REAL-TIME ISL TO TEXT TRANSLATOR USING SEQUENTIAL DEEP LEARNING ANS CNN-LSTM

Team Members:
Mayur R (4MH22CA024),
Kishan K M (4MH22CA019),
Nikhil M (4MH22CA028),
Manya B J (4MH22CA023)
Guide: Prof: Sahana H C
Project Mentor: Dr.Victor AI

Presentation Date: 26-05-2025

TEAM

Mayur R
Full Stack Developer

Kishan K M
Data Engineer

Nikhil M
Mobile App Developer

Manya B J
Research Analyst

GUIDE AND MENTOR

Dr.Victor AI
Project Guide

Dr.Hemanth S.R
Head of Department

Literature Survey on Indian Sign Language Research Papers

Paper	Model & Method	Accuracy	Strengths	Limitations
Sharma et al.	CNN Static ASL	92%	Robust preprocessing	No temporal modeling, ASL-focused
Iyer & Mehta	MobileNet V2 Transfer Learning	88%	Lightweight, efficient	Static only, limited data
Khan et al.	CNN-LSTM Dynamic ASL	Improved by 15%	Temporal dependencies modeled	High computational cost, ASL bias

RESEARCH PAPER-1 SIGN LANGUAGE RECOGNITION USING DEEP LEARNING BY SHARMA ET AL.

Problem Statement
Recognition of static American Sign Language (ASL) gestures using machine learning methods with high precision.

Objectives
Develop a robust CNN model to classify static ASL gestures with high accuracy and reliable preprocessing techniques.

Methodology
Applied convolutional neural networks (CNN) with intensive image preprocessing to classify static ASL signs.

Advantages & Limitations
1. Strong preprocessing enhanced model robustness.
2. Achieved 92% accuracy on static gestures.
3. Did not incorporate temporal modeling for dynamic gestures.
4. Focus limited to ASL, restricting multilingual generalization.

Proposed CNN-LSTM Architecture for Sequential Gesture Recognition by Research paper 1

Model Overview:

- Input: Sequence of gesture video frames.
- Conv3D Layers: Extract spatial and short-term temporal features.
- MaxPooling3D: Downsamples the data and reduces complexity.
- Batch Normalization: Speeds up convergence and stabilizes training.
- LSTM Layer: Models long-term temporal dependencies in gesture sequences.
- Dropout Layer: Prevents overfitting during training.
- Fully Connected + Softmax: Outputs gesture class probabilities.

Research Paper-2 Multilingual Sign Language Dataset & Model by Verma et al.

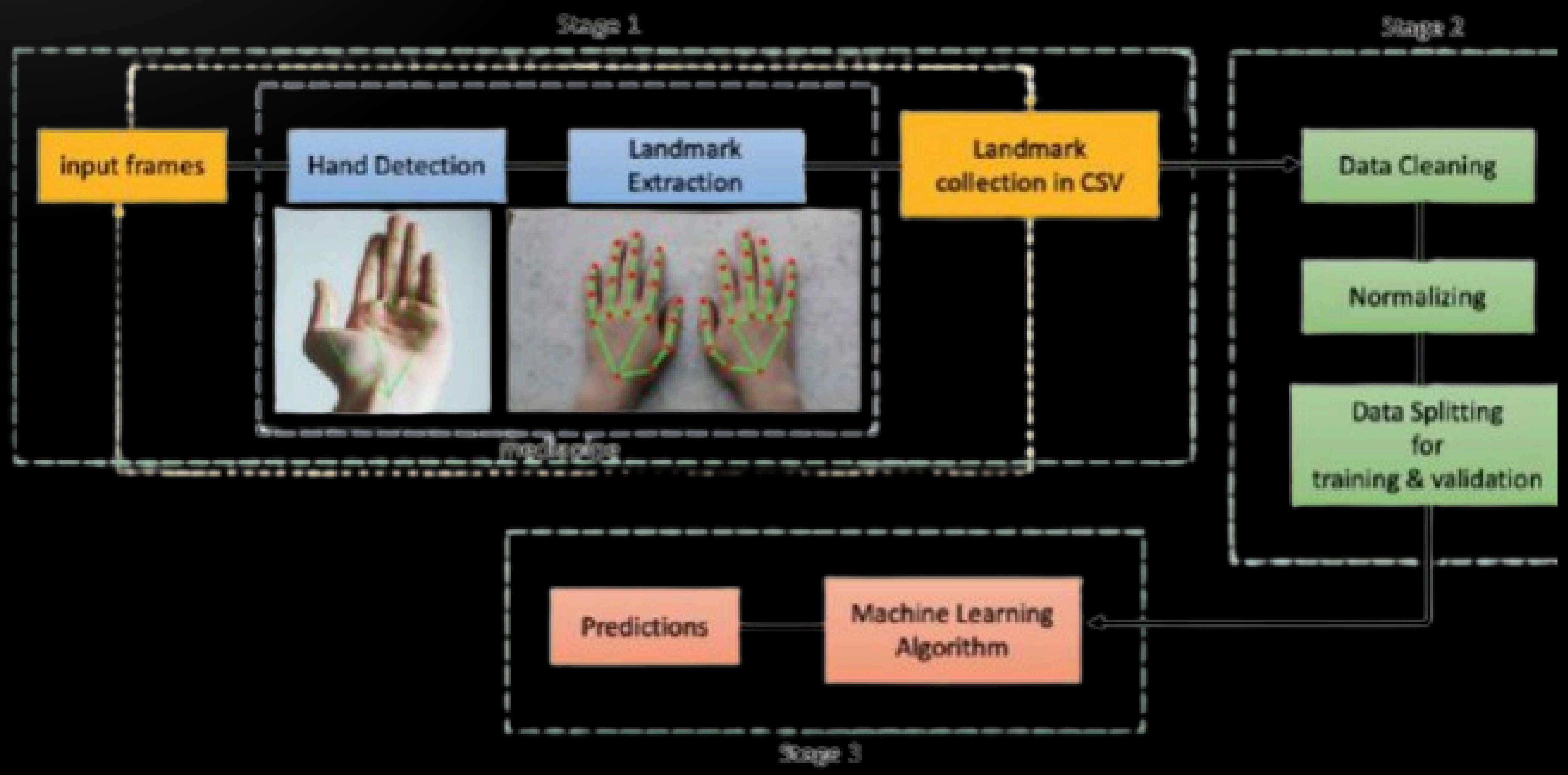
Problem Statement
Enabling sentence-level translation for Indian Sign Language and other languages using recurrent models.

Advantages & Limitations
Inclusion of multilingual data enhances applicability. Dataset remains small; model complexity challenges remain with RNN over simple RNN architectures.

Objectives
Create and utilize a multilingual dataset to support sentence-level gesture translation improving real-world application.

Methodology
Applied RNN-based architectures with preference for LSTM networks to model sequential dependencies.

Three-Stage Process for ISL Gesture Detection and Classification by Research Paper 2



- **Stage 1** – Data CollectionVideo frames are captured from the webcam.
- MediaPipe detects the hand in each frame and extracts 21 hand landmarks (like fingertip and knuckle positions). These x, y, z coordinates are then saved in a structured format as a CSV file for training.
- **Stage 2** – Data Preprocessing
- The collected CSV data is cleaned to remove errors or missing values. It is then normalized (scaled to a uniform range) for better model performance. The dataset is split into training and validation sets to prepare it for model learning.
- **Stage 3** – Model Training & Prediction
- A machine learning algorithm (like Random Forest, SVM, or LSTM) is trained on the preprocessed landmark data. Once trained, the model predicts hand gestures in real time or from test input, outputting the corresponding text.

RESEARCH PAPER-3 CNN-LSTM FOR SEQUENTIAL GESTURE RECOGNITION BY KHAN ET AL.

Problem Statement

Recognition of dynamic ASL gestures capturing temporal dependencies using advanced deep learning.

Methodology

Integrated CNN for spatial feature extraction with LSTM for temporal sequence modeling

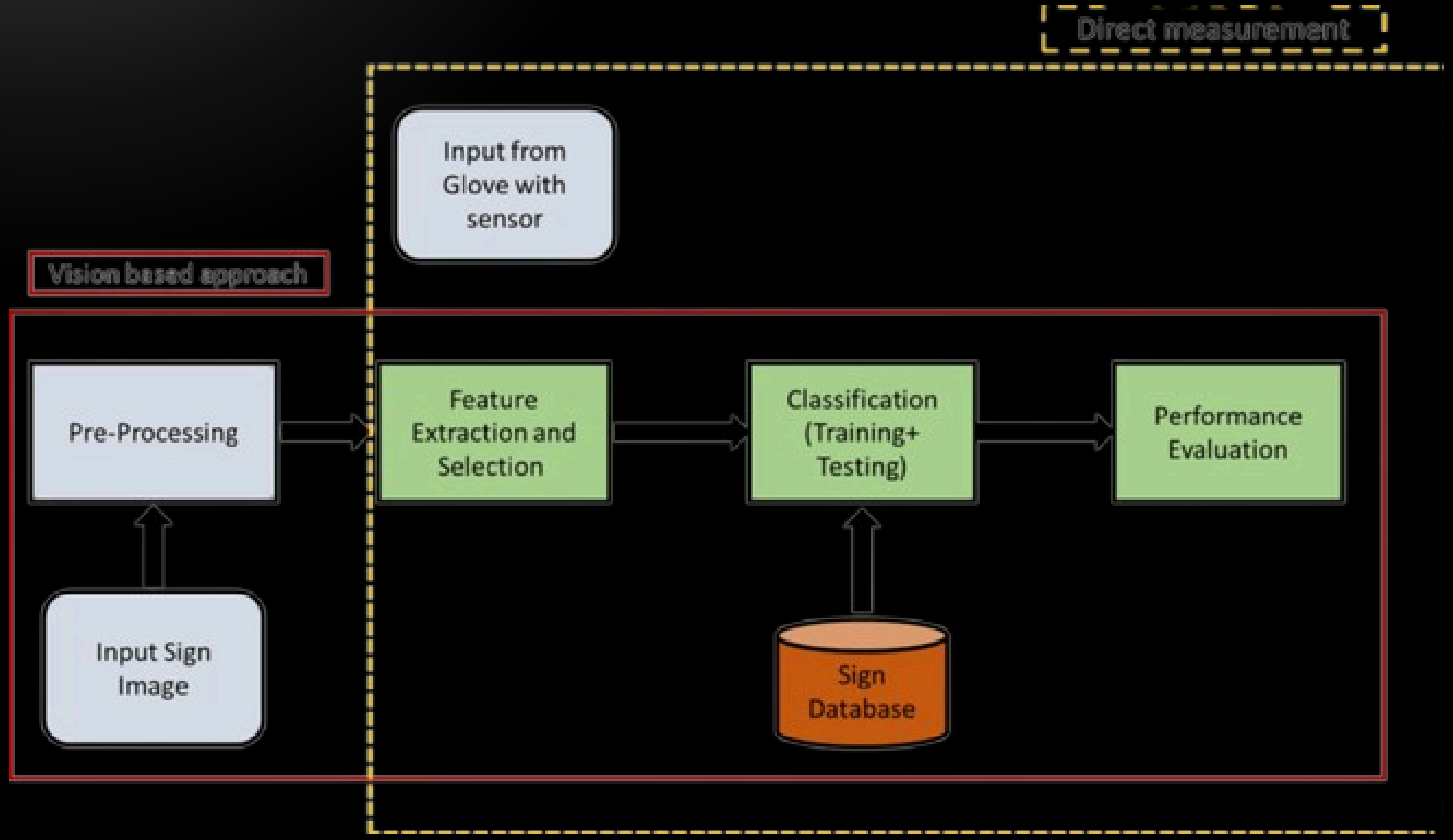
Objectives

Enhance accuracy in sequential gesture recognition by combining convolutional and recurrent networks.

Advantages & Limitations

Improved accuracy by 15% over CNN-only models. High computational cost and ASL-specific focus limit practical deployment and multilingual application.

System Workflow for Sign Language Recognition Using Image and Sensor Inputs



- Input Methods: There are two approaches shown —
- Vision-based approach (using images of hand signs)
- Direct measurement (using gloves with sensors)
- Pre-Processing: For vision-based input, sign images are pre-processed to enhance quality and remove noise.
- Feature Extraction and Selection: Key features are extracted from the input (e.g., hand shape, orientation) for accurate classification.
- Classification: The features are used to train and test a machine learning model to recognize specific signs.
- Sign Database: A database stores known signs and their features, helping the model learn and compare input signs.
- Performance Evaluation: Finally, the system's accuracy and reliability are assessed using test data.

Challenges in Sign Language Recognition Research

1. **Dataset Limitations**
   Small datasets and language-specific collections restrict generalizability and robustness.

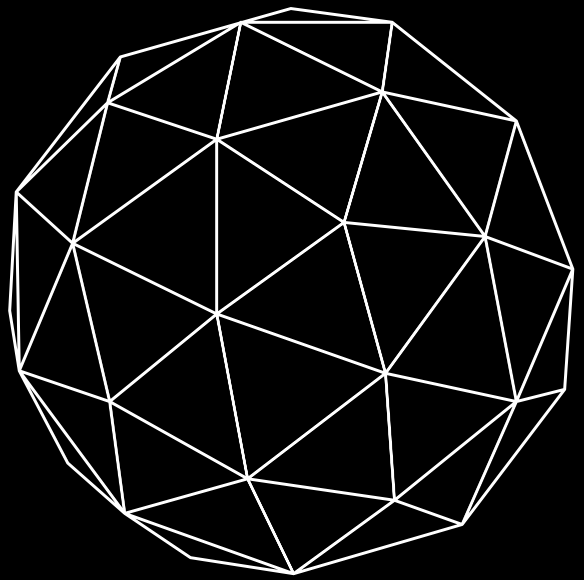
2. **Temporal Modeling Complexity**
   Capturing dynamic gestures requires computationally intensive recurrent models, raising deployment barriers.

3. **Multilingual Support**
   Most models focus on single languages, limiting accessibility for diverse sign language users.

4. **Computational Efficiency**
   Balancing model accuracy with resource efficiency remains a central challenge for real-time applications.

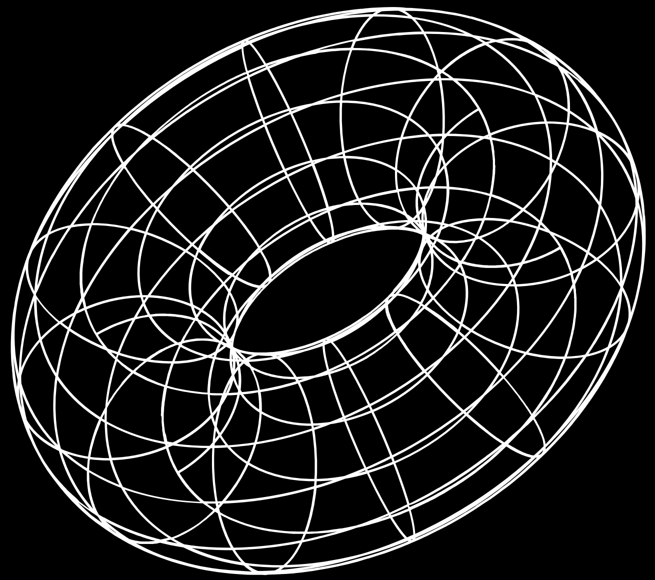
FEATURE 1

REAL-TIME ISL TO SPEECH



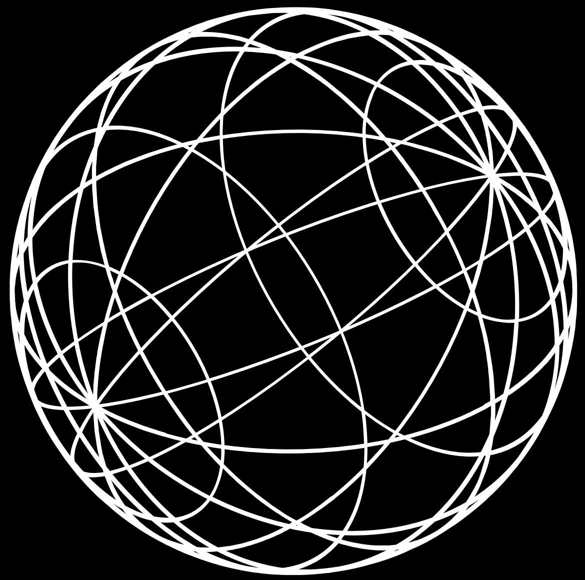
Real-Time Gesture Recognition

The system uses a webcam to continuously capture Indian Sign Language (ISL) hand gestures.



Gesture Prediction

The data is processed through a CNN-LSTM model.

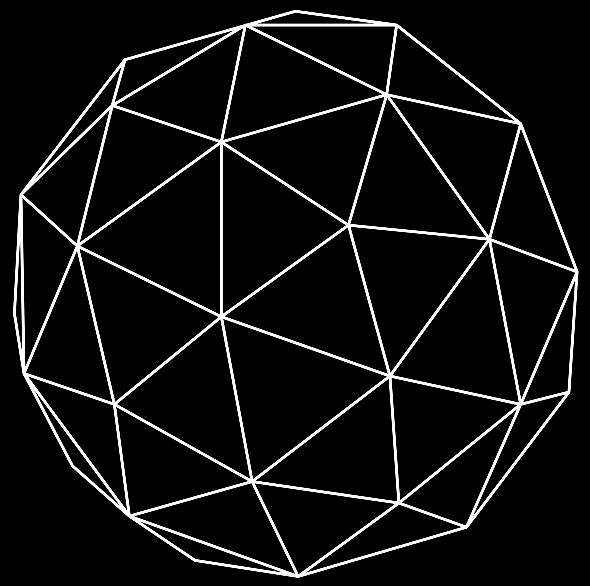


Predicted Text to Speech

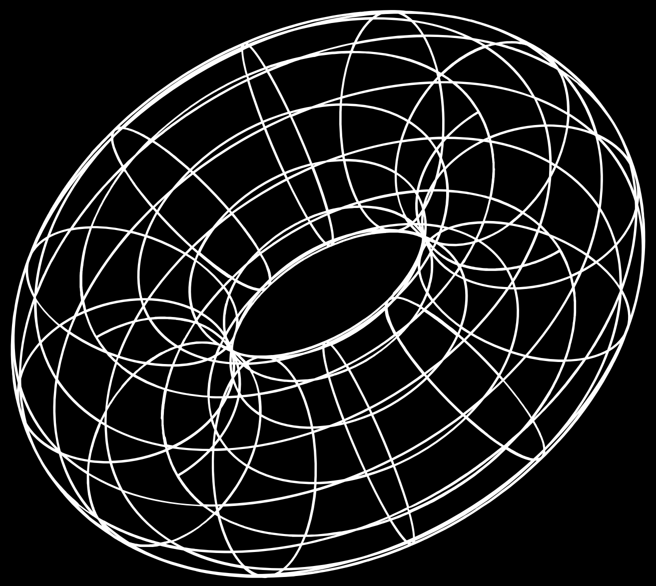
Once a gesture is recognized, it is instantly converted into readable text and then to Speech

FEATURE 2

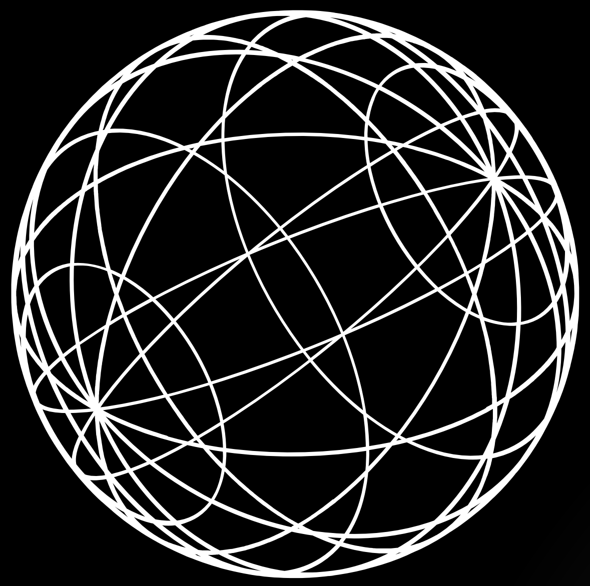
SPEECH TO SIGN-LANGUAGE CONVERSION



Voice Input Processing
The system captures the user's voice in real time using a microphone.



Text-to-Gesture Mapping
Each word or phrase is converted into a corresponding sign animation or gesture output.



Sign Language Output

- The output can be shown as either:
- Animated hand movement video
- On-screen ISL avatar

ADDITIONAL FEATURES

1. IMPLEMENTATION OF CHAT-BOT
2. MULTILINGUAL MODELS(ASL,BSL)
3. A USER-FRIENDLY APP IMPLEMENTATION

ROADMAP

Q1

- Problem Definition & Research Analysis
- ISL Dataset Collection (videos/images)
- Landmark Extraction using MediaPipe
- Data Labeling & Preprocessing

Q2

- CNN + LSTM Architecture Design
- Model Training & Hyperparameter Tuning
- Gesture-to-Text Mapping Logic
- Model Evaluation & Metrics (Accuracy, F1 Score)

Q3

- Flask Backend Integration
- Bolt Frontend Development (Gesture Input + Text Output)
- Real-Time Prediction via Webcam
- Final Testing, UI Polish & Demonstration

PRODUCT

Product 1

ISL Gesture Recognition Engine

- Uses CNN-LSTM to process real-time hand gestures and detect ISL signs from video frames.

Product 2

Text Conversion Module

- Translates detected gestures into meaningful English words or sentences.

Product 3

Text-to-Speech Converter

- Converts the output text into synthesized speech for better accessibility.

Product 4

User Interface (UI)

- Provides a simple, responsive web interface for live webcam input and output display using Flask + Bolt.

MODEL OVERVIEW

82%
Accuracy

0.73-0.81
F1 Score

THANK YOU