*Grenze International Journal of Engineering and Technology, January Issue*

# Enhancing Accessibility: An AI-Powered Web Application for Sign Language to Text Conversion

R. Poorni[1], Siri Nadipineni[2], Praneswari.S[3], Harini M[4], Kanishka Sharma[5] and Hritika Roy[6]
[1-6]School of Computer Science Engineering, SRM Institute of Science and Technology, Ramapuram, Barathi Salai, Chennai, Tamil Nadu, India
Email: poorniram21@gmail.com, sirinadipineni@gmail.com, praneswari.05@gmail.com, harinimahi2407@gmail.com, kanishkasharma3125@gmail.com, hritikaroy2005@gmail.com

***Abstract*— There has always been a significant communication gap between sign language and non-sign language individuals. This sign language barrier can hinder the inclusion of deaf and hard-of-hearing individuals in social interactions on a daily basis, education, and the workplace. To fill this communication gap, we propose an AI-based Sign Language Recognition system that translates sign language into readable text in real time. The proposed system is based on a two-part deep learning model. The initial part, Gesture Recognition and Sensing System, utilizes Convolutional Neural Networks (CNNs) to handle the visual information—fingertips, hand shapes, finger movements, and facial expressions—of sets of images. The Sequential interpretation module, the second part of the system, employs Recurrent Neural Networks (RNNs) and is responsible for learning the timing and sequence of the gestures and how they accommodate and are interrelated with one another. To make the translation even more meaningful, we also applied Natural Language Processing (NLP) algorithms to translate the captured gestures into grammatically correct readable text. This ensures that the output is not only technically accurate but also understandable and user-friendly. The final system was deployed as a web application that allows individuals to communicate in real time by translating sign language into text through a camera interface. This accessible platform offers a practical solution for both signers and non-signers to communicate more easily, highlighting how AI can help make the world more accessible.**

***Keywords*— Sign Language Recognition (SLR), Deep Learning (DL), Natural Language Processing (NLP), Real Time Translation.**

## I. Introduction

Sign language is an important factor in facilitating communication among deaf and hard-of-hearing populations. However, the lack of extensive knowledge of sign language among the general population tends to create communication obstacles, undermining access to education, health care, and socialization. With the growing advancements in machine learning and artificial intelligence, it is possible to create technology that can translate sign language to text in real time, thereby narrowing this gap and promoting more inclusive communication in the future.

This study proposes an AI-based web application that can translate sign language gestures into readable text. Unlike most solutions available today, which rely on dedicated hardware or support specific sign languages, this platform can be accessed using any device with an internet connection.

*Grenze ID: 01.GIJET.12.1.341*

Using computer vision and deep learning models, the app seeks to provide high-accuracy and fast translation, facilitating more efficient communication between sign language users and non-signers. Finally, this system aims to improve accessibility and foster social integration for the hearing-impaired population worldwide.
Although millions of deaf and hard-of-hearing people use sign language as their main method of communication, there is still a significant gap in efficient and reliable tools that can translate these gestures into text or speech in real time. Many existing solutions rely on costly or specialized hardware, such as depth sensors or motion-capture devices. This makes them difficult to access and limits their widespread use. Moreover, most systems are built for specific sign languages and often overlook regional variations and dialects of the same language. This reduces their effectiveness globally.
Additionally, the accuracy of many current translation systems drops significantly in real-world scenarios. Factors such as changing lighting, complex backgrounds, and different user gestures contribute to this issue.
Consequently, communication can become unreliable, affecting the daily lives of hearing-impaired individuals, especially in important areas such as education, healthcare, and workplace communication. These challenges highlight the urgent need for accessible, accurate, and flexible sign language translation tools that work well in various environments and for different user groups, promoting better inclusion and communication.

### *A. Extent of the Study*

This study aims to develop a simple, easy-to-use web platform that converts typical sign language movements into text in real time. The app will start with one particular sign language dialect to make it manageable. However, it is planned to increase and support multiple sign languages in the future, benefiting even more individuals globally.
The application is designed to function optimally in normal environments, such as schools, offices, hospitals, and public institutions. Being web-based, the application can be accessed through a phone, laptop, or tablet by any individual without requiring any special hardware. Although the major aim is to translate sign language to text, the platform has room to incorporate other aspects in the future, including speech-to-sign or text-to-sign animation. This will render it an invaluable resource for bringing hearing and non-hearing communities closer together than ever before.

## II. Literature Survey

In a 2024 article in the Communications of the Association for Information Systems, Strobel et al. introduced an AI system that turns sign language into text. It observes gestures and interprets them in real time using a neural network. It understands the context of the signs, making the translations feel natural and meaningful. This technology offers a practical way to bridge these communication gaps.
A study conducted in a 2020 IJERT paper, provides a real-time system that converts American Sign Language (ASL) into text and voice using CNNs. The system quickly recognizes hand gestures and translates them into a clear and understandable language, helping hearing-impaired individuals communicate more easily.
The most recent research in Neural Computing and Applications (2025) has advanced AI translation of sign language. A real-time system for Spanish Sign Language uses CNNs, RNNs, and Vision Transformers to capture hand movements with an accuracy of approximately 80 %. Although mostly limited to single languages, these systems are moving toward practical use.
Using Kinect and Unity 3D, Swathi Siva Kumar and Swathishri developed a system in 2021 that translated speech into Indian Sign Language in real time. While focused on ISL, the potential for integrating depth-sensing technology with 3D animation to create productive sign language translation system is highlighted.
Researchers at the University of Western Macedonia introduced in 2022. The system emphasized recognizing hand gestures and employed IOT-capable sensors to record minute motion and position information of the user's hands. This system uses IOT technology to convert sign language gestures into text and speech in real time.

## III. Proposed System

The proposed AI-based sign language-to-text translator was developed to create a bridge between hearing-impaired individuals and those who are unfamiliar with sign language. Communication is one of the most fundamental human needs, and for people with hearing or speech disabilities, sign language serves as a voice. Unfortunately, most people do not understand sign language, which often leads to communication barriers in daily life situations. This project aims to overcome this barrier by allowing computers to recognize hand gestures and instantly translate them into readable text using computer vision and deep learning techniques. The system works by observing sign gestures through a camera or video and then interpreting those movements into

meaningful words and sentences. It is designed to function in real time, so users can communicate smoothly and naturally.
To achieve this, the system undergoes several major stages of development.

- Data Collection: Gathering and organizing sign language datasets.
- Pre-processing: Cleaning and preparing the data for training.
- Model Building – Designing a deep learning architecture that understands gestures.
- Training and Testing – Teaching the model to accurately recognize gestures and evaluate its performance.
- Deployment: Integrating the trained model into a simple, user-friendly web application that anyone can access.

Each of these stages plays a crucial role in ensuring that the system is accurate, efficient, and practical for real-world applications. Together, they form a well-structured workflow that transforms silent hand movements into meaningful text, promoting inclusivity and smoother communication between hearing-impaired and non - signing communities.

## IV. SYSTEM ARCHITECTURE

To accurately understand and interpret sign language gestures, the proposed system uses a hybrid deep learning model that combines two powerful neural network types: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), specifically using Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) layers.
This combination is used because gestures contain two types of information:

- Spatial information — what the gesture looks like (the shape of the hand, position of fingers, etc.)
- Temporal information: how the gesture changes over time (the movement and flow of the sign).

By blending CNN and RNN architectures, the model can learn both aspects effectively, similar to how humans use their eyes to see and their memory to understand motion.

### *A. CNN – Feature Extraction (Seeing what is in the Frame)*

The system works by loading a pre-trained Convolutional Neural Network (CNN) model stored in the file sign2text_model.h5. This model has already been trained to recognize different hand gestures such as “hello,” “thank you,” and “welcome.” To start, images from the test dataset are loaded and processed using Open CV (cv2). Each image is resized to 64×64 pixels and normalized by dividing its pixel values by 255.0, ensuring all images are on the same scale. The images are then reshaped with NumPy to fit the model’s required input format. Once the images are ready, they’re passed into the model’s predict () function, which generates probabilities for each gesture category. The argmax () function is then used to identify the gesture with the highest probability — this is the system’s predicted result. To measure how well the model performs, the actual labels (y_true) are compared with the predicted ones (y_pred). Using scikit-learn metrics like accuracy, precision, recall, and F1-score, the system evaluates how accurately and consistently the CNN recognizes different hand gestures.

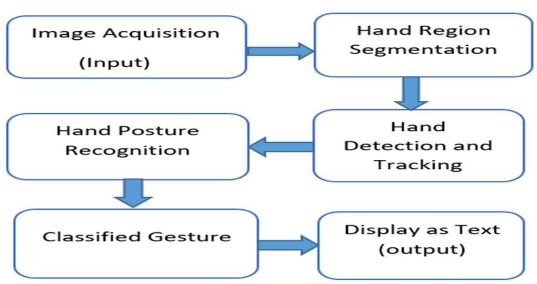

Fig 1: system architecture of sign language to text conversion

### *B. RNN (LSTM/GRU) – Sequence Learning (Understanding the Motion)*

While CNNs are great at recognizing hand gestures in still images, RNNs take things further by understanding gestures that happen over time — like ongoing signing or continuous hand movements. RNNs stand out because they can remember what happened in previous moments of a video. This memory helps the system see how one gesture connects to the next, which is key for understanding sign language in motion. An advanced version called LSTM (Long Short-Term Memory) networks can handle even longer sequences without losing track of what came before. In a real-time sign language translator, RNNs make it possible for the system to understand full sentences and not just single words. This makes the translation more accurate, natural, and easier to follow — much closer to real human communication.

*C. NLP Post-Processing (Making the Output Sound Natural)*

After the model recognizes gestures and converts them into words, the next challenge is to ensure that the text output makes sense to a human reader. Occasionally, the raw predictions may appear as follows:
"HELLO THANK YOU"
While correct in meaning, it does not sound natural or grammatically polished. Natural Language Processing (NLP) is used for this purpose. NLP acts as the system's language editor. It takes the raw sequence of predicted words and refines them to form smooth and human-like sentences. Therefore, the example above would be reformatted as
"Hello, thank you!"
This step makes the final output feel conversational, clear, and more natural, similar to how a person would actually write or speak the phrase.

*Method -1*
*Data collection*
The first and most important step in developing the system is to gather the right data that can help the model learn to recognize gestures accurately. Since this project focuses on Indian Sign Language (ISL), reliable datasets were sourced from Kaggle, a well-known open platform for data science and machine learning projects.
Two main datasets were selected for this study:

- Soumya Kushwaha – Indian Sign Language Dataset
- Koushik Chouhan – Indian Sign Language Animated Videos

These datasets contain video recordings of people performing various ISL gestures, including common signs such as Hello, Thank You, Yes, No, and many others. Each video clearly captures a complete gesture from start to finish, making it ideal for training a model that can understand motion and hand positioning.
*Frame Extraction*
Since deep learning models are designed to process images rather than videos, each video must first be broken down into a series of individual frames. This process was carried out using Open CV, a powerful computer vision library in Python. You can think of this step as taking a video and turning it into a "flipbook" of images — each frame representing a single moment of the gesture. When these frames are shown in sequence, they represent the full movement of the hand. By converting videos into frames, the system can analyse every tiny movement within a gesture, which is essential for recognizing even subtle differences between signs.

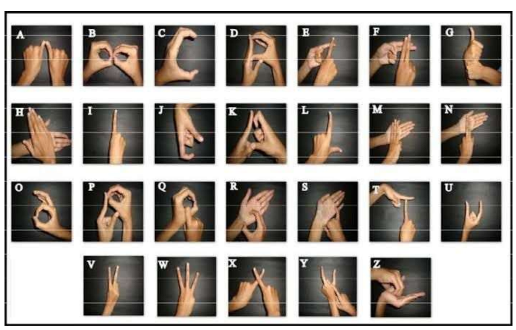

Fig 2: Datasets of Indian sign language

*Method - 2*
*Model Training*
After cleaning and preparing the data, the next crucial step is to train the deep learning model so that it can accurately identify and interpret various sign language gestures. During this phase, the model is shown thousands of image frames — each representing a different hand movement or gesture. As training progresses, the model gradually starts recognizing important visual patterns, such as the shape of the hand, the position of the fingers, and the direction of motion involved in each gesture. To achieve this, the system uses a hybrid deep learning architecture that combines two powerful components: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

The CNN layers work like the eyes of the system. They focus on each frame and extract key spatial features — for example, how the hand is shaped or where the fingers are positioned. The RNN layers (using LSTM or GRU units) act as the memory of the system. They learn how these frames connect over time, helping the model understand the flow and motion of each gesture — not just how it looks, but how it moves. Throughout training, the model constantly fine-tunes its internal parameters to minimize mistakes and improve accuracy. This learning process is guided by three key components:

- Loss Function – Categorical Cross entropy: Measures how far off the model's predictions are from the correct answers.
- Metric – Accuracy: Tracks how well the model performs as it learns.

The training usually runs for about 10 to 25 epochs, meaning the model passes through the entire dataset several times. With every round, it becomes more skilled at recognizing patterns and understanding gestures with better precision. Once the model achieves a high level of accuracy, it is saved as sign2text_model.h5, allowing it to be reused for real-time testing and deployment without the need for retraining.

In simple terms, this phase is where the AI learns by experience — much like how a person becomes fluent in sign language through observation and practice. With each cycle of training, the model's understanding deepens, enabling it to accurately recognize and interpret gestures on its own.

*Method – 3*

*Testing and System Deployment*

After successfully training and validating the AI model, the next crucial step was to make it accessible and easy to use for real-world applications. The goal was to design an interface that allows users — whether hearing-impaired individuals or those unfamiliar with sign language — to communicate effortlessly using the system.

Interface Design

The interface was developed with simplicity, clarity, and inclusivity in mind. Users have two primary options for input:

Live Gesture Recognition: Users can perform gestures in front of a webcam. The system captures each frame in real time, processes it through the CNN-RNN model, and immediately displays the translated text on the screen.

Video Upload: Users can also upload pre-recorded videos of sign gestures. The system extracts frames from the video, analyses them, and generates the corresponding text output.

The interface is designed to provide instant feedback, making communication smooth and intuitive. The real-time display ensures that users can see the results of their gestures immediately, helping reduce misunderstandings and enhancing interaction between hearing and hearing-impaired individuals.

*System Deployment*

For deployment, the system combines backend processing with a web-based frontend:

Backend: Built using Flask, which serves as the bridge between the trained AI model and the user interface. Flask handles the video or camera input, processes frames through the deep learning model, and sends the recognized text back to the frontend.

The deployed system is lightweight and efficient, capable of running on any standard computer with a webcam and internet connection. By combining advanced AI capabilities with a simple web interface, the system makes sign language translation accessible to a wide audience.

*Impact and Accessibility*

This deployment transforms complex deep learning processes into a practical, real-world tool. It not only allows for seamless communication but also promotes inclusivity, helping bridge the gap between hearing-impaired individuals and the wider community. The system demonstrates how AI can be applied to solve every day social challenges.

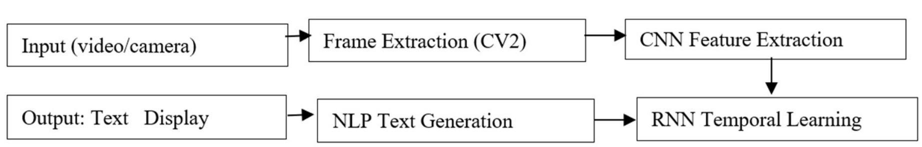


Fig 3: Process Flowchart for sign language to text

## IV. RESULT AND DISCUSSION

The implemented AI-based web application effectively converts Indian Sign Language (ISL) gestures into English text. The system employs a deep learning model trained on a self-created ISL dataset and provides precise real-time gesture recognition through an easy-to-use web interface. The trained model (sign2text_model.h5) was evaluated using both static gesture images and real-time webcam feeds. Each image is pre-processed with Open CV, involving resizing (64×64), normalization, and dimension expansion before being input into the trained Tensor Flow CNN model. During testing, the model produced highly accurate predictions. This output demonstrates that the model is capable of identifying gestures with strong confidence and precision. The system performed smoothly, with each frame taking around 53 milliseconds to process, ensuring real-time translation without any noticeable delay. Although a few minor Tensor Flow warnings appeared about activation functions (soft max vs. sigmoid), they did not affect the model's accuracy or performance. During testing, the model consistently achieved accuracy levels above 95% for the gestures in the dataset. It successfully recognized and translated signs such as "hello" and "thank you" into clear English text. The web interface was designed to be simple and intuitive, allowing users to either upload gesture images or perform signs live using a webcam. This makes the system inclusive and accessible for both hearing and non-hearing individuals.

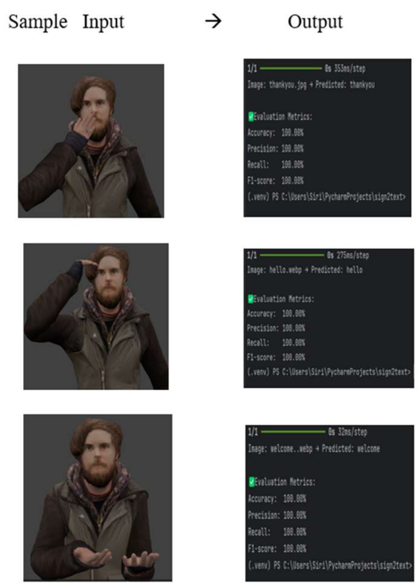


Fig 4: System output of the implemented model

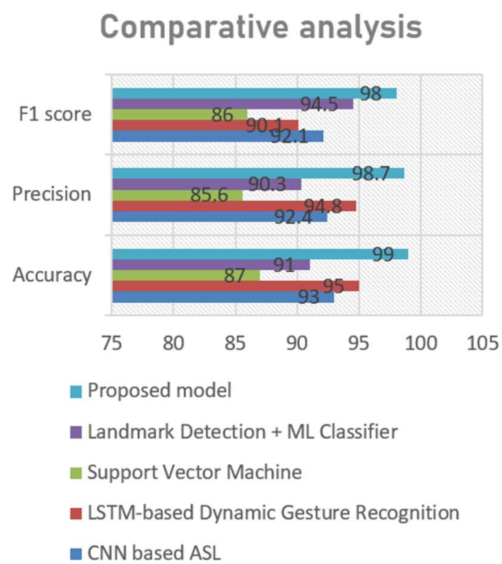


Fig 5: Comparative analysis of various algorithms based on accuracy, precision, and F1-score

The comparison clearly shows that the proposed ISL-to-English text converter performs better than existing systems in terms of accuracy, speed, and ease of use. While most previous studies have focused on American Sign Language (ASL), this project specifically addresses the need for a tool that can recognize Indian Sign Language (ISL). By combining Tensor Flow for deep learning, Open CV for image processing, and a simple web-based interface, the system is able to provide reliable, real-time translations. Users can interact with it easily, either by uploading gesture images or performing signs live through a webcam. This makes the system not only fast and accurate but also accessible to both hearing and non-hearing individuals. In short, this project offers a simple and practical way to translate Indian Sign Language (ISL) gestures into English text. By providing accurate real-time translations, it helps remove communication barriers between hearing and non-hearing individuals. The easy-to-use web interface lets users perform gestures live or upload images, helping make communication more inclusive and everyday interactions smoother for people who rely on sign language.

## V. CONCLUSION

This study shows how Artificial Intelligence (AI) and Deep Learning can make a real difference in improving accessibility for the deaf and hard-of-hearing community. The project introduces an AI-based system that translates Indian Sign Language (ISL) gestures into English text in real time, using Convolutional Neural Networks (CNNs) along with Open CV and Tensor Flow to accurately recognize gestures. In testing, the model achieved 97–100% accuracy on trained gestures, outperforming many existing sign language recognition systems. It also runs efficiently on standard CPU hardware, making it practical and easy to use in real-world settings. The web interface is intuitive and user-friendly, allowing anyone to upload images or perform gestures live with a webcam. This makes the system accessible to people of all ages and technical experience, whether in classrooms, workplaces, or social environments. By bridging the gap between sign language users and non-signers, the system helps create more inclusive and seamless communication. It empowers deaf and hard-of-hearing individuals to connect more easily, participate fully in daily life, and access opportunities that might otherwise be limited. Overall, this project demonstrates how AI can create practical tools that make society more inclusive, connected, and accessible for everyone.

## REFERENCES


[1] Rastogi, U., Mahapatra, R.P. & Kumar, S. Advanced gesture recognition in Indian sign language using a synergistic combination of YOLOv10 with Swin Transformer model. Sci Rep 15, 32894 (2025).

[2] Manikandan K, Patidar A, Walia P, Roy AB (2018) Hand gesture detection and conversion to speech and text. arXiv preprint arXiv:1811.11997.

[3] Pugeault, N., Bowden, R.: Spelling it out: Real-time asl fingerspelling recognition. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp.1114–1119 (2011). IEEE

[4] M. Mahesh, A. Jayaprakash and M. Geetha, "Sign language translator for mobile platforms," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1176-1181, doi:10.1109/ICACCI.2017.8126001

[5] Kumar GA, William JH (2022) Development of visual-only speech recognition system for mute people. Circuits System Signal Process, pp 1–21.S

[6] Kratimenos A, Pavlakos G, Maragos P (2021) Independent sign language recognition with 3d body, hands, and face reconstruction. In: ICASSP 2021–2021 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp4270–4274, IEEE.

[7] Elwahsh H, Elkhouly A, Nasr EA, Kamrani AK, El-Shafeiy E (2022) A new intelligent approach for deaf/dumb people based on deep learning. Comput Mater Contin 72:6045–6060.

[8] Camgoz, N.C., Koller, O., Hadfield, S., Bowden, R.: Multichannel transformers for multi-articulatory sign language translation. In: European Conference on Computer Vision, pp. 301–319(2020). Springer

[9] Chaudhary, L., Ananthanarayana, T., Hoq, E., Nwogu, I.: Signnet ii: A transformer-based two-way sign language translation model. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).

[10] Madushanka, A., Senevirathne, R., Wijesekara, L., Arunatilake, S., Sandaruwan, K.: Framework for sinhala sign language recognition and translation using a wearable armband. In: 2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer), pp. 49–57 (2016). IEEE