

# Sign Language Interpreter and Conversion to Text

Shraddha Anil Sarode<sup>1</sup>, Swarali Prakash Zore<sup>2</sup>, Akshada Rajendra Bhosale<sup>3</sup>

<sup>1,2,3</sup>Department of Computer Science and Engineering (Artificial Intelligence and Data Science) Padmabhooshan Vasantodada Patil Institute of Technology, Maharashtra, India

Guide: Prof. Mrs. M. R. Khare

\*\*\*

**Abstract** - This paper introduces a real-time Sign language interpretation system that converts hand gestures into text to support communication between individuals with hearing or speech impairments and healthcare professionals. Using computer vision and deep learning techniques, the system captures gestures from live video input and translates them into meaningful text instantly. This approach helps minimize communication barriers in clinical settings, enabling clearer interaction and improved patient care. The proposed system contributes to enhancing accessibility and promoting more inclusive healthcare services.

**Keywords**- Sign Language Recognition, Healthcare Communication, Gesture Recognition, Sign-to-Text Conversion, Text-to-Sign Conversion, Real-Time Communication, Real-Time Communication.

## 1. INTRODUCTION

Communication plays a vital role in healthcare, as it directly affects diagnosis, treatment, and overall patient care. However, for individuals with hearing or speech impairments, communicating with doctors can be difficult and sometimes stressful. Since most healthcare professionals are not familiar with sign language, this often creates a gap that can lead to misunderstandings and delays in treatment.

With advancements in technology, especially in computer vision and deep learning, it is now possible to develop systems that can understand and interpret human gestures. Sign language recognition systems aim to reduce this communication barrier by converting hand gestures into text, making interactions smoother and more effective.

This paper presents a real-time Sign Language Interpretation and Conversion to Text system designed to assist in healthcare settings. The system captures hand gestures using a camera, processes them using deep learning models, and converts them into readable text instantly. By doing so, it helps improve communication, enhances accessibility, and supports a more inclusive healthcare environment.

## 2. PROBLEM STATEMENT

In hospitals, clear communication between doctors and patients is very important. But for patients who use sign language, expressing their symptoms, pain, or concerns can be difficult when doctors do not understand sign language. At the same time, doctors may find it challenging to explain diagnoses, prescriptions, or instructions in a way the patient can understand.

Because of this gap, important information can be misunderstood, which may affect treatment and patient comfort.

This project focuses on creating a two-way communication system using sign language. The system will understand a patient's hand gestures and convert them into text for the doctor. It will also allow doctors to type messages, which will then be shown as sign language gestures for the patient.

The aim is to make communication in healthcare simpler, faster, and more accessible, so that every patient can clearly express themselves and understand their doctor without barriers.

## 3. OBJECTIVES

The main objective of this project is to create a smart and reliable two-way communication system that helps doctors and patients communicate easily using sign language. The system is developed to reduce communication difficulties faced by people with hearing or speech impairments during medical consultations and treatment.

The specific objectives of the proposed system are as follows:

- To identify and understand sign language gestures made by patients in real time.
- To convert patient gestures into readable text so that doctors can understand the message clearly.
- To transform the doctor's typed text into sign language gestures for patients.
- To reduce misunderstandings and improve communication in hospitals and clinics.

- To provide a simple and user-friendly system that can be used easily in healthcare environments.
- To make healthcare communication more inclusive, comfortable, and accessible with the help of modern technology

#### 4. LITERATURE SURVEY

Sign language recognition has become an important research area to improve communication for hearing-impaired individuals. Earlier systems mainly used sensor-based devices such as data gloves, which provided good accuracy but were costly and less practical for everyday use. To overcome these limitations, researchers shifted toward vision-based approaches using cameras for gesture detection.

With the development of deep learning, especially Convolutional Neural Networks (CNNs), the accuracy of gesture recognition systems has improved significantly. Tools like Keras enable efficient model development, while platforms such as Teachable Machine simplify the training process with minimal coding.

Most existing systems focus on one-way communication, converting sign language into text. However, recent research highlights the need for bidirectional systems that also convert text into sign language using stored video representations.

Despite progress, challenges such as varying lighting conditions, background noise, and differences in hand gestures still affect system performance. The proposed work addresses these issues by developing a two-way communication system using deep learning techniques.

#### 5. PROPOSED SYSEM

The proposed system is designed to provide an easy and effective way to convert sign language into text in real-time. The main focus of this system is to help people who use sign language communicates smoothly with others without needing any special devices.

In this project, a camera is used to capture hand gestures, and these gestures are processed using a deep learning model. The model is built using Keras and is based on a Convolutional Neural Network (CNN), which is well-suited for image and gesture recognition tasks. To simplify the model training process, Teachable Machine is used to collect and train gesture data, making it easier to create an accurate and efficient model.

Once the gesture is captured, the system analyzes it using

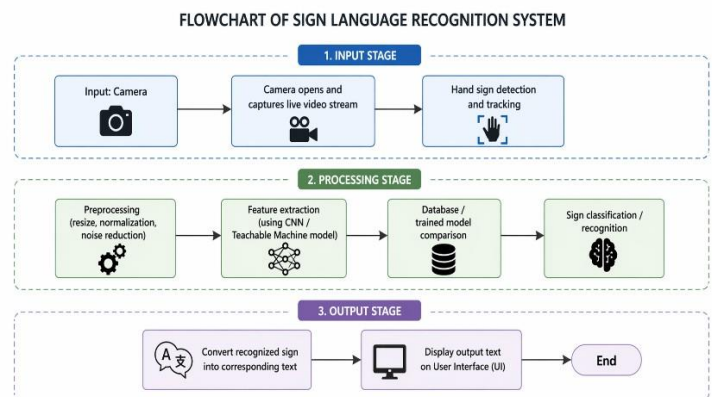
the trained CNN model and converts it into meaningful text instantly. This ensures real-time communication. In addition to this, the system also includes a feature that converts text into sign language using visual outputs like images or animations, enabling two-way communication.

The system is designed to be simple, cost-effective, and easy to use on devices like laptops or mobile phones, without requiring any extra hardware.

#### 6. IMPLEMENTATION

##### WORKING OF EXISTING SYSTEM

1. **Input Capture:** The system uses a camera to capture hand gestures in real time and converts the live video into frames for processing.
2. **Preprocessing:** The captured images are adjusted by resizing and cleaning to make them clearer and suitable for the model.
3. **Model Training:** (Teachable Machine + CNN) Gesture samples are trained using Teachable Machine, and the model is applied using a CNN built with Keras.
4. **Gesture Recognition:** The system analyzes each frame and identifies the hand gesture based on patterns learned during training.
5. **Conversion to Text:** Once the gesture is recognized, it is converted into meaningful text for the user.
6. **Text to Sign (Additional Feature):** The system can also take text input and convert it into sign language by showing related sign videos.
7. **Output Display:** The final result is displayed on the screen, either as text or as sign visual.



**Fig 1. Working Of Proposed System**

**WORKING OF REVERSE SYSTEM**

- I. *Text Input: The user enters text through a keyboard or graphical interface, which acts as the starting point of the process.*
- II. *Text Processing: The input text is cleaned by removing unnecessary characters and then divided into words or alphabets for easier conversion.*
- III. *Mapping with Sign Data: Each processed word or character is matched with its corresponding sign stored in the system's database.*
- IV. *Sign Sequence Generation: The system retrieves related sign visuals such as images or pre-recorded videos for each mapped word.*
- V. *Sequence Arrangement: The retrieved signs are arranged in the correct order to maintain the meaning of the original sentence.*
- VI. *Output Display: The final sequence of signs is displayed on the screen, forming a clear visual representation of the input text.*

REVERSE FLOW: TEXT TO SIGN CONVERSION SYSTEM

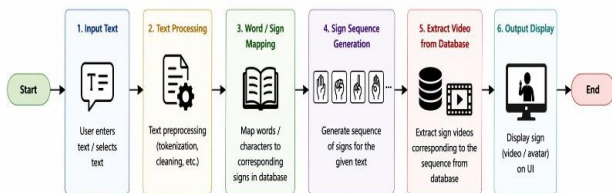


Fig 2. Working of Reverse Process

**7. TECHNOLOGIES USED**

**1. Convolution Neural Network:**

The main objective of using a Convolutional Neural Network (CNN) in this system is to accurately recognize hand gestures by extracting meaningful features from input images. Unlike traditional methods, CNN automatically learns patterns directly from the data, making it highly suitable for image-based classification tasks such as sign language recognition.

CNN is a deep learning model that consists of multiple layers designed to process visual information. The architecture used in this system includes convolutional layers, pooling layers, and fully connected layers, which work together to perform feature extraction and classification.

**Architecture of CNN**

1. *Convolutional Layers: The convolutional layers are responsible for extracting features from the input image. These layers apply small filters that slide over the image to detect patterns such as edges, curves, and hand shapes. As the layers deepen, more complex features are learned.*
2. *Activation Function (ReLU): After each convolution operation, the ReLU (Rectified Linear Unit) activation function is applied. It introduces non linearity into the model, allowing it to learn complex patterns efficiently and improving computational performance.*
3. *Pooling Layers: Pooling layers are used to reduce the spatial dimensions of the feature maps. This helps in decreasing computational complexity while retaining important features. Max-pooling is commonly used to select the most dominant features.*
4. *Fully Connected Layers: The extracted features are flattened and passed to fully connected layers. These layers combine all learned features to perform classification and determine the most relevant gesture class.*
5. *Output Layer (SoftMax): The final layer uses the SoftMax activation function to produce probability values for each gesture class. The class with the highest probability is selected as the predicted output.*

**CONVOLUTIONAL NEURAL NETWORK (CNN) ARCHITECTURE**

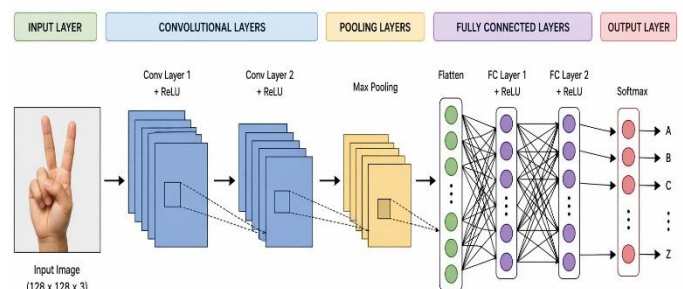


Fig 3. Convolution Neural Network(CNN) Architecture

### 1. Keras

Keras is used to build and manage the deep learning model in a simple and organized way. In this system, it helps in creating the CNN architecture by adding layers such as convolution, pooling, and dense layers step by step. During training, Keras adjusts the model's internal parameters to improve accuracy by learning from the dataset. It also provides functions to test and evaluate the model performance. Once the training is complete, the model is saved and later used to predict gestures in real time.

### 2. Teachable Machine

Teachable Machine is used to make the training process easier and faster. It allows users to collect gesture data directly using a camera and label each gesture class. The tool then automatically trains a model based on the provided data without requiring complex coding. After training, the model can be exported and integrated into the system. This makes it very useful for quickly building a custom gesture recognition model.

### 3. Python

Python acts as the backbone of the entire system. It is used to connect all components such as camera input, model prediction, and output display. With the help of different libraries, Python handles tasks like capturing video frames, preprocessing images, and passing them to the trained model. It also manages the system flow and ensure that both Gesture-to-text and text-to-sign processes run smoothly in real time.

## 8. RESULT AND DISCUSSION

The proposed sign language recognition system was tested using real-time hand gestures captured through a camera. The model successfully recognized most of the trained gestures and converted them into corresponding text with good accuracy. The use of CNN improved feature extraction, resulting in faster and more reliable predictions.

The reverse module (text-to-sign) also performed effectively by converting input text into appropriate sign language visuals using pre-recorded videos. The output was displayed clearly on the user interface, making it easy for users to understand. The system showed consistent performance under normal lighting conditions and with properly trained gesture datasets.

The results indicate that deep learning techniques can be effectively used for real-time sign language recognition. The integration of tools like Keras and Teachable

Machine simplified model development and improved accuracy.

However, some limitations were observed. The system performance may decrease under poor lighting conditions or when gestures are not clearly visible. Variations in hand size, orientation, and background can also affect prediction accuracy. Additionally, the system is limited to the gestures included in the trained dataset.

Despite these challenges, the proposed system provides a practical and efficient solution for bidirectional communication. With further improvements such as larger datasets and better preprocessing techniques, the system can achieve higher accuracy and robustness.

## 9. CONCLUSION

Good Communication between doctors and patients is important for proper treatment and care. However, patients who depend on sign language often experience difficulties while communicating in hospitals and clinics. This can create confusion, misunderstandings, and discomfort during medical consultations.

The proposed two-way sign language communication system is designed to make these interactions easier and more effective. The system helps doctors understand patient gestures by converting them into text, while also helping patients understand doctors through sign language gesture display from typed text.

By using technology to support communication in both directions, the project aims to create a more inclusive and accessible healthcare environment. It not only improves understanding between doctors and patients but also helps patients feel more confident, comfortable, and connected during their treatment process.

In the future, the system can be improved with better gesture recognition, support for different sign languages, and real-time voice features to make healthcare communication even more efficient and user-friendly.

## 10. REFERENCES

- Yang, D.; Martinez, C.; Visuña, L.; Khandhar, H.; Bhatt, C.; Carretero, J. Detection and Analysis of COVID-19 in medical images using deep learning techniques. *Sci. Rep.* 2021, 11, 19638. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
- Ministry of Statistics & Programme Implementation. Available online:

<https://pib.gov.in/PressReleasePage.aspx?PRID=1593253> (accessed on 5 January 2022).

- Camgoz, N.C.; Hadfield, S.; Koller, O.; Ney, H.; Bowden, R. *Neural Sign Language Translation*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018, Salt Lake City, UT, USA, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018*. [Google Scholar]
- JS. Arooj, S. Altaf, S. Ahmad, H. Mahmoud, A.S.N. Moha med. *Enhancing sign language recognition using CNN and sift: A case study on Pakistan sign language*. *J. King Saud Univ. Comput. Inf. Sci.*, 36 (2) (2024), Article 101934, [10.1016/j.jksuci.2024.101934](https://doi.org/10.1016/j.jksuci.2024.101934)
- R. Pigou et al., "Sign Language Recognition Using Convolutional Neural Networks," *European Conference on Computer Vision Workshops*, 2015.
- TensorFlow Documentation, <https://www.tensorflow.org/>, accessed June 2026.