www.ijarp.com ISSN 2456-9992 Page: 01-13

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 04

DESIGN AND DEVELOPMENT OF A SIGN LANGUAGE TO VOICE CONVERSION SYSTEM

[1]Manav Neemkar, [2]Sri Harshini Ramini, [3]Asker Abbas, [4]V. Lydia Roselyn, *[5]Dr. R. Uday Kumar

[1,2,3,4] B.Tech (MCT) Students, Dept. of Mechanical Engine Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad.

[5]Associate Professor, Dept. of Mechanical Engineering, Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad.

Article Received: 15 March 2026, Article Revised: 04 April 2026, Published on: 24 April 2026

***Corresponding Author: Dr. R. Uday Kumar**

Associate Professor, Dept. of Mechanical Engineering, Mahatma Gandhi Institute of Technology, Gandipet, Hyderabad.

DOI: https://doi-doi.org/101555/ijarp.7525

ABSTRACT

The proposed system is a lightweight, sensor-integrated wearable glove designed to translate hand gestures into audible speech in real time. It incorporates flex sensors and inertial measurement units (IMUs) embedded across the fingers and hand to capture fine-grained motion with high precision. Emphasizing portability, low-latency performance, and ergonomic comfort, the glove is optimized for everyday use. Its embedded-machine-learning architecture (TinyML) enables all signal processing and classification to run directly on-device, avoiding the limitations of camera-based systems such as lighting dependency, privacy concerns, and high processing overhead. The glove supports fast calibration and a customizable vocabulary, enabling users with different signing styles to adapt the system to their needs. By combining efficient gesture recognition with on-board speech synthesis, it serves as a practical and accessible assistive communication tool suitable for daily use.

The dual-glove configuration employs a total of ten flex sensors five on each glove and two MPU- 6050 IMUs to capture finger bend angles, wrist movement, and dynamic hand orientation. Sensor data is sampled at 50 Hz and transmitted wirelessly using the ESP-NOW protocol for low-power, low- latency communication. A hybrid 1D CNN–LSTM deep learning model operates on 100-sample sliding windows comprising 16 features (10 flex + 6 IMU axes), achieving gesture classification accuracy between 90–95% for Indian Sign

Language (ISL). After training, the model is quantized to INT8 format and deployed on the ESP32-S3 microcontroller using TensorFlow Lite Micro, reducing memory usage to approximately 0.15 MB. Recognized gestures are converted into audible English speech through a Bluetooth speaker via a text-to-speech engine. The system currently supports over 100 ISL words, with scalability to 500+ planned in future iterations, and maintains an end-to-end latency below 100 ms while operating fully offline. With a prototype system aims to deliver affordable, real-world assistance across educational, healthcare, workplace, and public-interaction environments. In addition to technical efficiency, the system is designed with user-centric considerations such as ease of operation, minimal hardware complexity, and long-term reliability. The modular structure allows future upgrades in vocabulary, gesture sets, and model accuracy without redesigning the entire hardware. This combination of scalability, affordability, and robustness makes the glove a promising foundation for next-generation assistive communication devices.

KEYWORDS: Sign Language Recognition, Flex Sensors, CNN-LSTM, Assistive Technology, Gesture-to- Speech System.

INTRODUCTION

Communication is the foundation of human interaction, enabling individuals to exchange ideas, express emotions, and participate meaningfully in society. For millions of people with hearing or speech impairments, sign language serves as their primary mode of communication. Sign languages, including Indian Sign Language (ISL), are rich, structured, and expressive visual languages that rely on combinations of hand gestures, facial expressions, and body movements. However, a significant communication barrier arises when sign language users interact with individuals unfamiliar with signing. This communication gap often restricts access to education, healthcare services, employment opportunities, and day-to-day social engagement, ultimately affecting the quality of life for the hearing- and speech-impaired community. Bridging this gap through accessible, affordable, and real-time technological solutions has become an increasingly important area of research in assistive technology (1-3). Traditional approaches to sign language translation have relied heavily on vision-based systems using cameras, computer vision, and image-processing algorithms. While these techniques can produce reasonable accuracy, they bring several limitations. Camera-based systems are sensitive to lighting conditions, background clutter, occlusion of hands, and require a fixed setup for reliable performance. They also raise privacy concerns,

especially in public environments, hospitals, or classrooms where continuous video recording is undesirable. Moreover, high-computation hardware and GPUs are often required, making these solutions costly and less portable (4-7). These challenges create a clear need for a more robust, user-friendly alternative capable of functioning efficiently in real-world environments without complex infrastructure. Sign Language (ISL), used by more than 7 million individuals across India, has limited technological support compared to American or British Sign Language. Research and development focused on ISL remain relatively sparse, creating an urgent need for tools that cater specifically to the Indian context. ISL consists of distinct gestures representing alphabets, words, and expressions, many involving both hands for accurate representation. The dual-glove approach in this project is designed to address the complexities of two-handed gestures, which are essential for authentic ISL recognition. Through the integration of flex sensors and IMUs on both gloves, the system captures a rich set of motion features necessary for accurate classification across a diverse vocabulary. The machine learning model used in this project follows a hybrid architecture combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) layers. CNNs are effective in extracting spatial patterns from sensor signals, while LSTMs specialize in understanding temporal dependencies. Together, they enable robust classification of dynamic gestures involving continuous motion. The model is trained on segmented sliding windows of sensor data and quantized to INT8 format to reduce memory usage for deployment on the ESP32-S3(8-12). Quantization ensures that the model retains high accuracy while occupying only 0.15 MB of storage, demonstrating the efficiency of TinyML deployment on resource-constrained devices.

METHODOLOGY

The system architecture of the sign language to voice conversion glove is designed to provide a fully integrated, real time, and user friendly solution that translates Indian Sign Language gestures into clear spoken English. The overall architecture combines sensing units, embedded processing modules, wireless communication pathways, audio generation components, and power management features into a compact wearable device. Each layer of the architecture plays a distinct but interconnected role to ensure that gesture recognition is accurate, fast, reliable, and suitable for everyday usage. The architecture follows a structured pipeline where sensor data is captured, processed, interpreted by a machine learning model, and finally converted into speech that can be heard by non signers.

The sensing layer forms the foundation of the gesture capture mechanism. Each glove is equipped with five flex sensors strategically placed along the fingers to measure bending or straightening movement. These sensors work on the principle of resistive variation, where resistance increases proportionally with the degree of bending. This property makes them ideal for detecting individual finger positions, which are essential for identifying static gestures and hand shapes used in Indian Sign Language. Whether the gesture involves curling the index finger, spreading the palm, or bending multiple fingers simultaneously, the flex sensors capture these fine articulations very accurately.In addition to flex sensors, each glove integrates two inertial measurement units. These IMUs, commonly represented by modules such as the MPU6050, include a three axis accelerometer and a three axis gyroscope. Together, these provide six axis motion data that describes hand orientation, tilt, rotation, and dynamic movement. While flex sensors capture finger level articulation, the IMUs capture entire hand movements across space, such as upward flicks, circular motions, forward pushes, and side sweeps. This dual sensing arrangement ensures that both static gestures and dynamic gestures are recorded with temporal and spatial precision. The combination of these sensors allows the system to distinguish between gestures that may look similar in finger shape but differ in motion, which is a critical requirement for authentic ISL gesture interpretation.

At the heart of the architecture lies the ESP32-S3 microcontroller, which serves as the central processing unit. This device includes a dual core processor and enhanced memory capacity, making it suitable for computational tasks such as running embedded neural network models. The ESP32-S3 offers integrated Wi Fi and Bluetooth functionality, supports low power operation, and includes sufficient storage for loading TensorFlow Lite Micro models. Its advanced ADC channels allow high resolution sampling of flex sensor outputs, while the I2C communication interface ensures synchronized and stable data acquisition from both IMUs. Raw sensor signals undergo a preprocessing pipeline within the microcontroller. This includes noise filtering, normalization, and temporal windowing to prepare data for the classification model. Filtering removes unwanted electrical noise and ensures that minor signal variations do not affect gesture interpretation. Normalization scales all sensor inputs to a uniform range, improving the consistency of model predictions. Sliding window segmentation organizes continuous data streams into structured samples that capture temporal change over short durations. The processed data is then fed to a hybrid deep learning model built using a combination of convolutional neural networks and long short term memory layers. The CNN layers extract spatial features from flex sensor and motion patterns, while

the LSTM layers identify temporal sequences and transitions. The model is deployed in quantized form through TensorFlow Lite Micro, enabling it to run efficiently within the limited memory of the microcontroller. This setup achieves fast inference speeds, allowing the system to maintain an end to end latency of around 100 milliseconds, which is sufficiently fast for natural human communication.

Once a gesture is recognized, the system uses its communication layer to relay the classified output to the audio subsystem. The architecture utilizes Bluetooth communication, specifically Bluetooth Low Energy, to ensure power efficient and reliable wireless transmission of recognized words. Unlike earlier systems that require a smartphone or computer to process or relay audio output, this architecture allows the microcontroller to connect directly to a portable Bluetooth speaker or headset. This direct communication minimizes hardware dependencies and simplifies the user's experience. The user only needs to turn on the glove and the speaker, after which the system automatically maintains the connection and sends recognized speech output seamlessly.

The audio layer is responsible for transforming recognized gestures into spoken English. This layer incorporates a text to speech engine that converts the gesture classification output into human understandable speech. The microcontroller encodes text strings into audio waveform instructions which are then streamed to the Bluetooth speaker. This process occurs in near real time so that speech output is synchronized closely with the gesture performed by the user. The audio subsystem ensures clarity, correct pronunciation, and natural sounding voice output. It also manages packet transmission, audio buffering, and synchronization to maintain smooth speech delivery without interruptions. In practical use, this allows fluent communication in social, educational, medical, and workplace environments, enabling ISL users to interact effortlessly with individuals who do not understand sign language.

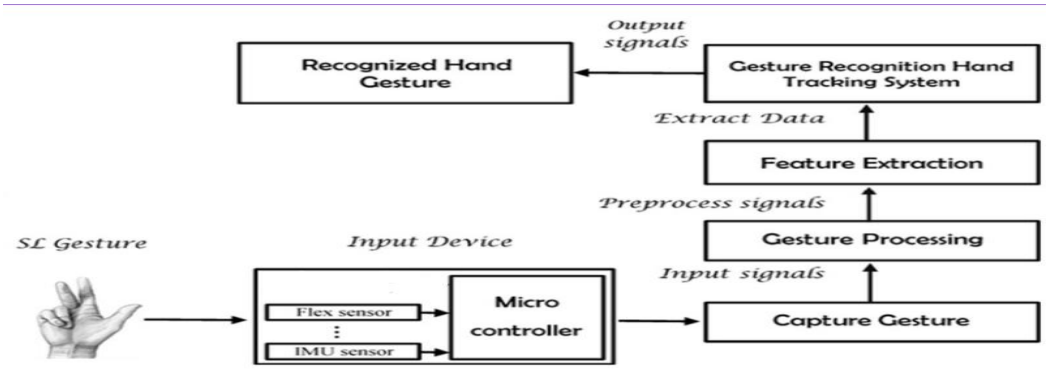

Fig.1 Block diagram.

RESULTS AND DISCUSSION

The initial stage of development focuses on constructing and validating a simplified prototype of the sign language recognition glove. This first-stage version uses a single glove equipped with five flex sensors and one inertial measurement unit and is designed to recognize only five basic gestures. This controlled and minimal setup enables reliable experimentation, stable data collection, and early verification of the system's sensing and classification logic before moving into more advanced multi-glove, machine learning, and vocabulary-expansion phases. The primary objective of the first-stage prototype is to establish a fully functional pipeline from sensing to gesture detection and Bluetooth communication while maintaining low complexity, predictable behavior, and ease of debugging.At this stage, the glove is embedded with five flex sensors, each attached to one finger: thumb, index, middle, ring, and little finger. These sensors provide analog values that change according to the curvature of each finger, allowing the system to capture distinct finger postures associated with basic Indian Sign Language gestures. In addition to the flex sensors, the glove incorporates a single MPU6050 module that provides acceleration and rotational motion readings along three axes each. Although the initial prototype relies mainly on flex data for classification, the IMU still contributes essential movement information for gestures that require a characteristic wrist motion, such as the signs for "YES," "NO," or "THANK YOU." Using only one glove drastically simplifies wiring, calibration, and interpretation while helping to isolate and refine the core sensing mechanism that will later support more sophisticated gestures.

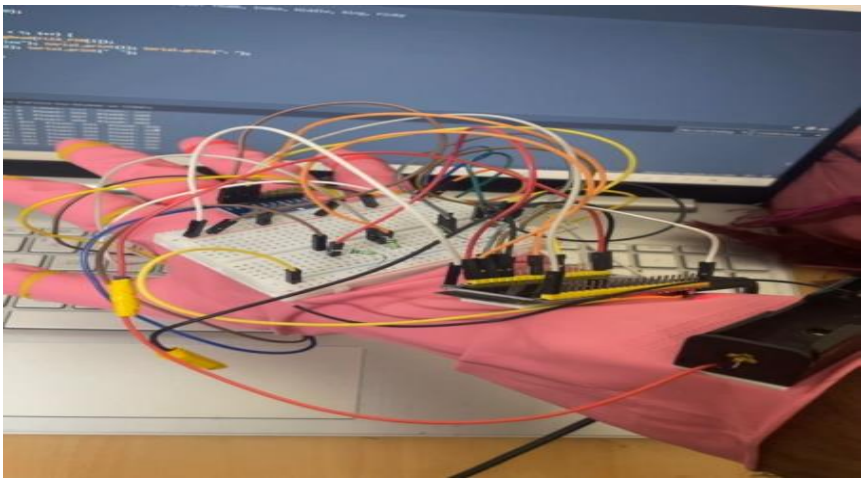

Fig.2 Prototype system.

The software for this prototype uses straightforward threshold-based logic instead of machine learning. Each gesture is defined as a unique pattern of finger bending combined, in some cases, with IMU movement values. For example, the sign for “HELLO” is detected when three fingers are extended and the remaining two remain bent, while “THANK YOU” is recognized when all fingers are extended and there is a slight wrist rotation. The gestures for “YES” and “NO” are intentionally designed to include distinct IMU-based motion patterns: a nodding-like motion for “YES” and a sideways movement for “NO.” The sign “HELP,” on the other hand, is characterized by an open palm combined with an upward acceleration spike. This threshold-based classifier ensures predictable and transparent operation, making it ideal for early testing where adjusting sensor thresholds and observing gesture boundaries are essential.

A significant benefit of this first-stage design is its simplicity, which allows the development team to focus on sensor behavior, noise handling, and gesture reproducibility. Through extensive trial and observation, the flex sensor thresholds are fine-tuned to accommodate natural differences in finger strength, glove tightness, and user comfort. The IMU values are also monitored to understand the typical range of wrist motions produced during intentional gestures versus unintentional movements. During early testing, the debug output provides continuous visual feedback through the serial monitor, showing raw flex sensor outputs, IMU measurements, and the currently detected gesture. This diagnostic tool is invaluable for understanding how different sensor readings interact and how reliably each gesture can be distinguished during real use. Once the glove can reliably detect the five basic signs, the Bluetooth transmission module ensures that each recognized gesture is sent wirelessly to a paired device. The encoded message includes both the gesture ID and the associated text phrase, which allows an external speaker or processing unit to generate audible speech. This final step validates the complete sensing and communication chain and demonstrates the glove’s practical utility in real-time interaction. The successful wireless transmission of recognized gestures proves that the glove can operate independently, without requiring a wired connection or external processing hardware, and sets the stage for future high-level integration of text-to-speech systems.

Fig.3 Prototype-Hand glove

The first-stage prototype also allows evaluation of ergonomic and structural considerations. With only one glove and minimal components, the system is lightweight and easy for users to wear, making it ideal for extended testing. The flexibility of the glove material, placement of wiring, and stability of sensor attachment are all examined during this phase. Observations from user tests help the development team identify comfort issues, sensor drift patterns, and potential design improvements that will be essential once more sensors, dual gloves, and additional electronics are integrated. Overall, this initial version of the glove provides a controlled and practical foundation for developing a fully scalable sign language recognition system. It confirms that the sensor arrangement, data acquisition routines, threshold-based classifier, and Bluetooth transmission pipeline all function reliably in real time. By validating these core elements at a small scale, the project ensures that future stages such as expanding gesture vocabulary, using machine learning models, and implementing dual-hand coordination are built upon a stable, tested, and well-understood baseline. The first-stage prototype therefore plays a critical role in shaping the subsequent design phases and reducing the complexity of future troubleshooting and system expansion.

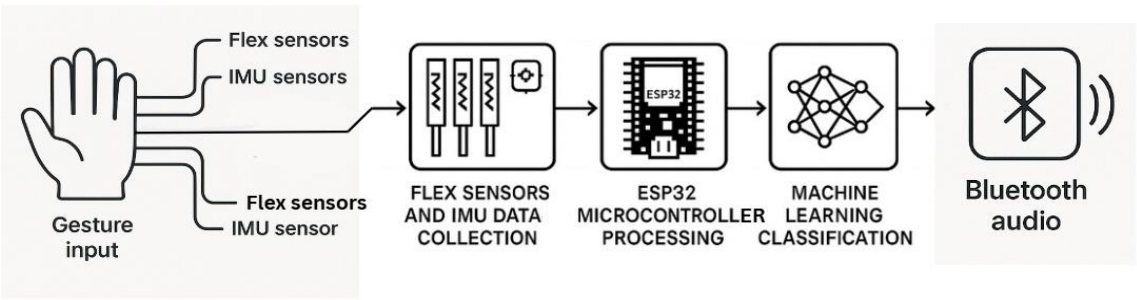

Fig.4 Gesture Input.

The working principle of the sign language to voice conversion glove is based on the integration of multiple sensing mechanisms, embedded processing, deterministic gesture classification, and wireless communication to convert Indian Sign Language gestures into meaningful spoken output. The glove functions as an intelligent wearable device that continuously monitors finger bending and hand movement, interprets this information using predefined logic, and transmits the interpreted gesture to a Bluetooth-enabled audio device. The entire workflow is designed to operate in real time, ensuring that hand movements are translated into spoken English with minimal delay. This section explains the complete operational sequence of the glove beginning from sensor activation to gesture detection and ending with wireless transmission of the recognized sign.

The system begins functioning as soon as the microcontroller is powered on. The initialization phase activates all onboard communication and sensing modules, including the serial interface for debugging, the HC05 Bluetooth module for data transmission, and the MPU6050 inertial measurement unit that monitors motion. The microcontroller then configures all five flex sensor pins as analog inputs so that bending data can be acquired continuously. A connection test is performed on the IMU to ensure that valid accelerometer and gyroscope data will be available during operation. If this test is unsuccessful, the glove enters an idle state and alerts the user through a serial message. Once initialization is complete, the glove enters its continuous monitoring mode, where the main loop executes repeatedly at a controlled sampling rate. During each iteration of the loop, the glove first collects real time bending information from all five flex sensors. These sensors act as variable resistors whose output voltage changes depending on the amount of curvature in each finger. The thumb, index, middle, ring, and little fingers each have their own sensor, allowing the system to capture hand configurations with high granularity. The analog readings from the sensors are stored in a dedicated array that is updated several times per second. This enables the glove to track both static hand shapes and transitions between gestures. Finger bending information forms the core of many sign language gestures, making these flex sensors primary contributors to gesture recognition.

Alongside flex sensor readings, the glove simultaneously collects motion and orientation data from the MPU6050 IMU. This sensor provides raw accelerometer values indicating linear movement along three axes and gyroscope values indicating rotational velocity around three axes. The raw readings are converted into standard units such as g-forces for acceleration and

degrees-per-second for angular rotation. These processed motion values allow the glove to detect wrist movements that accompany certain signs. For example, gestures such as “YES” and “NO” involve distinct dynamic wrist motions that cannot be recognized through finger bending alone. The IMU therefore enhances the glove’s ability to distinguish between gestures that would otherwise appear visually similar or identical in static form.

Once flex and IMU data are collected, the gesture recognition engine evaluates the sensor patterns to determine whether the user is performing any of the defined signs. In this first-stage prototype, the engine uses threshold-based logic, where each gesture is associated with a specific combination of bending levels and motion characteristics. A gesture is recognized when the real-time sensor values match the predefined criteria for that gesture. For example, the sign for “HELLO” is identified when the thumb, index, and middle fingers are extended beyond their thresholds while the ring and little fingers remain bent. The sign for “THANK YOU” requires that all fingers be extended and supported by a slight rotational motion of the wrist. Similar logic is applied to detect “YES,” which includes a nodding-like up-and-down hand motion captured through gyroscope data, and “NO,” which involves a lateral side-to-side motion. The gesture “HELP” is recognized when the palm is open and an upward acceleration spike is detected through the accelerometer. This threshold-based approach provides transparent and predictable classification behavior, making it ideal for an early-stage prototype. Once a gesture is identified, the system compares it with the previously detected gesture to avoid redundant or repeated outputs. If the gesture has changed, the glove constructs a formatted message that includes both the gesture ID and the corresponding word mapped from the gesture library. This message is sent wirelessly to the paired Bluetooth module using a predefined communication protocol. The Bluetooth device handles the transmission in a simple and reliable manner, ensuring that each recognized sign results in a corresponding audio output. The external speaker, mobile device, or audio unit that receives the transmitted message generates the final speech output, effectively converting the user’s sign language gesture into spoken English. The real time nature of this process ensures that communication remains fluid and natural for both the signer and the listener.

Throughout operation, the glove also produces diagnostic output through the serial monitor. This debug channel prints flex sensor readings, IMU values, and the currently recognized gesture, allowing continuous monitoring and fine-tuning of thresholds. This feature is crucial for calibrating the glove for different users, as finger strength, hand size, and natural

movement patterns vary. The debugging output helps refine accuracy and ensure that each gesture is detected consistently across multiple trials. The working principle of the first-stage prototype demonstrates how sensor-driven logic can be used effectively to recognize a small vocabulary of sign language gestures. By establishing a reliable pipeline from sensing to classification and wireless communication, the system validates the feasibility of using a low-cost, embedded solution for gesture recognition. This initial functionality forms a strong foundation for advancing toward more sophisticated versions of the glove. Subsequent stages will incorporate machine learning models to replace the threshold-based classifier, expand the vocabulary to hundreds of gestures, and integrate dual-hand sensing for full Indian Sign Language support. Nonetheless, the underlying working principle remains consistent: capturing human intent through hand movement, interpreting it through embedded processing, and enabling communication through real-time speech output.

CONCLUSIONS

The development of the single-glove sign language to voice conversion system marks an important first step toward creating an accessible, low-cost, and user-friendly communication aid for individuals with hearing and speech impairments. This prototype successfully demonstrates the feasibility of using a combination of flex sensors and inertial measurement units to capture finger articulation and hand motion patterns, which together form the fundamental components of Indian Sign Language gestures. By implementing threshold-based logic and a streamlined decision engine, the system is able to reliably distinguish between five essential gestures and transmit their corresponding meanings through a Bluetooth interface to generate spoken output. The complete pipeline spanning sensor acquisition, gesture interpretation, message formatting, and wireless transmission operates in real time and validates the core functional concept of gesture-to-speech conversion on an embedded platform.

Beyond verifying the technical foundation, the prototype offers valuable insights into the practical challenges associated with gesture recognition. It highlights the need for careful sensor calibration, stable mounting of components, and consistent gesture patterns across users. The debugging and diagnostic capabilities incorporated into the system further support refinement by enabling precise monitoring of sensor behavior and recognition accuracy. These learnings establish a strong groundwork for future iterations of the project, where the focus will shift toward replacing threshold logic with machine learning models, improving

reliability, expanding vocabulary, and integrating dual-hand sensing for complete ISL coverage. Overall, the first-stage glove demonstrates a functional, compact, and efficient assistive device that transforms gestures into meaningful speech output. While still in its early form, the system proves that sensor-driven recognition can serve as a viable bridge between signers and non-signers, promoting inclusivity and making everyday interactions more accessible. The success of this prototype sets the direction for more advanced stages of development, ultimately contributing toward a scalable and impactful communication solution.

REFERENCES

1. Natarajan Vijayaraj, Mohan Nalini, Kinol A Mary, R M Bommi, Smart Glove for Impaired People to Convert Sign into Voice with Text, Proceedings of ICCEBS, 2023
2. Zhou Z, Chen K, Li X and co authors, Sign to Speech Translation Using Machine Learning Assisted Stretchable Sensor Arrays, Nature Electronics, volume 3,2020
3. Khan R U, Khattak H, Wong W S, AlSalman H, Mosleh M A A, Mizanur Rahman S M, Intelligent Malaysian Sign Language Translation System Using Convolutional Based Attention Module with Residual Network, Computational Intelligence and Neuroscience, volume 2021, 2021
4. Rami Aldahir, Ronald R Grau, Using Convolutional Neural Networks for Visual Sign Language Recognition Towards a System that Provides Instant Feedback to Learners of Sign Language, Proceedings of the International Web for All Conference, 2024
5. K Shenoy, T Dastane, V Rao, D Vyavaharkar, Real Time Indian Sign Language Recognition, Proceedings of the International Conference on Computing Communication and Networking Technologies, specified, 2018
6. Swaroop Gudi, Chinmay Inamdar, Yash Divate, Sunil Tayde, Sign Language Detection Using Gloves, International Journal for Research in Applied Science and Engineering Technology.
7. Hind Bitar, Ohoud Alzamzami, Dimah Alahmadi, Amal Barsheed, Amal Alghamdi, Hanadi Almshjary, Intelligent Gloves An Information Technology Intervention for Deaf Mute People, Journal of Intelligent Systems, volume 32, issue not specified, 2023
8. Urav Dalal, Aasmi Thadhani, Mahek Upadhye, Shreya Shah, Meera Narvekar, Nilesh Patil, Deep Learning Enabled Smart Glove for Real Time Sign Language Translation, Journal of Electrical Systems, volume 20, issue 10 special issue, 2024

9. P Das, R De, S Paul, M Chowdhury, B Neogi, Analytical Study and Overview on Glove Based Indian Sign Language Interpretation Technique, Michael Faraday IET International Summit Proceedings, 2015
10. Authors from Sangmyung University, Application of Wearable Gloves for Assisted Learning of Sign Language Using Artificial Neural Networks, Processes, volume 11, 2023
11. Authors from Punjab Engineering College, ISL Recognition System Using Integrated Mobile Net and Transfer Learning Method, Expert Systems with Applications, volume 221, 2023
12. Abhishek Tandon, Amit Saxena, Keshav Mehrotra, Khushboo Kashyap, Harmeet Kaur, A Review Paper on Smart Glove Converts Indian Sign Language into Text and Speech, International Journal for Scientific Research and Development, volume 4, issue 8, November 2016.