Real-Time Indian Sign Language Pose Recognition Using Deep Convolutional Neural Networks

Report submitted to GITAM (Deemed to be University) as a partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Electronics & Communication Engineering.



DEPARTMENT OF ELECTRICAL, ELECTRONICS AND COMMUNICATION ENGINEERING

GITAM SCHOOL OF TECHNOLOGY

GITAM (DEEMED TO BE UNIVERSITY)

BENGALURU -561203

NOV 2025

**K. Vignan Uttej-BU22EECE0100519**

**C.Dinesh Karthik-BU22EECE0100353**

**K.P. Balachandra - BU22EECE0100350**

**DECLARATION**

We declare that the project work contained in this report is original and it has been done by me under the guidance of my project guide.

**Name & Reg No:**

K. Vignan Uttej-BU22EECE0100519

C.Dinesh Karthik-BU22EECE0100353

K.P. Balachandra - BU22EECE0100350

**Department of Electrical, Electronics and Communication Engineering GITAM School of Technology, Bengaluru-561203**

****

**CERTIFICATE**

This is to certify that, K. Vignan Uttej, C. Dinesh Karthik & K.P. Balachandra has satisfactorily completed Mini Project Entitled in partial fulfillment of the requirements as prescribed by University for VIIth semester, Bachelor of Technology in “Electrical, Electronics and Communication Engineering” and submitted this report during the academic year 2025-2026.

[Signature of the Guide] [Signature of HOD]

**Table of contents**

**Chapter 1: Introduction**

* 1. Overview of the problem statement
  2. Objectives and goals

**Chapter 2 : Literature Review**

**Chapter 3 : Strategic Analysis and Problem Definition**

* 1. SWOT Analysis
  2. Project Plan - GANTT Chart
  3. Refinement of problem statement

**Chapter 4 : Methodology**

* 1. Description of the approach
  2. Tools and techniques utilized
  3. Design considerations

**Chapter 5: Conclusion**

**Chapter 6 : Future Work**

**Chapter 1: Introduction**

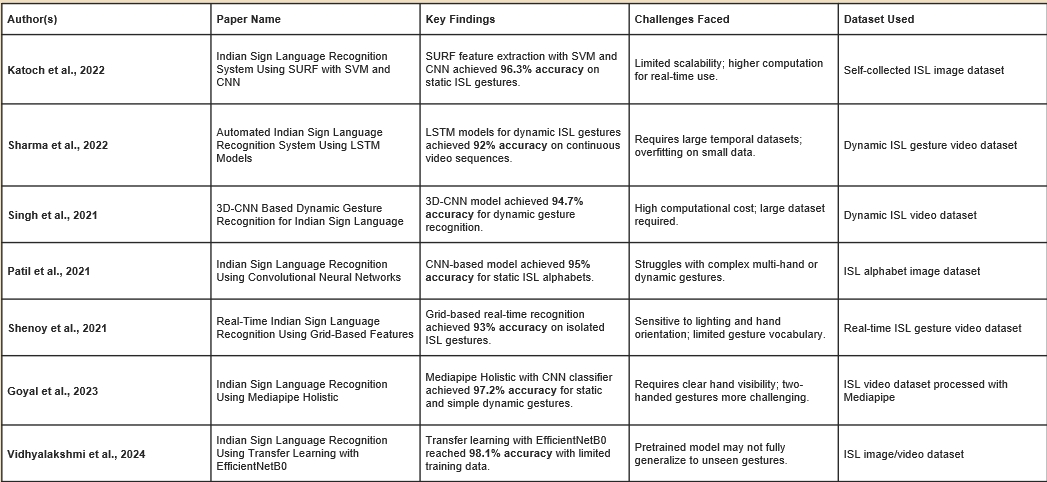
* 1. Overview of the problem statement

Communication between hearing-impaired individuals and the general population remains a significant challenge due to the absence of a shared language. Indian Sign Language (ISL) is a crucial medium of interaction for the deaf and hard-of-hearing community, yet most people are not trained to understand it. This knowledge gap often leads to social exclusion and communication barriers in everyday situations such as education, healthcare, and public services.

With the rapid advancements in deep learning and web technologies, there is now an opportunity to create a real-time, web-based system that can translate ISL gestures into text or speech, enabling instant understanding and promoting inclusivity and accessibility for everyone.

* 1. Objectives and goals
* To design a real-time ISL translation system using a web-based platform that captures hand gestures directly through a webcam.
* To train a Deep Convolutional Neural Network (CNN) for accurate recognition of Indian Sign Language poses and movements.
* To ensure robustness of the system under varying lighting conditions, different backgrounds, and diverse hand orientations.
* To convert recognized signs into textual or auditory output for clear and instant communication.
* To promote inclusivity and accessibility by providing an assistive technology that bridges the communication gap for the hearing-impaired community.

**Chapter 2 : Literature Review**



The literature survey highlights a steady progression in Indian Sign Language (ISL) recognition research, with a clear shift from traditional machine-learning methods toward deep learning models for higher accuracy and real-time performance. Earlier studies achieved good results but were often limited by static gesture recognition, small vocabularies, or difficulty handling dynamic movements and environmental variations. Recent papers demonstrate that advanced architectures such as CNNs, 3D-CNNs, LSTMs, Media pipe tracking, and transfer learning significantly improve recognition of both static and dynamic signs.

Our project, “Real-Time Indian Sign Language Pose Recognition Using Deep Convolutional Neural Networks,” aligns with these advancements by adopting a deep CNN approach deployed on a web platform. This strategy addresses previous limitations by maintaining high accuracy even with lighting changes, background clutter, and varied hand orientations, while also enabling instant predictions directly in a browser. By processing live video feeds and providing text or speech output, the system extends the capabilities of earlier work and lays the foundation for future developments such as sentence-level translation, larger gesture vocabularies, and wider public accessibility through scalable web deployment.

**Chapter 3 : Strategic Analysis and Problem Definition**

* 1. SWOT Analysis

Strengths

* Real-time processing with low latency for instant gesture recognition.
* Web-based deployment enables access from any browser without installations.
* Deep CNN model ensures high accuracy and robustness to lighting and background changes.
* Inclusive solution that bridges communication between hearing-impaired individuals and the general public.

Weaknesses

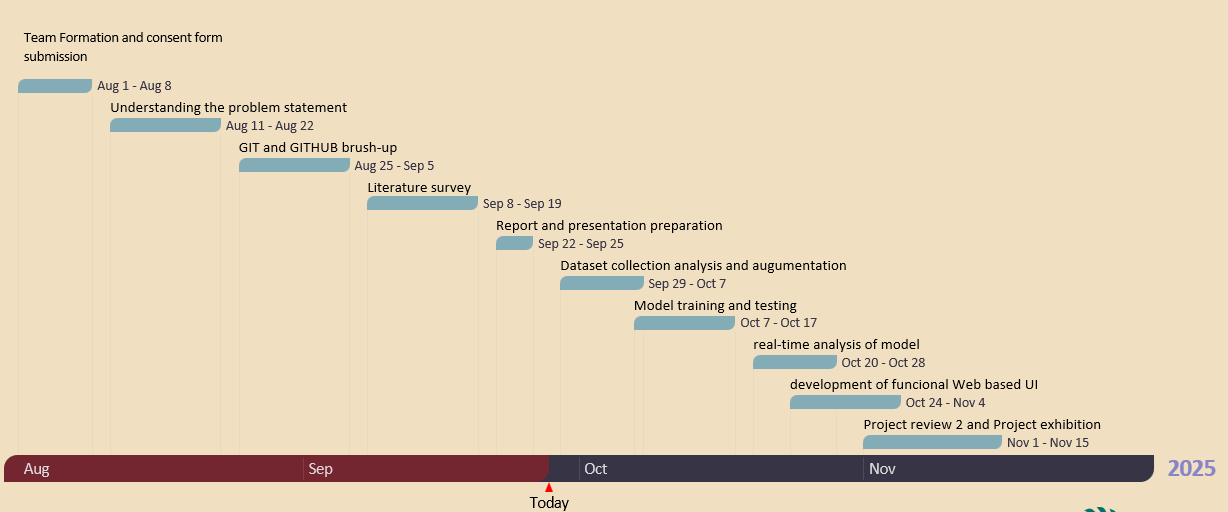
* Dataset dependency – performance relies heavily on the quality and diversity of training images
* High computational requirements during training and for continuous live inference.
* Limited vocabulary – currently focuses on single gestures (alphabets/numbers) rather than full sentence structures.

Opportunities

* Expansion to sentence-level recognition and continuous sign translation.
* Integration with assistive technologies or public service platforms (schools, hospitals, government services).
* Potential adoption by NGOs, educational institutions, and accessibility programs to promote inclusivity.

Threats

* Variability in user gestures such as different signing speeds, hand sizes, or camera angles may affect accuracy
* Network or browser limitations could impact real-time performance on low-end devices.
* Emerging competing solutions using wearables or sensor-based approaches may challenge adoption.
  1. Project Plan - GANTT Chart



* 1. Problem statement

To develop a real-time, web-based Indian Sign Language (ISL) recognition system using deep learning that accurately translates hand gestures into text or speech, while remaining robust to diverse lighting, background, and hand orientations, thereby bridging the communication gap for the hearing-impaired community.

**Chapter 4 : Methodology**

* 1. Description of the approach
* Capture live hand gestures using a webcam through a browser-based interface.
* Preprocess input frames by resizing, normalizing, and applying background filtering to extract clear hand regions
* Train a Deep Convolutional Neural Network (CNN) on the Indian Sign Language dataset to recognize alphabet and numeral gestures.
* Deploy the trained model on a web platform for real-time inference accessible through any browser.
* Convert predictions into text or speech output to enable instant communication.

4.2 Tools and Techniques Utilized

Programming Language: Python

Deep Learning Framework: TensorFlow / PyTorch

Model Type: Convolutional Neural Network (CNN)

Dataset: Indian Sign Language (ISL) dataset collected from Kaggle (42,000 images, 35 classes)

Platform: Web-based deployment using Flask/Django or similar framework

Tools: OpenCV for image preprocessing, HTML/CSS/JavaScript for the web interface

4.3 Design Considerations

* Accuracy vs. Latency: Optimize the model to maintain high recognition accuracy while ensuring low delay for real-time interaction.
* Scalability: Design the system so that new gestures or additional classes can be added in the future without major changes.
* Robustness: Ensure reliable recognition under different lighting conditions, varied backgrounds, and diverse hand orientations or skin tones.
* User-Friendliness: Provide a simple and intuitive web interface for smooth user interaction and easy accessibility across devices.
  1. Tools and techniques utilized
  + **Programming Language:** Python
  + **Deep Learning Frameworks:** TensorFlow / PyTorch
  + **Model Type:** Convolutional Neural Networks (CNN)
  + **Dataset:** Curated Indian Sign Language dataset (or self-collected data if required).
  + **Platform:** Android Smartphone with integrated camera
  + **Tools:** OpenCV for image processing, Android Studio for mobile deployment
  1. Design considerations
* Accuracy vs. Latency: Optimize the model to maintain high recognition accuracy while ensuring low delay for real-time interaction
* Scalability: Design the system so that new gestures or additional classes can be added in the future without major changes.
* Robustness: Ensure reliable recognition under different lighting conditions, varied backgrounds, and diverse hand orientations or skin tones.
* User-Friendliness: Provide a simple and intuitive web interface for smooth user interaction and easy accessibility across devices.

**Chapter 5: Conclusion**

Our project successfully developed and trained a deep learning model capable of recognizing Indian Sign Language (ISL) gestures using a web-based platform.

The trained Deep Convolutional Neural Network (CNN) demonstrated reliable recognition performance on the collected dataset, validating the core objective of creating a real-time ISL translator.

By leveraging deep learning and web technologies, we have established a strong proof-of-concept for bridging the communication gap between the deaf and hard-of-hearing community and mainstream society.

This initial success lays the groundwork for future phases, including real-time deployment, sentence-level recognition, and a fully scalable browser-accessible application that promotes inclusivity and accessibility.

**Chapter 6 : Future Work**

Since the dataset has been successfully collected, the next phase of this project will focus on building and deploying the recognition system.

Key areas of future work include:

* Model Development and Training – Design and train a Deep Convolutional Neural Network (CNN) using the collected Indian Sign Language dataset to accurately recognize gestures.
* Model Optimization – Fine-tune hyperparameters and preprocessing techniques to achieve high accuracy and low latency suitable for real-time performance.
* Real-Time Web Integration – Integrate the trained model into a browser-based application that captures live video from a webcam and delivers instant predictions.
* Output Conversion – Implement text and optional speech output for recognized signs to enable clear communication.
* Robustness and Scalability – Expand the dataset and test the system under different lighting conditions, backgrounds, and hand orientations, ensuring the model remains reliable and scalable for future enhancements such as sentence-level recognition.