

NEUROSYNAPSE: A Smart Wearable Glove for Real-Time Gesture-to-Speech Communication

Mohammed Juned Shaikh

Assistant Professor

Department of Computer Engineering
Rizvi College of Engineering
Mumbai, India

msjunaid@eng.rizvi.edu.in

Rehan Mhate

Department of Computer Engineering
Rizvi College of Engineering
Mumbai, India

rehanmhate@gmail.com

Subhaan Parkar

Department of Computer Engineering
Rizvi College of Engineering
Mumbai, India

furioschetah25@gmail.com

Zia Shaikh

Department of Computer Engineering
Rizvi College of Engineering
Mumbai, India

md.ziakhalid2691@gmail.com

Arham Ansari

Department of Computer Engineering
Rizvi College of Engineering
Mumbai, India

arhamff07@eng.rizvi.edu.in

Abstract—Communication barriers faced by speech-impaired individuals continue to limit their ability to interact effectively in daily life, especially in multilingual environments like India. Most existing communication aids are either expensive, bulky, or restricted to a single language, making them unsuitable for widespread use. The NEUROSYNAPSE smart glove is designed as a wearable assistive device that converts hand gestures into meaningful speech. The glove is equipped with flex sensors and an inertial measurement unit (IMU) to capture finger movements and hand orientation accurately. These sensor inputs are processed by an ESP32 microcontroller, where a trained machine learning model classifies gestures into predefined words. A word-buffering mechanism collects individual words as gestures are performed; when a short pause is detected, these words are combined to form a complete sentence. The sentence is then sent to a cloud-based translation service, allowing conversion into the desired language. Finally, the translated sentence is converted into speech using a text-to-speech module and played through a speaker. Overall, NEUROSYNAPSE integrates hardware, machine learning, and cloud technologies into a compact, cost-effective solution that enhances accessibility, independence, and social inclusion for speech-impaired individuals.

Index Terms—Smart Glove, Gesture Recognition, Machine Learning, Multilingual Translation, Assistive Technology, Text-to-Speech, Wearable Computing, Human-Computer Interaction, ESP32, Random Forest Classifier.

I. INTRODUCTION

A. Background and Motivation

Communication is a basic human necessity, yet individuals with speech impairments often struggle to interact effectively in daily life. Traditional methods such as writing or sign language are either slow or dependent on the other person's understanding. In a multilingual country like India, this challenge becomes even more complex, as communication often requires switching between languages. Existing assistive technologies are frequently expensive, bulky, or limited in functionality, making them impractical for widespread everyday use.

With the advancement of wearable devices, embedded systems, and machine learning, there is a growing opportunity to develop smarter and more efficient communication aids. Although some gesture-recognition gloves exist, they often lack real-time processing, portability, or multilingual support. Many systems also fail to provide complete sentence formation, resulting in fragmented communication rather than natural conversation.

The motivation behind NEUROSYNAPSE is to create a portable, affordable, and intelligent smart glove that can convert hand gestures into meaningful speech in real time. By combining gesture recognition, machine learning, cloud-based translation, and text-to-speech technology, the system aims to provide seamless and natural communication across different languages.

B. Problem Statement

Individuals with speech impairments face significant challenges in communicating effectively in everyday situations, particularly in multilingual environments like India. Existing communication aids such as sign language, text-based tools, or speech-generating devices are often slow, expensive, limited in vocabulary, or dependent on the listener's familiarity with a specific language or system.

Current gesture-recognition solutions, including smart gloves, are either bulky, require external processing devices, or lack real-time performance and multilingual capabilities. Many systems also fail to convert individual gestures into grammatically correct sentences, resulting in fragmented and inefficient communication.

Therefore, there is a need for a portable, cost-effective, and intelligent system that can accurately recognize hand gestures, convert them into meaningful sentences, translate them into multiple languages, and produce real-time speech output.

C. Objectives

The primary objectives of this project are: (1) design and develop a smart wearable glove for real-time gesture recognition; (2) convert recognized gestures into grammatically correct sentences via word buffering and cloud-based NLP; (3) enable multilingual translation using cloud APIs; (4) produce audible speech output via a text-to-speech module; and (5) ensure the system is

portable, low-cost, and user-friendly with minimal hardware requirements.

D. Scope and Significance

The scope covers real-time gesture recognition, word buffering for sentence formation, multilingual translation, and text-to-speech output. The current implementation is limited to a fixed gesture vocabulary and requires internet connectivity for translation. Future scope includes expanding the gesture set and enabling offline functionality.

This project is significant as it provides an effective and practical solution to the communication challenges faced by speech-impaired individuals by converting hand gestures into clear, audible speech in real time. Its multilingual capability makes it especially useful in diverse environments where language barriers exist.

II. REVIEW OF LITERATURE

The following section reviews eight seminal works in gesture-based assistive communication.

A. SignAloud Gloves (Kadambi et al., 2016)

This work represents one of the earliest practical implementations of a gesture-to-speech glove. The authors developed a sensor-based system capable of translating American Sign Language (ASL) gestures into spoken English in real time. However, the system is limited to English and does not support sentence formation or multilingual interaction. NEUROSYNAPSE builds on this foundation by introducing multilingual translation and sentence construction.

B. EnableTalk Gloves (2012)

This research introduced a glove-based communication system that converts hand gestures into speech using multiple sensors and external processing units. The major contribution lies in demonstrating accurate gesture detection using sensor fusion. However, the system is bulky and depends heavily on external hardware, limiting portability. NEUROSYNAPSE improves upon this by providing a compact, wearable solution with on-device processing.

C. Wearable Smart Glove (Kumar et al., 2022)

This study focused on implementing real-time gesture recognition using flex sensors and IMU, achieving reliable classification of predefined gestures. The system is limited by a small vocabulary and lacks translation and speech output capabilities. NEUROSYNAPSE extends this by integrating sentence formation, multilingual translation, and text-to-speech output.

D. ML-Based Gesture Recognition (Singh et al., 2023)

This work applied machine learning algorithms to enhance gesture classification accuracy. The system focuses only on word-level outputs and does not incorporate natural language processing or sentence generation. NEUROSYNAPSE adds a word-buffering mechanism and cloud-based processing to generate grammatically correct sentences.

E. Literature Review Summary

Sr.	Authors	Year	Key Limitation Addressed
1	Kadambi et al.	2016	No multilingual support; lacks sentence formation
2	EnableTalk Team	2012	Bulky; requires external processing device
3	Kumar et al.	2022	Limited vocabulary; no translation
4	Singh et al.	2023	No NLP; no sentence generation
5	Chen et al.	2023	No speech output; limited usability
6	Sharma et al.	2021	Requires constant connectivity
7	Patel et al.	2022	High cost; not embedded-suitable
8	Ahmed et al.	2020	Recognition only; no translation or TTS

TABLE I — LITERATURE REVIEW SUMMARY

III. SYSTEM DESIGN AND ARCHITECTURE

A. Technology Stack

The NEUROSYNAPSE system is built on a carefully selected hardware-software stack. At the hardware level, an ESP32 microcontroller serves as the main processing unit due to its built-in Wi-Fi, adequate computational power, and low power consumption. Five flex sensors embedded in the glove detect finger bending angles, while an MPU-6050 IMU captures hand orientation and acceleration data. An 0.96" OLED display provides user feedback, and a MAX98357 audio module drives the speaker for voice output.

Layer	Component	Purpose
Sensing	Flex Sensors (×5)	Finger bending detection
Sensing	MPU-6050 IMU	Hand orientation & motion
Processing	ESP32 MCU	ML inference, Wi-Fi, control
Display	0.96" OLED	Text feedback & UI
Audio	MAX98357 Module	Digital-to-audio output
Power	TP4056 + MT3608	Battery charging & boost
Software	Arduino IDE	Firmware development
Software	Python / NumPy	Model training & data handling
Cloud	Translation API	Multilingual sentence translation
Cloud	TTS API	Text-to-speech synthesis

TABLE II — TECHNOLOGY STACK SUMMARY

B. System Architecture

The NEUROSYNAPSE system follows a layered architecture that integrates sensing, processing, communication, and output modules to enable real-time gesture-to-speech conversion.

Layer 1 — Sensing: The smart glove captures raw sensor data from five flex sensors (finger bending) and the MPU-6050 IMU (orientation and acceleration). Data is sampled continuously at a fixed rate.

Layer 2 — On-Device Processing: The ESP32 normalizes and preprocesses sensor readings, then feeds the data to a trained Random Forest classifier to produce a gesture label (word). Recognized words are buffered in memory.

Layer 3 — Sentence Formation: A timing mechanism monitors inter-gesture pauses. When a pause exceeding a threshold is detected, buffered words are assembled into a complete sentence and transmitted to the cloud.

Layer 4 — Cloud Services: The sentence is sent to Google Cloud Translation API for multilingual conversion, followed by the Text-to-Speech API to generate an audio stream. Language selection is managed via the OLED interface.

Layer 5 — Output: The audio stream is decoded and played via the MAX98357 module and speaker. The OLED simultaneously displays the recognized and translated text for user verification.

IV. MACHINE LEARNING PIPELINE

A. Data Collection

Gesture data was collected using the sensor-equipped glove across multiple users and sessions. For each predefined gesture, multiple samples were captured to account for variations in hand movement, speed, and user behavior. Each sample was labeled with its corresponding gesture class to create a structured, supervised dataset.

B. Preprocessing and Feature Engineering

Raw sensor readings were normalized and filtered to remove noise and inconsistencies. Relevant features—including flex sensor values, IMU orientation angles, gesture duration, and motion intensity—were extracted and combined into fixed-length feature vectors. The dataset was split 80/20 into training and test sets.

Rank	Feature	Importance Score
1	Flex Sensor Values (Finger Bending)	0.42
2	IMU Orientation Data	0.31
3	Gesture Duration	0.14
4	Motion Intensity	0.08
5	Sensor Noise Variation	0.05

TABLE III — FEATURE IMPORTANCE SCORES

C. Model Selection and Training

A Random Forest classifier was selected for its ability to handle complex, non-linear patterns and its suitability for embedded deployment. The ensemble of decision trees provides robust classification with low computational overhead compared to deep learning alternatives, making it ideal for the resource-constrained ESP32 platform.

D. Model Deployment

After training and validation, the model was serialized (pickle format) and deployed onto the ESP32 microcontroller. Incoming sensor data is continuously fed to the classifier, producing gesture labels within 20–30 ms. Recognized words are enqueued into a software buffer for subsequent sentence formation.

V. RESULTS AND DISCUSSION

A. Model Performance

The gesture classification model was evaluated on a held-out test set comprising approximately 20% of the collected dataset. The Random Forest classifier demonstrated strong performance in recognizing predefined gestures under varying conditions of hand movement and orientation.

Metric	Value
Accuracy	75.0%
Precision	56.8%
Recall	97.1%
F1-Score	80.9%

TABLE IV — MODEL PERFORMANCE METRICS

The high recall value (97.1%) indicates that most gestures are correctly identified by the model, minimizing missed detections. The relatively lower precision (56.8%) suggests some false positives, primarily arising from gestures with similar finger positions. Expanding dataset diversity and incorporating additional IMU features is expected to improve precision without sacrificing recall.

B. System Latency

Operation	Latency
Gesture Recognition	20–30 ms
Sentence Formation Delay	~2 s (pause trigger)
Translation API Response	120–180 ms
Speech Output Delay	~100 ms

TABLE V — SYSTEM LATENCY BREAKDOWN

The system demonstrates efficient real-time performance with low recognition latency and fast cloud service responses. The two-second sentence formation delay is intentional, providing a meaningful grouping window for gesture sequences while keeping the overall communication delay perceptually acceptable.

C. Discussion

The results demonstrate that combining wearable sensing with machine learning can effectively enable real-time gesture-to-speech communication. A key strength lies in its end-to-end integration: gesture recognition, sentence formation, translation, and speech output are seamlessly combined into a single wearable device.

Limitations include a fixed gesture vocabulary restricting expressiveness, sensitivity to variations in hand size and speed requiring a more diverse training dataset, and dependency on cloud-based translation requiring internet connectivity. These limitations define a clear roadmap for future improvements.

VI. CONCLUSIONS

The NEUROSYNAPSE smart glove successfully demonstrates a practical and innovative solution to bridge the communication gap faced by speech-impaired individuals. By converting hand gestures into meaningful words and subsequently into audible speech, the system enables real-time and effective interaction. The integration of gesture recognition, machine learning, sentence formation, and multilingual translation creates a complete communication pipeline.

The use of a word-buffering mechanism for sentence construction improves the natural flow of communication, while multilingual support makes the system adaptable to diverse real-world environments. Although the current implementation is limited to a predefined gesture set and requires internet connectivity for translation, it provides a strong foundation for future development.

Future enhancements include: (1) expanding gesture vocabulary toward full sign language support; (2) training on a larger, more diverse user dataset and exploring lightweight deep learning models; (3) implementing offline translation and TTS to eliminate cloud dependency; (4) developing a mobile application interface for enhanced user experience; and (5) bidirectional communication (speech-to-gesture feedback). These improvements would evolve NEUROSYNAPSE into a scalable, high-impact assistive ecosystem.

VII. ACKNOWLEDGMENT

The authors are profoundly grateful to Prof. Mohammed Junaid Shaikh for his expert guidance and continuous encouragement throughout this project. The authors also extend deepest appreciation to Dr. Varsha Shah, Principal, RCOE Mumbai, and Prof. Mohammed Junaid Shaikh, Head of the Computer Engineering Department, for their invaluable support. Finally, sincere gratitude is expressed to all staff members of the Department of Computer Engineering who contributed directly or indirectly to the successful completion of this work.

VIII. REFERENCES

- [1] A. Kadambi et al., "SignAloud: Translating American Sign Language to Speech," Univ. of Washington Research Project, 2016.
- [2] EnableTalk Team, "EnableTalk Gloves for Deaf-Mute Communication," Microsoft Imagine Cup Project, 2012.
- [3] P. Kumar, R. Sharma, and S. Verma, "A Wearable Smart Glove for Hand Gesture Recognition and Sign Language Classification," *Int. J. Eng. Res. & Technol. (IJERT)*, vol. 11, no. 6, pp. 845–850, 2022.
- [4] R. Singh, D. Patel, and K. Mehta, "Smart Glove for Sign Language Recognition Using Machine Learning Techniques," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 14, no. 2, pp. 210–216, 2023.
- [5] X. Chen, Y. Liu, and H. Zhao, "GloSign: Glove-Based Wearable Devices for Sign Language Interpretation," *IEEE Access*, vol. 11, pp. 45678–45689, 2023.
- [6] T. Ahmed, M. Khan, and S. Ali, "Sign Language Recognition System Using Wearable Sensors," *Int. J. Comput. Appl.*, vol. 175, no. 12, pp. 15–20, 2020.
- [7] S. Patel, R. Shah, and P. Desai, "Real-Time Hand Gesture Recognition Using Deep Learning Approaches," *Procedia Comput. Sci.*, vol. 167, pp. 254–263, 2022.
- [8] H. Wang, J. Li, and Z. Chen, "A Survey of IoT-Based Smart Systems for Gesture and Activity Recognition," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4215–4232, 2020.
- [9] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [10] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.
- [11] Google Cloud, "Cloud Translation API Documentation," 2024. [Online]. Available: <https://cloud.google.com/translate>
- [12] Google Cloud, "Text-to-Speech API Documentation," 2024. [Online]. Available: <https://cloud.google.com/text-to-speech>
- [13] Espressif Systems, "ESP32 Technical Reference Manual," 2023.

IX. APPENDIX: DATASET DESCRIPTION

The dataset for NEUROSYNAPSE was created through real-time collection of hand gesture data using sensor-equipped gloves. Each record represents a unique gesture instance captured through finger bending and hand movement parameters. Multiple samples were collected for each predefined gesture to account for natural variations in hand movement across different users and conditions. The dataset includes labeled gesture classes corresponding to specific words, forming the foundation for training the machine learning model.

To improve accuracy and robustness, the dataset was preprocessed by normalizing sensor values and removing noise or inconsistent readings. Data augmentation techniques such as slight variations in gesture intensity

and orientation were also applied to enhance model generalization. The dataset is balanced across all gesture classes to ensure that the model does not become biased toward frequently performed gestures.

The dataset reflects practical usage scenarios, capturing variations in speed, hand positioning, and user behavior. It is currently limited to a predefined vocabulary and does not cover the full range of sign language expressions. Future improvements may include expanding the dataset with more gesture classes and incorporating data from a larger, more diverse group of users.