

Smart Glove Based Indian Sign Language Translator with ML Integration

Iniyavan R

*Dept. of Electronics and Communication Engineering
B.S. Abdur Rahman Crescent Institute of Science and Technology
Chennai, Tamil Nadu, India
iniyavan@crecident.education*

Sharon Christina K

*Dept. of Electronics and Communication Engineering
B.S. Abdur Rahman Crescent Institute of Science and Technology
Chennai, Tamil Nadu, India
220051601107@crecident.education*

Sharanya T

*Dept. of Electronics and Communication Engineering
B.S. Abdur Rahman Crescent Institute of Science and Technology Chennai, Tamil
Nadu, India
220051601106@crecident.education*

Zulfa Muneer M

*Dept. of Electronics and Communication Engineering
B.S. Abdur Rahman Crescent Institute of Science and Technology Chennai, Tamil
Nadu, India
220051601122@crecident.education*

Abstract—Communication remains a critical challenge for individuals with hearing and speech impairments in India, where Indian Sign Language (ISL) is the primary mode of interaction yet lacks widespread adoption. This paper presents a smart glove-based ISL translation system that integrates multimodal sensing with deep sequential learning for real-time gesture classification. The proposed system employs resistive flex sensors to capture finger joint articulation and an MPU-6050 IMU to acquire accelerometer and gyroscopic data for modeling hand orientation and motion dynamics. Sensor data is acquired via an ESP32 microcontroller, where preprocessing steps such as noise filtering, normalization, and temporal segmentation are performed prior to transmission. Temporal dependencies in gesture sequences are modeled using a Long Short-Term Memory (LSTM) network within Deep Learning, enabling effective learning of sequential patterns and motion trajectories. The model is trained on a custom ISL dataset and achieves a classification accuracy of 92.4%, outperforming Random Forest (88.7%), Support Vector Machine (79.3%), and K-Nearest Neighbors (74.5%). The system achieves an end-to-end inference latency of <150 ms, satisfying real-time processing requirements. Recognized gestures are mapped to textual outputs and synthesized speech via a React-based interface, with lightweight communication facilitated through the MQTT protocol. The proposed architecture demonstrates a robust, low-latency, and scalable assistive framework for ISL interpretation, with potential deployment in real-world assistive communication systems.

Keywords—Indian Sign Language, smart glove, LSTM, gesture recognition, flex sensor, MPU-6050, ESP32, MQTT, TensorFlow, assistive technology

I. INTRODUCTION

Communication remains a critical challenge for individuals with hearing and speech impairments, particularly in regions where Indian Sign Language (ISL) is the primary mode of interaction but is not widely understood by the general population. This limitation significantly impacts social inclusion, restricting access to essential services such as education, healthcare, and public assistance systems [1]. Consequently, the development of automated and real-time sign language translation systems has become an important research direction in assistive technologies.

Existing sign language recognition systems are broadly classified into **vision-based** and **sensor-based** approaches. Vision-based systems leverage image and video processing techniques, including convolutional neural networks (CNNs)

and pose estimation models, but are often constrained by environmental dependencies such as illumination variations, background clutter, occlusion, and camera positioning [2]. Moreover, such systems typically require high computational resources, limiting their deployment in real-time and embedded scenarios. In contrast, **wearable sensor-based systems** provide a more reliable alternative by directly capturing hand kinematics and finger articulation, thereby ensuring robustness against environmental variations and enabling continuous real-time operation [3].

Recent advancements in wearable sensing have facilitated the development of smart glove architectures for gesture recognition. These systems commonly employ **flex sensors** to measure finger bending through resistance variation, and inertial sensing units such as the MPU-6050 to capture 3-axis acceleration and angular velocity for modeling hand orientation and dynamic motion patterns. The integration of these sensing modalities enables effective **multimodal sensor fusion**, improving the discriminative capability of the system, particularly for gestures with similar finger configurations but differing motion trajectories [4]. Sensor acquisition, preprocessing, and wireless transmission are efficiently handled using embedded platforms such as the ESP32, which provides a balance between computational efficiency and real-time performance [5].

A key challenge in gesture recognition lies in accurately modeling the **temporal dependencies** inherent in sequential gesture data. Conventional machine learning algorithms, including Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), primarily rely on static feature representations and are therefore limited in capturing dynamic temporal patterns. To overcome this limitation, deep learning approaches, particularly **Long Short-Term Memory (LSTM)** networks within Deep Learning, have been widely adopted. LSTM networks incorporate gated memory units that enable the retention of long-term dependencies in time-series data, making them well-suited for modeling complex gesture sequences [6].

Motivated by these advancements, this paper proposes a **smart glove-based ISL translation system** that integrates multimodal sensing, efficient edge-level preprocessing, and LSTM-based sequence modeling for real-time gesture recognition. The proposed system captures gesture data using flex sensors and IMU measurements, performs preprocessing and normalization

on the embedded platform, and utilizes a trained LSTM model for classification. The recognized gestures are mapped to textual representations and synthesized speech through a React-based interface, while communication between system components is facilitated using the lightweight MQTT protocol, ensuring low-latency and scalable data exchange.

Experimental evaluation demonstrates that the proposed system achieves a classification accuracy of **92.4%**, outperforming traditional machine learning approaches, while maintaining an end-to-end latency of **less than 150 ms**, thereby satisfying real-time operational constraints.

The main contributions of this paper are summarized as follows:

- Design of a wearable smart glove integrating flex sensors and IMU for multimodal gesture acquisition.
- Implementation of an efficient edge-based data acquisition and preprocessing pipeline using the ESP32 microcontroller.
- Application of LSTM-based deep learning for modeling temporal dependencies in gesture sequences.
- Achievement of high classification accuracy (92.4%) with low inference latency (<150 ms).
- Development of a complete end-to-end assistive communication system with text and speech output.

II. RELATED WORK

A. Sensor-Based Sign Language Systems

Hrithik et al. [1] presented a machine learning-assisted smart glove system for converting sign language into speech, demonstrating strong real-time gesture classification. Ji et al. [2] introduced a wearable data glove using IMU sensors emphasizing motion-based gesture recognition for continuous sign sequences. Patil and Deshpande [3] designed a smart glove using flex sensors and supervised learning classifiers with gesture-to-text conversion capability.

Islam and Hossain [4] proposed a wearable sensor system focusing on feature extraction optimization for improved classification accuracy. Sharma and Verma [5] developed a real-time ISL recognition system with emphasis on minimal processing delay. Kumar et al. [6] combined flex and motion sensors with ML-based classification, while Rao and Reddy [7] demonstrated supervised learning on live smart glove sensor streams.

B. Identified Gaps

Across reviewed systems, several consistent limitations persist: (1) restricted gesture vocabularies not representative of actual ISL linguistic structure; (2) reliance on isolated gesture recognition without temporal sequence modeling; (3) insufficient handling of dynamic and continuous gestures due to lack of recurrent architectures; and (4) absence of ISL-specific datasets with adequate diversity in signing styles.

The proposed system addresses these gaps through LSTM-based sequence modeling, multi-sensor fusion combining flex and IMU data, a custom ISL dataset with at least 30 samples per gesture class, and a real-time MQTT-enabled deployment pipeline.

III. SYSTEM ARCHITECTURE

The proposed ISL Translator is designed as a five-layer architecture: Input Layer, Data Acquisition Layer, Machine Learning Layer, Application Layer, and Output Layer. This modular design ensures scalability and enables seamless integration between hardware and software components for real-time operation.

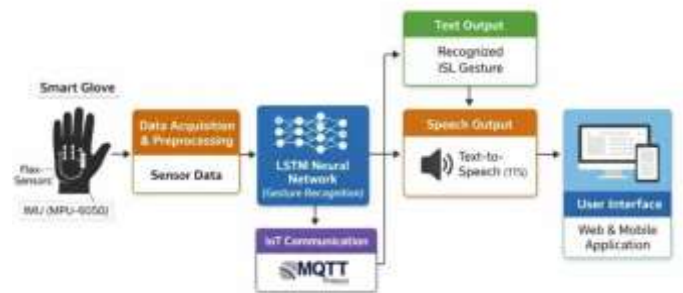


Fig. 1. Proposed System Block Diagram — ISL Smart Glove Translator

The system block diagram (Fig. 1) shows the data flow from the smart glove sensors through the ESP32 microcontroller, machine learning inference engine, and IoT communication layer to the final text and speech outputs.

A. Hardware Components

The ESP32 microcontroller (Fig. 2) serves as the central processing unit, featuring a dual-core Xtensa LX6 processor at up to 240 MHz, integrated Wi-Fi (802.11 b/g/n), and Bluetooth 4.2 support. Its multiple ADC channels interface with the flex sensors, and its I2C bus communicates with the MPU-6050 IMU.



Fig. 2. ESP32 Microcontroller

Five 2.2-inch flex sensors (Fig. 3) are mounted along the outer surface of each finger. Each sensor has a flat resistance of approximately 25 kΩ and maximum bent resistance of 45–125 kΩ, connected to ESP32 ADC pins via 10 kΩ voltage divider circuits, yielding 12-bit digital readings (0–4095).



Fig. 3. Flex Sensors

The MPU-6050 IMU (Fig. 4) combines a 3-axis gyroscope and 3-axis accelerometer, providing orientation, acceleration, and angular velocity data. Communication uses I2C at 400 kHz. Together with the five flex sensor readings, the IMU produces a 22-dimensional feature vector per frame that forms the input to the LSTM classifier.



Fig. 4. MPU-6050 IMU Sensor

The assembled smart glove (Fig. 5) houses the ESP32 in a wrist-mounted enclosure attached via velcro, powered by a 3.7V Li-Po battery. A trigger button on the index finger initiates and terminates gesture recording sessions, enabling clean segmentation of gesture sequences from continuous sensor streams.



Fig. 5. Assembled Smart Glove Hardware

IV. METHODOLOGY

The system pipeline (Fig. 6) transforms raw sensor data through six stages: gesture data collection, preprocessing and normalization, feature extraction, LSTM model training, real-time inference, and multimodal output generation.



Fig. 6. Methodology Flow Diagram

A. Data Collection

Gesture data is collected using a Python script with the smart glove hardware. Users perform predefined ISL gestures while holding the trigger button, with the ESP32 streaming 22 sensor values per frame at 50 Hz over USB serial. The script segments sequences between button presses and saves each recording as a CSV file. The dataset covers gestures including

Hello, Thank You, How Are You, Good Morning, and other common ISL phrases, with a minimum of 30 samples per class.

flex_thumb	flex_index	flex_middle	flex_ring	flex_pinky	pressure_x	pressure_y	label
610	700	704	698	684	0	0	Hello
614	783	704	701	685	0	0	Hello
617	780	704	699	685	0	0	Hello
607	784	704	722	686	0	0	Hello
595	784	706	700	683	0	0	Hello
596	769	692	699	686	0	0	Hello
611	784	705	699	684	0	0	Goodbye
618	784	704	701	682	0	0	Goodbye
592	779	699	699	683	0	0	Goodbye
615	780	702	698	679	0	0	Goodbye
614	778	699	700	677	0	0	Goodbye
597	779	698	699	682	0	0	Goodbye
612	767	690	699	679	0	0	Goodbye
619	778	699	701	677	0	0	Goodbye

Fig. 7. Collected Dataset Sample (flex and IMU readings per frame)

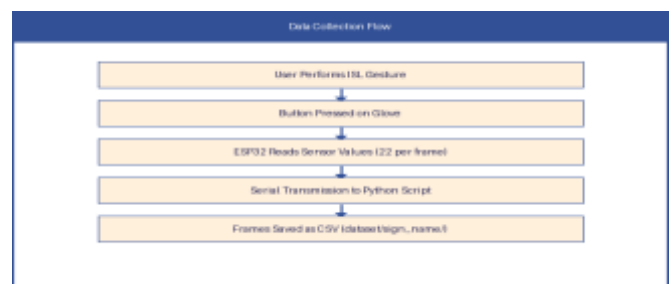


Fig. 8. Data Collection Flow

B. Data Preprocessing

Raw sensor data undergoes: (1) min-max normalization scaling all features to [0, 1]; (2) moving average noise filtering to smooth electrical interference; and (3) sequence standardization to a fixed length of 150 frames via zero-padding or truncation. The final dataset structure is a 3D NumPy array of shape (n_samples, 150, 22). Labels are one-hot encoded for multi-class classification.

C. LSTM Model Architecture

The LSTM model (Fig. 10) is implemented in TensorFlow/Keras. The network accepts input sequences of shape (150, 22) and processes them through two stacked LSTM layers with 128 and 64 units respectively. A dropout layer (rate=0.3) follows the first LSTM layer to reduce overfitting. A Dense layer with 32 ReLU units precedes the final Softmax output layer. The model is compiled with the Adam optimizer (lr=0.001) and categorical cross-entropy loss.

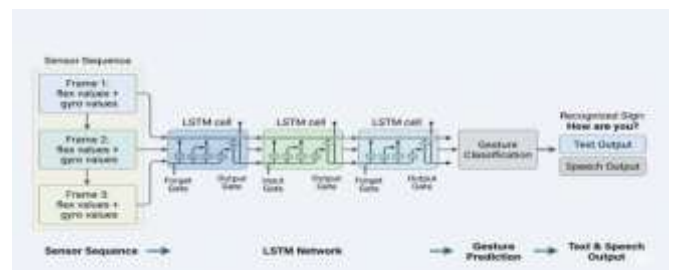


Fig. 9. LSTM Neural Network Architecture

D. Training and Validation

The dataset is divided 70%/15%/15% for training, validation, and testing respectively. Training proceeds for up to 50 epochs with batch size 32, using early stopping (patience=10) on validation loss. The trained model is saved as sign_language_model.h5 with the label encoder (label_encoder.pkl) for deployment in the real-time inference pipeline.

V. PERFORMANCE EVALUATION

A. Model Comparison

The LSTM model is benchmarked against Random Forest, SVM, and KNN classifiers trained on the same dataset. Figures 11–14 illustrate the comparative performance across accuracy, precision, recall, and F1-score. The LSTM consistently outperforms all alternatives across all metrics, confirming the advantage of temporal sequence modeling for gesture data.

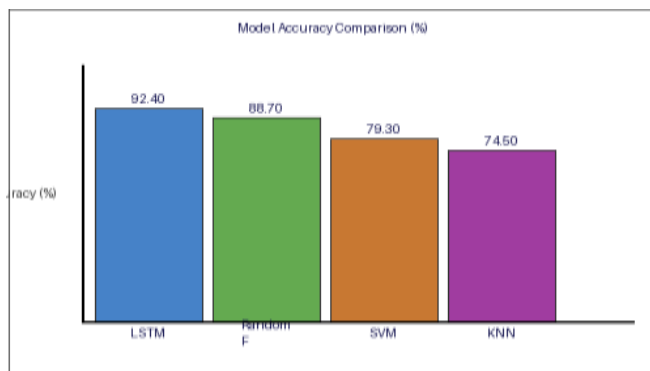


Fig. 10. Accuracy Comparison of ML Models (%)

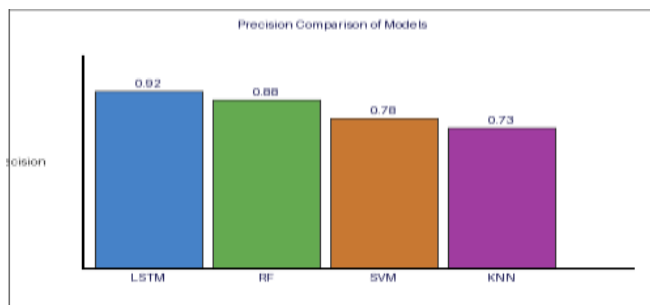


Fig. 11. Precision Comparison of ML Models

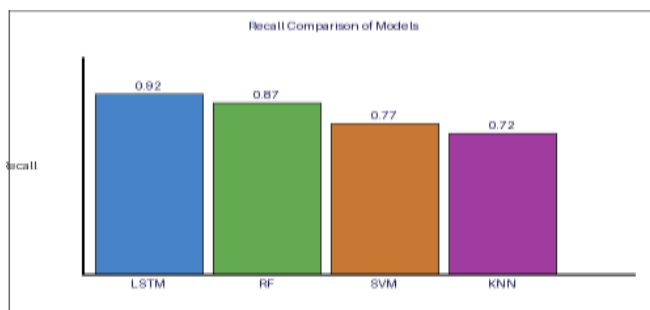


Fig. 12. Recall Comparison of ML Models

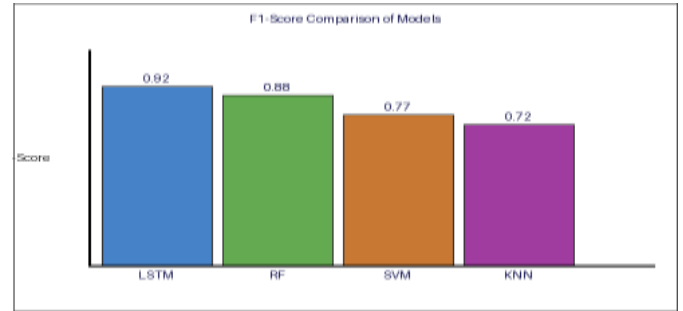


TABLE I. PERFORMANCE COMPARISON OF ML MODELS (ISL GESTURE RECOGNITION)

Model	Accuracy(%)	Precision	Recall	F1 Score
LSTM	92.4	0.921	0.918	0.919
Random Forest	88.7	0.879	0.874	0.876
SVM	79.3	0.781	0.768	0.774
KNN	74.5	0.732	0.718	0.725

B. System Performance

System-level performance was evaluated under integrated operation with the smart glove, Flask-SocketIO backend, and React frontend running simultaneously. Key performance metrics are summarized in Table II. The sub-150ms LSTM inference latency combined with the 25 FPS camera pipeline and MQTT delivery under 80ms ensures smooth real-time interaction. Confidence scores for correctly recognized gestures consistently exceeded 0.85, with common gestures such as Hello, Thank You, and How Are You achieving confidence values above 0.93.

TABLE II. SYSTEM PERFORMANCE SUMMARY

Performance Metric	Result
LSTM Accuracy	92.4%
Avg. Inference Time (LSTM)	< 150 ms
Camera Frame Rate	~25 FPS
Gesture Confirm Latency	~0.5 s
MQTT Delivery Latency	50–80 ms
Registered Gestures	15 ISL Phrases
TTS Response Time	< 200 ms

VI. SOFTWARE IMPLEMENTATION

A. Backend Architecture

The backend is implemented in Python using Flask and Flask-SocketIO for real-time bidirectional communication. The SignInterpreter class manages model loading, landmark extraction, sequence building, and gesture prediction. MQTT integration via paho-mqtt enables wireless reception of hardware glove predictions for display in the unified web interface.



Fig. 14. Gesture Registration Interface

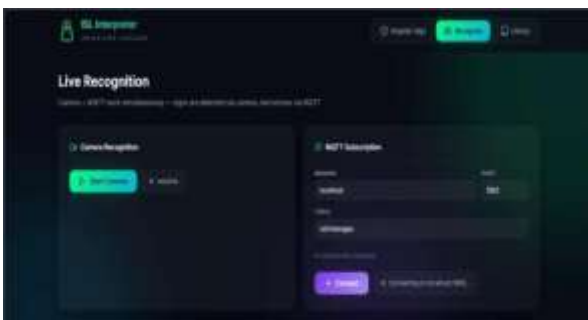


Fig. 15. Real-Time Recognition Interface

B. MediaPipe Camera Pipeline

A parallel camera-based recognition pipeline uses Google MediaPipe Hands to detect and track 21 hand landmarks per frame in 3D. A 126-dimensional feature vector (21 landmarks $\times$ 3 coordinates $\times$ 2 hands) is extracted per frame. A Random Forest classifier with a sliding 30-frame window and 5-prediction majority voting buffer (confidence threshold 0.5) provides confirmed gesture detection. A cooldown of 50 frames prevents duplicate detections.

VII. RESULTS AND DISCUSSION

A. ISL Alphabet Recognition

The system accurately recognized ISL alphabets during testing sessions. Representative results for alphabets F (95% confidence), M (96%), R (97%), and Y (98%) are shown in Fig. 17, demonstrating consistent high-confidence recognition with matched sensor readings visible in the prediction history panel.



Fig. 16. ISL Alphabet Recognition Results (F)

B. Word and Sentence Recognition

Common ISL words including Hello (98%), No (95%), and Goodbye (96%) were recognized with high confidence, as shown in Fig. 18. Sentence-level recognition for phrases such as 'Hello everyone' and 'I like to walk' demonstrated end-to-end system operation including text display and text-to-speech conversion (Fig. 17).



Fig. 17. Word Recognition — Hello (98%)



Fig. 18. System Output — Sentence-Level ISL Translation

C. MQTT Communication

MQTT communication between the smart glove inference script and the web interface achieved delivery latency of 50–80 ms, enabling near-real-time wireless gesture display. The system confirmed 24 messages received from the ESP32_Glove source during integrated testing (Fig. 19), demonstrating reliable IoT-enabled operation.

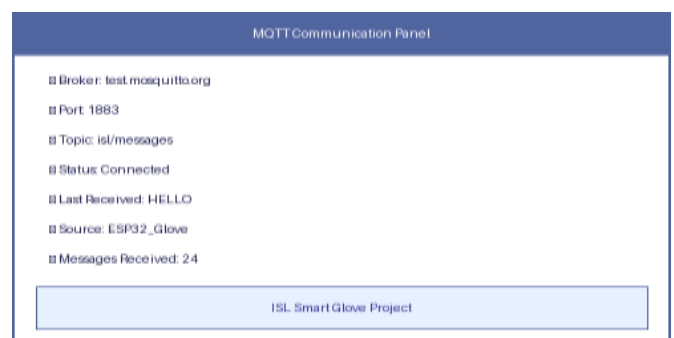


Fig. 19. MQTT Communication Panel

VIII. CONCLUSION

This paper presented a Smart Glove-Based Indian Sign Language Translator integrating ESP32-based wearable sensing, LSTM deep learning, IoT communication via MQTT,

and a React-based web interface. The system achieves 92.4% gesture recognition accuracy with sub-150ms inference latency, enabling practical real-time ISL translation with both text and speech outputs.

The dual-pathway architecture combining hardware glove inference via MQTT with camera-based MediaPipe recognition enhances versatility. The system demonstrated stable real-time performance with gesture confirmation latency under 0.5 seconds. Future work will focus on expanding the ISL vocabulary, implementing transformer-based sequence models, bilateral two-glove support, NLP-based sentence construction from gesture sequences, and edge deployment of the inference model directly on the ESP32 microcontroller.

IX. REFERENCES

- [1] Hrithik T. H., Rhethika S., Akhil K. H., and Deepa K., "ML Assisted Sign Language to Speech Conversion Gloves for the Differently Abled," IEEE IATMSI, 2024.
- [2] X. Ji, Y. Zhao, and X. Yu, "Dataglove for Sign Language Recognition of People with Hearing and Speech Impairment via Wearable Inertial Sensors," IEEE Sensors Journal, 2023.
- [3] S. K. Patil and R. R. Deshpande, "Smart Glove Based Sign Language Recognition System Using Machine Learning," IEEE ICSSIT, 2022.
- [4] M. R. Islam and M. A. Hossain, "Wearable Sensor Based Sign Language Recognition Using Machine Learning," IEEE ICBECA, 2021.
- [5] P. K. Sharma and A. Verma, "Real-Time Indian Sign Language Recognition System Using Machine Learning Techniques," IEEE ICCCA, 2020.
- [6] A. Kumar, S. Mehta, and R. Jain, "Smart Wearable System for Sign Language Interpretation Using Machine Learning," IEEE ICAIS, 2023.
- [7] N. S. Rao and P. S. Reddy, "Hand Gesture Recognition Using Sensor Based Smart Glove and Machine Learning," IEEE ICECA, 2022.
- [8] S. Banerjee and A. Mukherjee, "Indian Sign Language Recognition Using Wearable Sensors and Machine Learning Algorithms," IEEE ICSPC, 2021.
- [9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10] F. Zhang et al., "MediaPipe Hands: On-device Real-time Hand Tracking," arXiv:2006.10214, 2020.
- [11] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning," Proc. OSDI, 2016.
- [12] A. Banks and R. Gupta, "MQTT Version 3.1.1," OASIS Standard, 2014.