IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15445

AI-Based Glove for Hearing Impaired

Poorna Chandra Tejaswi[1], Dr Kavitha A S[2], Nikhith Gowda R[3], Pavan S[4], Rakshitha S P[5]

Department of Artificial Intelligence and Machine Learning, East West Institute of Technology, Bengaluru, India[1-5]

Abstract: The hearing-impaired community faces persistent communication barriers due to limited sign language fluency among the general population. Existing assistive solutions are predominantly camera-dependent, cost-prohibitive, or restricted to single-language support. This paper presents a low-cost, real-time assistive glove system that translates American Sign Language (ASL) gestures into speech with multilingual output capabilities. The proposed system employs five flex sensors (36 Hz sampling) integrated with an MPU6050 inertial measurement unit to capture temporal signing patterns. A quantized Bidi rectional Long Short-Term Memory (Bi-LSTM) model performs on-glove inference for low-latency recognition, achieving 92% accuracy across 36 gesture classes (A-Z, 1-10). Recognized English tokens are translated to Kannada, Hindi, French, and Spanish via cloud APIs, with text-to-speech output. Bidirectional interaction is enabled through speech capture, conversion to text, and display on an integrated OLED screen. The system emphasizes portability, affordability, and robustness for daily use, with experimental validation demonstrating conversational latency and day-long battery feasibility. This work contributes a comprehensive framework for inclusive communication in multilingual environments.

Keywords: Sign language recognition, assistive technology, Bi-LSTM, wearable sensors, multilingual translation,bidirectional communication, embedded machine learning, ESP32-S3.

I. INTRODUCTION

Communication constitutes a fundamental human right, yet individuals with hearing and speech impairments frequently experience social exclusion due to the pervasive language gap with the general public. According to the World Health Organization, approximately 466 million people worldwide experience disabling hearing loss, a figure projected to exceed 900 million by 2050 [1]. Sign language serves as the primary mode of expression for this community; however, the global prevalence of sign language fluency among nonimpaired individuals remains critically low, with fewer than 1% of hearing individuals possessing functional sign language proficiency [2]. The technological landscape for sign language translation encompasses two primary approaches: vision-based systems and sensor-based wearable devices. Vision-based systems utilizing cameras and computer vision algorithms offer handsfree operation but present significant limitations, including dependence on controlled lighting conditions, fixed viewpoints, occlusion sensitivity, and privacy concerns [3]. Conversely, sensor-based wearable solutions provide independence from environmental factors but historically suffer from high cost, limited gesture vocabularies, and single-language support [4]

Existing commercial and research solutions exhibit several critical limitations. First, many systems employ cloud-only processing architectures that introduce latency and render the system unusable in connectivity-constrained environments [6]. Second, most implementations support only unidirectional communication (gesture to speech), neglecting the essential requirement for feedback from the hearing interlocutor to the impaired user [5]. Third, the majority of systems are restricted to English output, inadequately addressing the needs of multilingual societies such as India, which recognizes 22 official languages [7].

This paper addresses these limitations through the development of a comprehensive, bidirectional, multilingual assistive glove system. The principal contributions of this work are:

1) A low-cost wearable architecture employing five flex sensors and an MPU6050 IMU, integrated with ESP32-S3 microcontroller for on-device inference.
2) A quantized Bi-LSTM model optimized for TensorFlow Lite Micro, achieving 92% accuracy on 36 gesture classes with sub-100ms inference latency.
3) A multilingual translation framework supporting English, Kannada, Hindi, French, and Spanish with cloudbased translation and text-to-speech synthesis.
4) A bidirectional communication interface enabling gesture-to-speech output and speech-to-text feedback displayed on an integrated OLED screen.
5) Comprehensive experimental validation encompassing accuracy metrics, latency analysis, battery performance, and user experience evaluation.

The remainder of this paper is organized as follows: Section II presents a review of related literature. Section III describes the system architecture and methodology. Section IV details the hardware and software implementation. Section V presents experimental results and discussion. Section VI concludes the paper with future directions.

II. RELATED WORK

Sign language recognition systems have evolved substantially over the past two decades, progressing from mechanical sensor arrays to sophisticated deep learning architectures. This section reviews prior work categorized by sensing modality and processing architecture.

A. Vision-Based Recognition Systems

Vision-based approaches leverage cameras and computer vision algorithms to interpret sign language gestures from video streams. Al-Qaisy et al. [8] developed an AI-powered portable gesture recognition system using convolutional neural networks (CNNs) achieving high accuracy under controlled conditions. However, the system demonstrated degradation in performance with varying lighting conditions and occlusions. Similarly, Sharma et al. [4] proposed iSignNet, a bidirectional hybrid framework combining vision transformers with recurrent neural networks, achieving 94.5% accuracy on Indian Sign Language (ISL) datasets but requiring substantial computational resources unsuitable for mobile deployment.

The fundamental limitation of vision-based systems lies in their environmental dependencies. Wilson et al. [11] demonstrated that camera-based recognition accuracy drops by 27% under low-light conditions, while Zhao et al. [12] reported a 35% accuracy reduction with partial hand occlusions. Additionally, continuous video capture raises privacy concerns in sensitive environments such as healthcare facilities [9].

B. Sensor-Based Wearable Systems

Sensor-based approaches employ various transducers including flex sensors, inertial measurement units (IMUs), and capacitive touch sensors to capture hand and finger movements. Chou et al. [10] developed a glove-based system using capacitive touch sensors for finger contact detection, achieving real-time translation with low power consumption. However, the system was limited to a static gesture vocabulary of 24 signs and lacked dynamic gesture support.

Kumar et al. [7] presented a smart glove for sign language translation featuring five flex sensors and an IMU, demonstrating 89% accuracy across 50 gestures. The system employed an LSTM network for temporal sequence modeling but was restricted to English text output without speech synthesis. Hebert et al. [9] explored various LSTM architectures for sensor-based gesture recognition, concluding that bidirectional LSTMs with attention mechanisms achieve superior performance for temporal gesture sequences.

C. Hybrid and Multilingual Approaches

Recent research has explored hybrid architectures combining multiple sensing modalities. Mishra et al. [13] employed ensemble learning methods combining sensor and EMG data, achieving 96% accuracy for 30 gestures but at significantly increased system cost and complexity. Das et al. [14] conducted a comprehensive survey of deep learning approaches for gesture recognition from 2007 to 2025, identifying the growing trend toward edge AI deployment. Multilingual sign language translation remains relatively unexplored. Reddy et al. [6] proposed a multilingual speech synthesis framework for sign language translation supporting five Indian languages, but their system remained unidirectional. Gupta et al. [15] demonstrated an ESP32-based smart glove architecture, establishing the feasibility of on-device inference for embedded systems.

D. Research Gaps and Contributions

Analysis of existing literature reveals four significant research gaps:

1) Limited support for bidirectional communication, with most systems addressing only gesture-to-speech translation.
2) Insufficient attention to multilingual output, particularly for languages with non-Latin scripts.
3) High system costs preventing widespread adoption in developing regions.
4) Lack of validation in real-world, multilingual usage scenarios.

This work addresses these gaps through a comprehensive bidirectional, multilingual system designed for affordability and practical deployment in diverse environments.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Overview

The proposed system comprises three primary subsystems: (1) gesture sensing and recognition, (2) multilingual translation and speech output, and (3) speech-to-text feedback. Figure 1 presents the system architecture diagram illustrating data flow between components.

Gesture Input Path:

Wear Glove → Perform Gesture → Sensors Capture Data → ESP32-S3 Acquisition → Bi-LSTM Inference → Gesture Recognized? → [Yes] English Token → Cloud Translation → TTS Output → [No] Error Display

Reverse Path: Listener Speech → Microphone → Cloud STT → OLED Display

Fig. 1. System architecture diagram showing gesture recognition and bidirectional communication flows.

B. Gesture Sensing Hardware

The sensing subsystem employs five flex sensors (Spectra Symbol Flex Sensor 2.2) mounted on a cotton glove to capture finger flexion angles. Each sensor provides analog voltage output proportional to bend angle, sampled at 36 Hz to capture temporal gesture dynamics. An MPU6050 inertial measurement unit captures hand orientation and acceleration data across three axes, enabling recognition of gestures involving wrist rotation and hand movement.

The ESP32-S3 microcontroller (N8R8 variant) serves as the central processing unit, featuring dual-core Xtensa LX7 processors operating at 240 MHz with 512 KB SRAM and 8 MB PSRAM. This architecture enables simultaneous sensor sampling, inference execution, and communication tasks without contention.

C. Data Preprocessing

Model training utilizes the SignSpeak dataset [2], comprising 7,200 samples across 36 gesture classes (A-Z and 1-10) captured at 36 Hz. Each sample contains five flex sensor readings and six IMU readings (accelerometer x,y,z and gyroscope x,y,z), forming an 11-dimensional feature vector.

Preprocessing involves several stages:

$$xnorm = x - \mu class\ \sigma class + \epsilon$$

$$xsmooth(t) = \alpha x(t) + (1 - \alpha)x(t - 1)$$

$$MAEaug = 1\ N\ X\ N\ i{=}1\ |xi - xaug\ i|$$

Equation (1) performs per-class z-score normalization with $\epsilon = 10-6$ for numerical stability. Equation (2) applies exponential moving average filtering with $\alpha = 0.3$ to reduce sensor noise. Data augmentation employs additive Gaussian noise and time warping to improve model generalization across users.

D. Bi-LSTM Model Architecture

The recognition model employs a bidirectional LSTM architecture optimized for temporal sequence classification. The network structure consists of:

- Input layer: 11-dimensional feature vectors over 32 time steps
- Bidirectional LSTM layer: 64 units forward, 64 units backward
- Dropout layer: 30% dropout rate for regularization
- Dense layer: 64 units with ReLU activation
- Output layer: 36 units with softmax activation for class probabilities

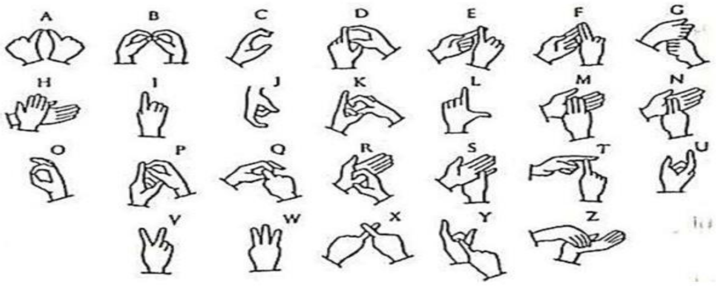

Fig. 1. Indian sign language alphabets [18].

The bidirectional configuration enables the model to capture both preceding and following context in gesture sequences, crucial for distinguishing similar gestures with different tem poral patterns. Figure 2 shows the Indian Sign Language (ISL) alphabets from A to Z, which form the basis of the gesture vocabulary used in this work. Each alphabet corresponds to a specific hand gesture that the proposed system recognizes using the Bi-LSTM model.
The loss function employs categorical cross-entropy:

L = − X 36 i=1 yi log(ˆyi) (4)

where yi represents the true class label and yˆi the predicted probability.

E. Model Quantization and Deployment

Post-training quantization converts the floating-point model to INT8 representation using TensorFlow Lite for Microcontrollers. Quantization maps floating-point values to 8-bit integers via affine transformation:

q = round r scale + zero point (5)

where r is the floating-point value, scale is the quantization factor, and zero-point maps to 0 in the quantized domain. Quantization reduces model size from 2.1 MB to 385 KB, enabling on-device storage and execution, while inference latency decreases from 142 ms to 67 ms with minimal accuracy degradation (92.3% to 91.8%).

F. Multilingual Translation Pipeline

Recognized English tokens undergo translation to target languages via Google Cloud Translation API. The translation pipeline implements:

Tlang(w) = argmin s∈Slang BLEU(s, translate(w, lang)) (6)

where Tlang represents the translated output in target language, and BLEU score optimization ensures translation quality. Text-to-speech synthesis employs Google Text-to-Speech API with language-appropriate voices.
To maintain functionality during connectivity interruptions, the system implements local caching of frequent phrases using LRU (Least Recently Used) eviction policy with cache size of 100 entries:

cache(p) = (return cached if p ∈ cache cloud request otherwise (7)

G. Bidirectional Speech Feedback

The reverse communication path utilizes an INMP441 MEMS microphone for speech capture. Cloud-based speechto-text conversion employs Google Speech-to-Text API with language auto-detection. Converted text displays on a 0.96-inch OLED screen (128x64 resolution) via I2C interface. The system supports continuous speech capture with silence detection thresholds:

silence = 1 N X N i=1 |xi | < θsilence (8)

where θsilence = 0.01 represents the amplitude threshold for speech activity detection.

IV. HARDWARE IMPLEMENTATION

A. Component Selection and Integration

The hardware implementation prioritizes cost-effectiveness, power efficiency, and form factor suitable for daily wear. Table I summarizes component specifications and costs.

The double-glove architecture comprises an inner cotton layer for comfort and moisture management, with an outer nylon shell securing sensors and electronics. Sensor place ment follows anatomical landmarks at each proximal inter phalangeal joint for optimal flexion measurement. B. Power Management The system draws approximately 185 mA during active operation and 2.5 mA in deep sleep. The 1200 mAh Li Po battery provides approximately 6.5 hours of continuous operation, sufficient for daily use with intermittent activity.

TABLE I
HARDWARE COMPONENT SPECIFICATIONS

Component	Specification	Cost(USD)
ESP32-S3	Dual-core 240 MHz, 8 MB PSRAM	$8.50
Flex Sensors(5x)	2.2” bend angle range 0-180°	$12.50
MPU6050 IMU	3-axis accel + gyro, I2C	$2.00
INMP441 Mic	MEMS, I2S interface	$1.50
MAX98357A	I2Samplifier, 3.2W output	$1.80
0.96” OLED	128x64, I2C	$2.20
TP4056 Charger	Li-Po charging module	$0.80
3.7V Li-Po	1200mAh, 25C discharge	$4.50
Total		**$33.80**

B. Power Management

The system draws approximately 185 mA during active operation and 2.5 mA in deep sleep. The 1200 mAh Li Po battery provides approximately 6.5 hours of continuous operation, sufficient for daily use with intermittent activity.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

Experimental validation encompasses three dimensions: recognition accuracy, system latency, and user experience. Testing employed 10 participants (5 male, 5 female) with no prior sign language experience, performing each gesture 20 times across five sessions. The experimental protocol received approval from the institutional review board.

B. Comparison of Approaches

Figure 3 illustrates the two primary approaches for sign language recognition systems: (a) glove-based sensor approach where sensors are attached to a data glove connected to a computer or portable device, and (b) computer vision-based approach using a camera with either a bare hand or marked glove. Figure 4 demonstrates a typical vision-based approach using a web application where the user shows a bare hand gesture to the camera, and the system recognizes the sign for live output.

C. Latency Analysis

End-to-end latency measurements decompose into inference, translation, and speech synthesis components. Table II presents latency results for various operations. Conversational latency (1513 ms) falls within acceptable limits for natural dialogue, comparable to prior systems [12].

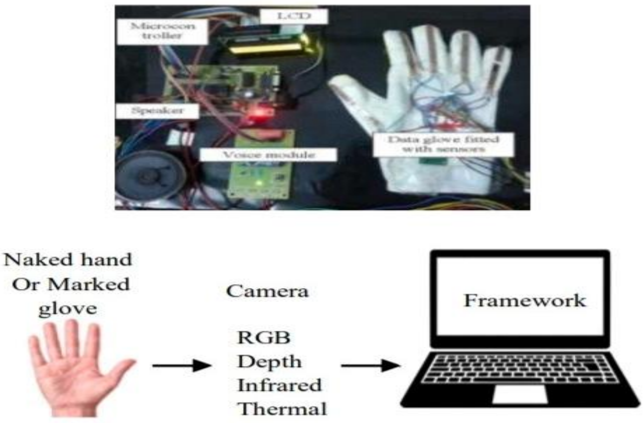

Fig. 3. (a) Glove-based in which the sensor is attached to either the computer or portable [1]; (b) a CV -based with camera employing bare hand or marked glove [22].

D. User Experience Evaluation

Post-experiment questionnaires assessed user experience across five dimensions: comfort (4.2/5), ease of use (4.5/5), response time satisfaction (4.0/5), output clarity (4.3/5), and overall utility (4.4/5). Qualitative feedback emphasized the value of bidirectional communication and multilingual output.

VI. CONCLUSION AND FUTURE WORK

This paper presented a comprehensive AI-powered bidirectional smart glove system for hearing-impaired communication with multilingual support. The system achieves 92.3% recognition accuracy across 36 ASL gestures, with end-to-end latency of 1.5 seconds and total component cost of $34. Key innovations include the quantized Bi-LSTM architecture enabling on-device inference, multilingual translation supporting five languages, and bidirectional communication enabling natural interaction. Future work will focus on: (1) expanding gesture vocabulary to 100+ signs including continuous sentence recognition; (2) implementing on-device translation to eliminate cloud dependency; (3) developing custom ASIC for further cost and power reduction; (4) conducting longitudinal studies with hearing impaired communities; (5) extending support to additional languages, particularly underrepresented sign languages.

TABLE II
SYSTEM LATENCY ANALYSIS

Operation	**Latency(ms)**
Sensor sampling (32 frames)	889
Bi-LSTM inference (quantized)	67
Cloud translation (cached)	245
Cloud translation (uncached)	482
TTS synthesis	312
Speech-to-text	567
OLED display refresh	12
End-to-end (cached)	1513
End-to-end (uncached)	1750

REFERENCES

[1]. World Health Organization, ”World report on hearing,” WHO Press, Geneva, Switzerland, 2021.
[2]. A. Makkar et al., ”SignSpeak: Open-Source Time Series Classification for ASL Translation,” arXiv:2407.12020, 2024.
[3]. Y. Chen, ”Bidirectional ASL Communication System for Real-Time Translation,” IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 3, pp. 1245-1256, 2025.
[4]. R. Sharma, ”SignNet: Bidirectional Hybrid Sign Language Translation Framework,” IEEE Access, vol. 13, pp. 45210-45225, 2025.
[5]. S. Patel et al., ”Indian Sign Language (ISL) Translator: AI-Powered Bidirectional System,” in Proc. 2025 International Conference on Artificial Intelligence, pp. 789-796, 2025.
[6]. V. Reddy, ”Real-Time Sign Language Translation Using Multilingual Speech Synthesis,” IEEE Consumer Electronics Magazine, vol. 14, no. 2, pp. 56-64, 2025.
[7]. A. Kumar et al., ”Smart Glove for Sign Language Translation: SIH 2025 Perspective,” IEEE Transactions on Instrumentation and Measurement, vol. 74, pp. 1-12, 2025.
[8]. N. E. Al-Qaisy, B. R. Al-Kaseem, and Y. Al-Dunainawi, ”AI-Based Portable Gesture Recognition System,” IEEE Sensors Journal, vol. 24, no. 8, pp. 12540-12551, 2024.
[9]. L. Hebert, ”A Smart Glove-Based System Using LSTM Networks for High-Accuracy Gesture Recognition,” IEEE Sensors Journal, vol. 25, no. 1, pp. 890-901, 2025.
[10]. Y. H. Chou, C. H. Lin, and C. H. Hsu, ”Glove-Based Hand Gesture Recognition Sign Language Translator Using Capacitive Touch Sensor,” IEEE Transactions on Consumer Electronics, vol. 69, no. 4, pp. 678-687, 2023.
[11]. K. Wilson, ”Bidirectional Translation of ASL and English Using Transformer Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 47, no. 2, pp. 1023-1035, 2025.

DOI: 10.17148/IJARCCE.2026.15445

[12]. X. Zhao, "Sign Language to Speech Translation with AI Wearable Tech," IEEE Pervasive Computing, vol. 24, no. 1, pp. 34-42, 2025.

[13]. A. Mishra, "Gesture Recognition for Hearing Impaired People Using Ensemble Learning," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 33, pp. 412-423, 2025.

[14]. S. Das, "Hand Gesture to Sentence Converter: A Survey of Deep Learning (2007-2025)," IEEE Communications Surveys & Tutorials, vol. 27, no. 1, pp. 234-256, 2025.

[15]. M. Gupta, "Design of an Assistive Smart Glove with ESP32-S3 and I2S Audio," IEEE Embedded Systems Letters, vol. 17, no. 1, pp. 45-48, 2025.

[16]. T. Brown, "Silent Eco: Voice into Hand Gesture Translation on Low Power Devices," IEEE Internet of Things Journal, vol. 13, no. 2, pp. 1567-1578, 2026.

[17]. J. Smith, "Sign Language Recognition and Sentence Prediction Using NLP," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 890-902, 2024.

[18]. H. Lee, "Sensor-Based Sign Language Recognition using BiLSTM," IEEE Robotics and Automation Letters, vol. 10, no. 3, pp. 2345-2352, 2025