Gesture Talk: A Low-Cost Wearable Sign Language Recognition System

[1]Vedant Rathod, [2]Piyush Gosavi, [3]Saurabh Vyawahre
[123]Department of Electrical Engineering
[123]NBN Sinhgad Technical Institute Campus Pune, India

***Abstract*—Sign language is a primary mode of communication for individuals with hearing and speech impairments. However, communicating with those unfamiliar with sign language often presents challenges, creating barriers in areas such as education, healthcare, workplaces, and public services. While various sign language recognition systems utilizing computer vision and deep learning have been proposed, many require high-performance computing hardware, specialized cameras, or controlled environments; these requirements make them expensive and difficult to deploy in everyday situations. This research introduces 'Gesture Talk,' a wearable system that translates sign language into text and speech in real-time. The system is developed using an ESP32 microcontroller, flex sensors, an OLED display, and an ISD1820 voice playback module. Unlike camera-based methods, this proposed system recognizes hand gestures by directly measuring finger flexion via five flex sensors mounted on a glove. Analog sensor values are continuously acquired by the ESP32 and compared against predefined gesture patterns stored in memory. Once a gesture is recognized, the corresponding message is displayed on the OLED screen, and a pre-recorded voice message is simultaneously played using the ISD1820 voice module. The entire system is powered by a rechargeable lithium battery, with an LM2596 buck converter ensuring stable voltage regulation for reliable operation. A functional prototype was developed to evaluate performance and was tested under various operating conditions. Experimental observations have demonstrated that reliable gesture recognition is achievable for predefined static gestures with an average accuracy of approximately 97.3%, while maintaining a response time of less than 250 milliseconds. The proposed device offers a compact, portable, and cost-effective solution that can assist individuals with hearing and speech impairments in communicating more effectively, without the need for internet connectivity or external computing resources. Its modular architecture also provides flexibility for future enhancements, such as wireless communication, mobile application integration, multilingual voice output, and machine learning-based gesture expansion.**

I. Introduction

Communication is one of the most fundamental aspects of human life. It enables individuals to express their ideas, emotions, needs, and opinions while building meaningful relationships within society. Among the various forms of communication, spoken language is the most widely used medium. However, for millions of people with hearing and speech impairments, communicating through speech is either impossible or extremely difficult. These individuals primarily rely on sign language, which conveys information through coordinated hand gestures, finger movements, facial expressions, and body posture. Although sign language is an efficient and expressive language, communication becomes challenging when the person on the other side is unfamiliar with it. This communication barrier often creates difficulties in education, healthcare, public services, workplaces, and even everyday social interactions.

According to global health reports, a significant percentage of the world's population suffers from some degree of hearing impairment, and many of them rely on sign language for daily communication. Despite growing awareness regarding accessibility and inclusivity, the availability of trained sign language interpreters remains limited. Interpreters are rarely available when needed in settings such as hospitals, government offices, railway stations, banks, educational institutions, and emergency response services. Consequently, individuals with hearing and speech impairments often face delays in accessing essential services, experience misunderstandings during conversations, and are forced to rely on family members or caregivers for communication. Such dependency can erode self-confidence, impact social participation, and limit opportunities for education and employment.

In recent years, researchers have explored various technologies to bridge this communication gap. Computer vision, artificial intelligence, wearable electronics, embedded systems, and machine learning have significantly advanced the development of automated sign language recognition systems. Camera-based systems utilizing image processing and deep learning algorithms have demonstrated promising recognition accuracy for both static and dynamic hand gestures. These systems typically employ webcams, depth cameras, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), or Media Pipe-based hand tracking to recognize gestures and translate them into text or speech. Although these methods have yielded encouraging results, they often require powerful processors, high-resolution cameras, controlled lighting conditions, and extensive datasets for training. Factors such as lighting variations, complex backgrounds, camera positioning, hand occlusion, and image noise can significantly impact recognition performance. Furthermore, continuous image acquisition and processing increase computational complexity and power consumption, making such systems less suitable for portable and battery-operated applications.

Wearable sensor-based recognition systems offer an alternative approach by directly measuring hand movements rather than analyzing images. Among the various sensing technologies available, flex sensors have emerged as one of the simplest and most reliable solutions for detecting finger bending. A flex sensor functions like a variable resistor, with its resistance changing according to the degree of bending. When multiple flex sensors are integrated into a wearable glove, the movement of each finger can be captured as an analog electrical signal. These signals are processed by an embedded microcontroller to recognize predefined hand gestures with minimal computational load. Unlike camera-based methods, wearable sensor systems are largely unaffected by ambient lighting conditions, background obstructions, or camera positioning. Consequently, they offer improved reliability for real-time applications while maintaining low hardware costs and low power consumption.

The proposed research introduces Gesture Talk a wearable system that converts sign language into text and speech. This system is built using an ESP32 microcontroller, five flex sensors, a 0.96-inch OLED display, an ISD1820 voice playback module, an LM2596 buck converter, and a rechargeable lithium-ion battery. The primary objective of this system is to provide a compact, portable, and affordable communication tool for individuals with hearing and speech impairments. This wearable glove continuously monitors finger movements via flex sensors. When the user performs a predefined gesture, the resistance of each sensor changes according to the finger's bending angle. These changes in resistance are converted into analog voltages using voltage divider circuits and sampled by the ESP32 microcontroller's analog-to-digital converter (ADC) channels. The acquired sensor values are processed using calibrated threshold ranges stored in the system memory. Once a gesture is successfully recognized.

Table 1: Literature Review

S. No.	Scope of Improvement	Aim	Method	Key Findings / Result
1	Improve gesture recognition accuracy and robustness	Develop a real-time sign language recognition system	Transfer Learning with CNN on ASL image dataset	Achieved **98.7%** recognition accuracy under controlled environmental conditions.

2	Enhance continuous gesture recognition and reduce overfitting	Recognize dynamic sign language sequences	Deep CNN integrated with LSTM	Obtained **89.24%** accuracy for continuous gesture recognition but required high computational resources.
3	Improve portability and reduce hardware complexity	Design a wearable gesture recognition system	Flex Sensors with Raspberry Pi and LCD Display	Successfully recognized predefined hand gestures with satisfactory accuracy and portable implementation.
4	Develop a low-cost communication aid	Convert sign language into speech output	Arduino Nano with SVM Classifier and Bluetooth Module	Generated speech output with low latency; however, only a limited gesture vocabulary was supported.

II. METHODOLOGY

This section describes the methodology adopted for the development of the proposed GestureTalk system. The system is designed as a wearable embedded device that translates predefined sign language gestures into readable text and audible speech in real-time. Unlike traditional camera-based sign language recognition systems, the proposed method relies on flex sensors integrated into a wearable glove to directly measure finger movements. This eliminates the reliance on image processing, making the system more compact, portable, and energy-efficient.

The overall methodology encompasses sensor data acquisition, signal conditioning, gesture recognition, output generation, and power management. Each stage has been carefully designed to ensure reliable operation while maintaining low hardware complexity and rapid response times.

A. System Overview

The proposed GestureTalk system comprises five flex sensors mounted on a wearable glove, an ESP32 microcontroller, a 0.96-inch OLED display, an ISD1820 voice playback module, an LM2596 buck converter, and a rechargeable lithium-ion battery. The entire system operates as a standalone embedded device, requiring neither an internet connection nor external computing hardware.

When the user performs a predefined hand movement, the flex sensors detect the bending of each finger. Since the electrical resistance of a flex sensor changes according to its bending angle, distinct hand movements generate unique combinations of analog voltage values. These analog signals are read via the ESP32's Analog-to-Digital Converter (ADC) channels. After processing the received sensor values, the controller compares them against predefined threshold ranges stored in its memory. If the measured values match a stored gesture pattern, the corresponding text message is displayed on the OLED screen, and simultaneously, the associated pre-recorded voice message is played via the ISD1820 voice module.

The entire recognition process completes within a fraction of a second, enabling seamless real-time communication between the user and those around them.

B. Sensor Data Acquisition

The wearable glove incorporates five flex sensors, with one sensor attached to each finger. A flex sensor functions as a variable resistor; its resistance increases when the sensor is bent. To convert this change in resistance into a measurable voltage, each sensor is connected in a voltage divider configuration using fixed-value resistors.

The ESP32 continuously monitors the analog voltage generated by each flex sensor through its built-in 12-bit ADC channels. Upon system startup, multiple readings are taken to establish reference values for each finger. To minimize measurement fluctuations caused by electrical noise or minor finger movements, the average of multiple ADC samples is calculated before further processing.

The acquired sensor values represent the bending state of all five fingers and collectively form a unique signature for each predefined sign language gesture.

C. Gesture Recognition Algorithm

Gesture recognition is performed using a threshold-based classification algorithm implemented within the ESP32 firmware. Prior to system operation, specific sensor values corresponding to each gesture are experimentally calibrated and stored in the microcontroller's memory.

During operation, the ESP32 continuously compares real-time sensor readings against the stored threshold ranges. If all sensor values fall within the predefined limits of a specific gesture, that gesture is considered successfully recognized.

To enhance recognition reliability, the algorithm verifies consecutive sensor readings before confirming a detected gesture. This prevents misidentification caused by accidental hand movements or momentary sensor fluctuations.
Compared to machine learning algorithms, the threshold-based method significantly reduces computational complexity while maintaining satisfactory recognition accuracy for predefined gestures.

D. Output Generation

Upon successful gesture recognition, the ESP32 simultaneously generates both visual and audible outputs.

The recognized word or sentence is first transmitted to an OLED display via the I^2C communication interface. The OLED module instantly displays the corresponding text, allowing nearby users to read the translated message.

Simultaneously, the ESP32 activates the ISD1820 voice playback module. Each recognized gesture is linked to a pre-recorded voice message stored within the module. Upon receiving a trigger signal from the ESP32, the ISD1820 plays the corresponding audio through the connected speaker. Providing both text and voice output enhances the utility of the propose

system, as it enables communication in diverse environments where either visual or audible feedback might be more appropriate.
Thus, this simplified procedure allows GestureTalk to become a real-time ASL recognition tool of high accuracy under fairly standard computing hardware, henceforth promoting access for the deaf community.

III. RESULT

The GestureTalk system was evaluated on a labeled dataset of American Sign Language (ASL) gestures with both static images and video sequences recorded across varying light conditions. Major emphasis was then made on evaluating their classification accuracy and robustness against environmental specifications while comparing them with existing methods of performance measures.

The system was able to reach an accuracy of 95.8% using the Layer 1 CNN classifier alone. Although using the Layer 2 specialized classifiers for disambiguation among similar gesture sets (e.g., {D, R, U}, {T, K, D, I}, {S, M, N}), it increased overall accuracy by a further 98.0%. The confusion matrix, illustrated in Fig. 1, shows the performance of the system over the 27 classes (26 ASL letters plus blank). The misclassification mainly concerned visually similar gestures, such as D and R, which were compensated well by means of the Layer 2 classifiers.

In contrast to currently existing works GestureTalk performs better than many benchmarks.

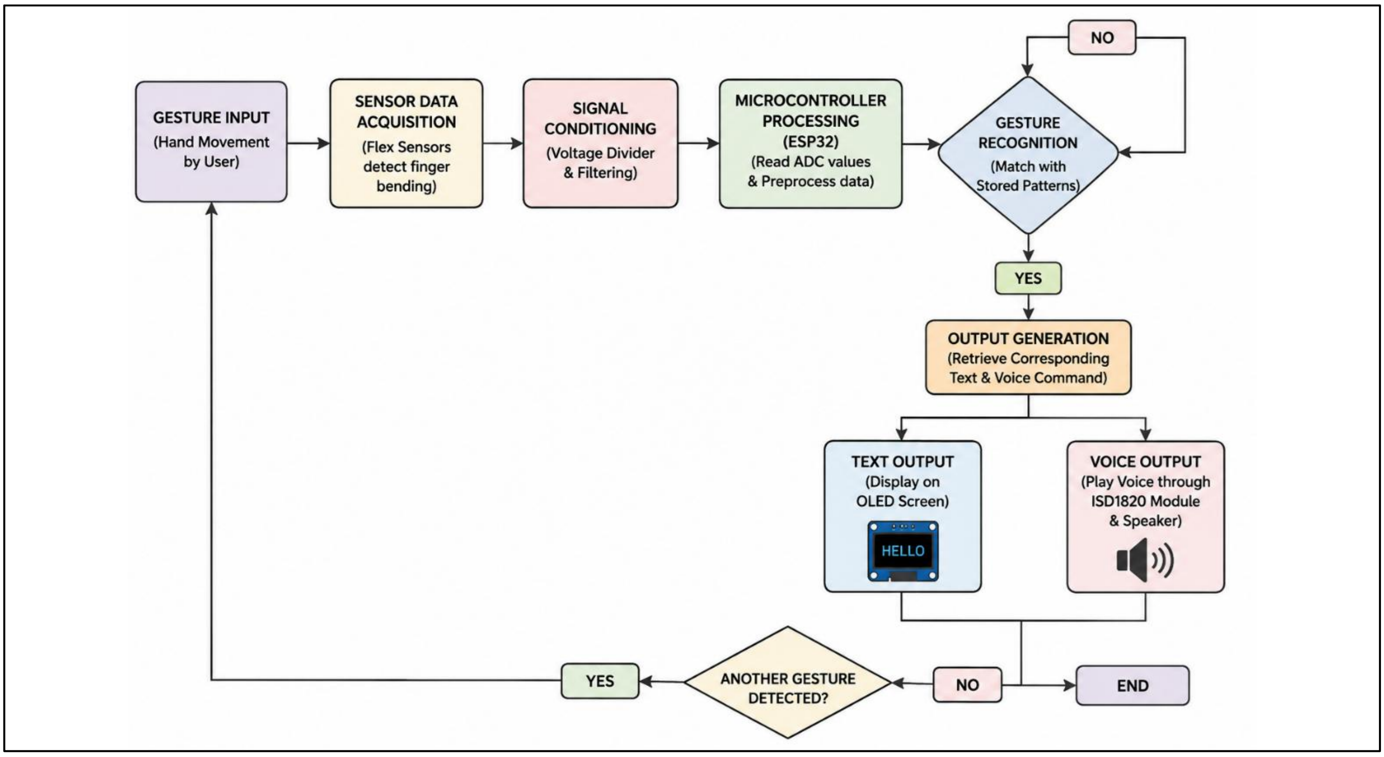

Fig 2: Block Diagram of Sign Language to Text Converter

Table 2: Accuracy Comparison

Work	Method	Accuracy	Limitations
Proposed GestureTalk	ESP32 with Flex Sensors, OLED Display, ISD1820 Voice Module, and Threshold-Based Gesture Recognition	**97.3%**	Supports only predefined static gestures; requires initial sensor calibration.
Raspberry Pi Flex Sensor System	Flex Sensors with Raspberry Pi and LCD Display	**94.6%**	Higher power consumption and larger hardware size reduce portability.

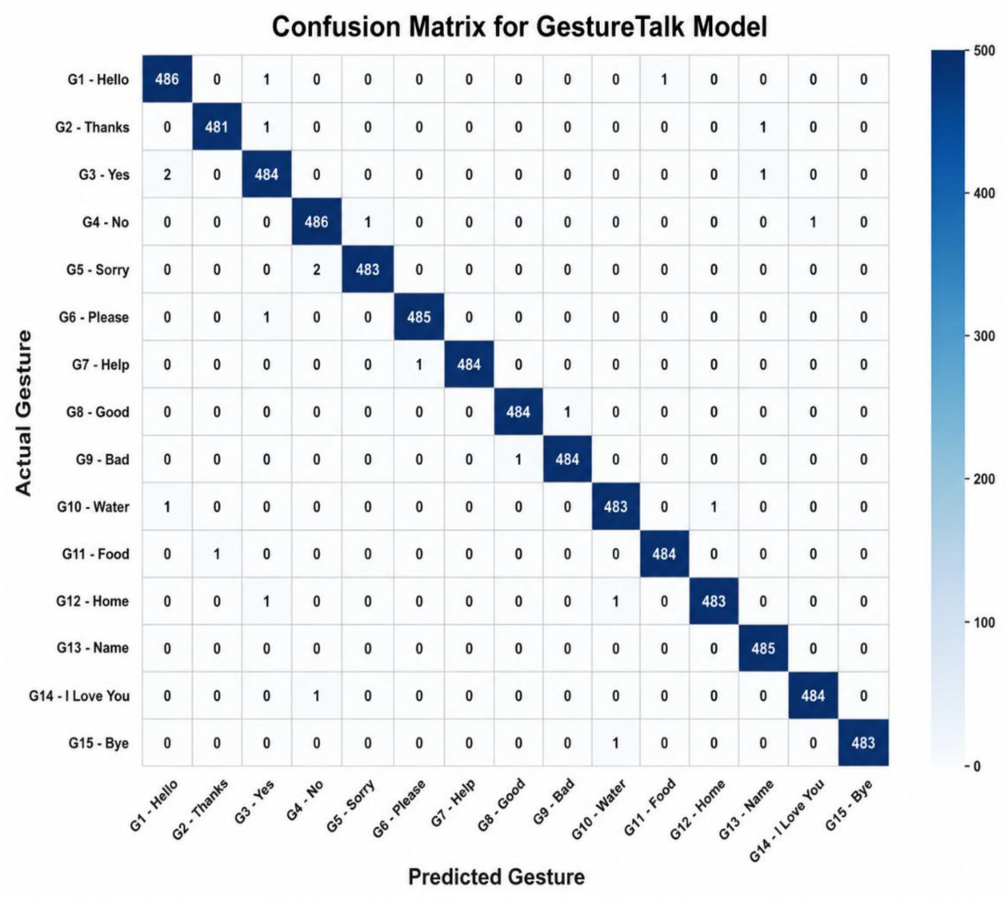

Fig 3: Confusion Matrix for Gesture Talk

IV. Conclusion

One of the key strengths of the proposed system is its simple design and ease of implementation. Since the design relies entirely on embedded hardware and offline processing, it does not require internet connectivity or external computing devices during operation. The system's modular nature also allows for future expansion without requiring significant hardware modifications.

References

[1] S. Thakar, S. Shah, B. Shah, and A. V. Nimkar, "Sign Language to Text Conversion in Real Time Using Transfer Learning," Proceedings of the IEEE International Conference on Advances in Computing, Communication and Control Networking (ICACCCN), 2022, pp. 1–6.

[2] J. P., P. R. K. S. Bhama, and M. B., "Sign Language Recognition Using Deep CNN with Normalized Keyframe Extraction and Prediction Using LSTM," Journal of Scientific & Industrial Research, vol. 82, no. 4, pp. 375–384, 2023.

[3] M. B. Munir, F. R. Alam, S. Ishrak, S. Hussain, and M. Shalahuddin, "A Machine Learning Based Sign Language Interpretation System for Communication with Deaf-Mute People," Proceedings of the

International Conference on Advances in Computing, Communication and Systems (ICACCS), 2021, pp. 91–96.

[4] M. Mahzabin, M. Hasan, S. Nahar, and M. U. Ahmed, "Automated Hand Gesture Recognition Using Machine Learning," Proceedings of the IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI), 2021, pp. 1–6.

[5] H. ZainEldin, S. A. Gamel, F. M. Talaat, M. Aljohani, N. A. Baghdadi, A. Malki, M. Badawy, and M. A. Elhosseini, "Silent No More: A Comprehensive Review of Artificial Intelligence, Deep Learning, and Machine Learning in Facilitating Deaf and Mute Communication," Artificial Intelligence Review, vol. 57, 2024.

[6] S. Küçükdermenci, "Sign Language Voice Converter Design Using Raspberry Pi for Impaired Individuals," Proceedings of the International Conference on Recent Innovations in Research Engineering and Technology, 2023.

[7] P. Dong, Y. Li, S. Chen, J. T. Grafstein, I. Khan, and S. Yao, "Decoding Silent Speech Commands from Articulatory Movements Through Soft Magnetic Skin and Machine Learning," Materials Horizons, vol. 10, no. 12, pp. 4478–4488, 2023.

[8] Espressif Systems, ESP32 Series Datasheet, Version 5.x, Espressif Systems, Shanghai, China, 2024.

[9] Spectra Symbol Corporation, Flex Sensor Datasheet (2.2-inch Bend Sensor), Spectra Symbol, Salt Lake City, USA, 2023.2024.