 Jan 1; 86: 257–60.
10. Varsha M, Nair CS. Indian sign language gesture recognition using deep convolutional neural network. In 2021 IEEE 8th International Conference on Smart Computing and Communications (ICSCC). 2021 Jul 1; 193–197.

