JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 02
2026
Research Article
https://doi.org/10.56979/1002/2026/1201

Bridging the Gap: Real-Time American Sign Language Recognition Using a Somatosensory Glove

Rawal Khan[1], Nadia Sultan[1&3]*, and Joddat Fatima[2&3]

[1]Department of Electrical Engineering, Bahria University, Islamabad, Pakistan.
[2]Department of Software Engineering, Bahria University, Islamabad, Pakistan.
[3]Centre of Excellence in Artificial Intelligence (CoE-AI), Bahria University, Islamabad, Pakistan.
*Corresponding Author: Nadia Sultan. Email: nadiaimran.buic@bahria.edu.pk

Received: November 04, 2025 **Accepted:** January 30, 2026

Abstract: Sign Language (SL) is a main language for millions of Deaf and Hard-of-Hearing (DHH) people – yet a huge communication barrier still exists, as almost all hearing people do not know SL. Vision-based SLR methods have come a long way, but they still face problems like illumination variations, background clutter, hand occlusion and privacy issue whereas commercial glove-based devices are often expensive and not as portable. This paper introduces a somatosensory glove-based ASL recognition system with wireless capability, able to recognize both static and dynamic American Sign Language (ASL) gestures by flex and inertial sensing fusion. The data were collected by a wired interface to allow noise-free and high-fidelity signal acquisition. Two custom datasets of 19 gestures including 15 static and 4 dynamic were collected from 16 participants respectively on the order of about 8000–9500 labelled samples. Three machine learning based models, XGBoost, RF and MLP were used to train the gesture classifier. For them, XGBoost obtained the most robust performance, achieving sample-level cross-validated accuracies of 97.6% and 99.2% for static and dynamic gestures, respectively. RF and MLP gave competitive baseline results. The results emphasize the power of low-cost wearable sensing and machine-learning-based classification and provide a viable, privacy-sensitive path to scalable near-real-time ASL recognition systems.

Keywords: American Sign Language (ASL); Wearable Sensors; Somatosensory Glove; Flex and IMU Sensors; Machine Learning; XGBoost

1. Introduction

Sign Language (SL) is the first mean of communication for millions of Deaf and Hard-of-Hearing (DHH) people around the world. Despite its importance, hearing people rarely know Sign Language (SL), resulting in major obstacles in education, health services, work and everyday life communication. Prior work by [1] and [2] highlights that the lack of shared sign language knowledge limits accessible communication, motivating assistive SLR solutions. Human interpreters can assist, but they are costly, limited in quantity and unable to provide live real-time interpretation over an extended period of time as noted by [3]. Motivated by this gap, there is an increasing interest in automatic Sign Language Recognition (SLR) technology that can mitigate and close this communication divide.

Vision-based SLR has made significant advances in recent years with CNN, RNN and Transformer architecture. [4] presented strong results based on a hybrid CNN–Transformer architecture, [5] reviewed advances, challenges, and opportunities in continuous sign language recognition. [4] achieved high accuracy using a hybrid CNN–Transformer architecture for vision-based sign language recognition. Different from successes, vision-based methods do have practical issues in reality: affected by lighting change, different scene backgrounds and occlusions, their high computation complexity and privacy problems (especially for classroom or workplace or public scenario) [4] and [13].

To address these challenges, studies have recently focused on wearable and sensor-based SLR. Gloved and arm-banded systems using flex sensors, accelerometers, gyroscopes, IMUs or sEMG can record finger articulation and hand movement directly, are robust to the environment and provide increased privacy [6,13]. [6] reached 93% static gesture accuracy with a consumer Manus Prime X glove and [7] designed a flex-sensor-based assistive glove for recognizing alphabetic and digit postures in ASL. IMU- and glove-based approaches have also been successful for dynamic gestures with accuracies over 95% [3] and [11].

There's a new generation of sensing technology coming, too. [9] reported TENG gloves that do not require outboard power. [2] invented a haptic feedback glove to support learning of ASL as well as [10] showed a low cost IMU gloves for Tanzanian Sign Language. [11] already introduced GRU models to continuous gesture spotting, where they developed flex sensors gloves. [3] combined IMU and sEMG signals using Myo armbands for real-time word recognition, achieving 99.875% accuracy. [13] reviewed the trade-offs of traditional ML representations like k-NN and DTW, ensemble model systems, and deep learning techniques.

In spite of this progress, some issues are still to be addressed. Many popular datasets are small, homogeneous or only contain either static or dynamic gestures, limiting the downstream generalizability [5,13]. Some research described in the literature deal with only static gestures [7], while some focus only on dynamic behavior patterns [8,9], and there still exist gaps in the availability of combined instances. [11] and [9] indicated that a wide range of hand sizes, gesture types and sampling conditions are almost never considered, while [4] observed that deep learning models remain computationally unaffordable to deploy on power-constrained wearable platforms. Furthermore, commercial gloves, such as those that are of high precision (but typically expensive) and with a small set of gestures [6].

To overcome these challenges, in this work we introduce a new dataset containing 19 American Sign Language (ASL) signs (15 static and 4 dynamic), recorded from 16 participants at a sample rate of 5Hz using an affordable somatosensory glove with integrated flex sensors and an IMU. The dataset captures static handshapes as well as temporal motion patterns in a reproductible and affordable manner.

We compare three different machine-learning classifiers, XGBoost, Random Forest (RF), and a Multilayer per ceptron (MLP), to find out the best model for near-real-time ASL recognition. The experimental results indicate that XGBoost offers the best overall solution, as it obtains high accuracy along with low computational burden and is thus appropriate for deployment on wearable and embedded systems for SLR system.

2. Dataset Description

The dataset was collected from a total of 16 participants (Both male and female). We sought to have differentiable hand sizes and gesture performance styles. The participants were 20, 25 years old and were young adults who could consistently perform the required gestures. Furthermore, such diversity was also critical since different users instinctively execute a gesture with slightly varying finger bending strength, hand motion speed or general orientation. By including participants with different features, it provides a more general dataset and better represents real-world scenarios where people with different demographics may use such a system. The dataset consisted of 19 distinct gestures, divided into 15 static gestures and 4 dynamic gestures.

Static Gestures: A, B, C, D, F, G, H, L, S, U, 1, 2, 4, 5, 7.

These gestures mainly represented a mix of American Sign Language (ASL) alphabets and numbers. Since static gestures rely on finger posture, they were recorded as instantaneous 13-dimensional feature vectors. Specifically, the glove provided five flex sensor readings (one for each finger), three accelerometer signals (ACCX, ACCY, ACCZ), three gyroscope signals (GYRX, GYRY, GYRZ), and two orientation values (pitch and roll). Together, these 13 features captured the essential finger-bending patterns and hand orientation required for static gesture recognition.

Dynamic Gestures: HELLO, YES, NO, LIKE.

These gestures were temporal movements, that made them more complex. The raw signals have been captured with a sliding window of 2 seconds (5 Hz sampling rate) and translated into 110-dimensional feature vectors (Each vector is formed by flattening 10 frames × 11 channels (10×11=110). Dynamic gestures were chosen since they are a subset of signs, which are more frequently present in everyday signing tasks than the static ones.

The dataset was recorded from 16 participants performing 19 ASL gestures, as shown in Fig. 1. Each participant provided 30–35 samples per gesture, which, when combined the number of collected samples for a single gesture was approximately 500–550 across participants. Overall, the dataset contained around 8,000-9,500 samples making it diverse and reliable enough for training and testing.

Figure 1. Real-time data collection from participants using the somatosensory glove during ASL gesture recording. Each participant performed 19 gestures, with signals captured through flex and IMU sensors.

This dataset represents one of the only real-time glove based ASL collections collected under controlled settings that contain static and dynamic gestures. Compared to vision-based datasets massively impacted by lighting variations, background clutter, and hand occlusion, the glove-based approach reports environment-free direct sensor measurements which allow for capturing fine levels of finger motion and hand orientation with high correlation.

The dataset consists of windowed sensor samples collected from multiple participants; however, explicit subject identifiers were not retained during aggregation. As a result, individual samples cannot be uniquely associated with specific users. This design reflects a sample-centric dataset structure commonly used for initial feasibility analysis in wearable sensing systems.

The creation of this dataset is an important contribution to the research field of wearable computing, machine learning, and assistive technology. It has a balanced structure and a variety of features, making it an ideal dataset for model training and performance comparison. In this work, the dataset is used to evaluate three classifiers (i.e. XGBoost, Random Forest (RF), and Multilayer Perceptron (MLP)) and decide the most suitable model for real-time ASL recognition.

3. Methodology

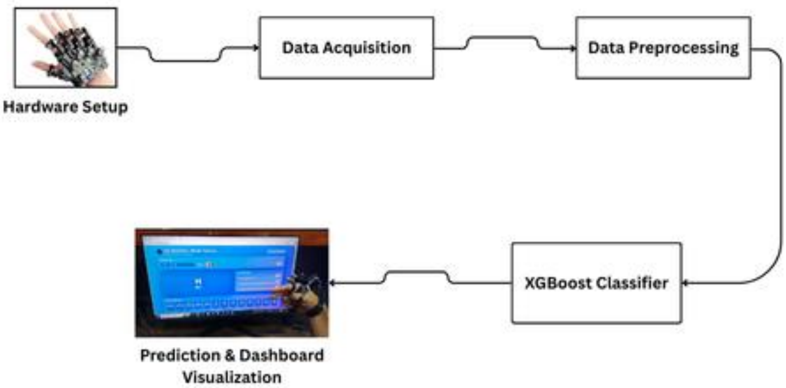

Figure 2. Methodology pipeline showing the end-to-end workflow from sensor-based data collection using the somatosensory glove to preprocessing, XG Boost classification, and final gesture prediction

3.1. Hardware Setup

An end-to-end pipeline, from sensor data collection to pre-processing and classification, is illustrated in Fig. 2. The experimental setup was a modified commercially available somatosensory glove to establish ASL gesture acquisition. Five flex sensors, one for each finger handle, were incorporated in the glove to detect the curling patterns of fingers. Hand movement and orientation sensing were done by using a MPU6050 IMU, with 3-axis accelerometer signal and 3-axis gyroscope signal. An Arduino Nano microcontroller controlled the data acquisition and transmission. Since the glove has a wireless Bluetooth connection through the HC-05 module in its design, we used a wired USB connection for dataset collection, to keep signal transfer as reliable and synchronous as possible. And the receiving end was a laptop, which archived raw sensor streams for further analysis.

3.2. Data Acquisition

The data were recorded with a 5 Hz sampling Rate and from 16 participants. For static gestures, each occurrence was extracted into an instantaneous 13-dimensional feature vector, which is the concatenation of 5 flex sensors and 8 IMU features (ACCX, ACCY, ACCZ, GYROX, GYROY, GYROZ and orientation estimates). For dynamic motions, a sliding window approach was used: each motion was cut into 2s windows with a stride of 1s and this led to approximately 110-dimentional vectors per sample. Each dynamic feature vector was constructed by flattening 10 consecutive time frames × 11 sensor channels, resulting in a 110-dimensional representation. Overlapping windows were used with a stride of 1 second (50% overlap) to increase sample density.

The dataset of 15 static gestures (A, B, C, D, F, G, H, L, S, U,1, 2, 4, 5, 7) and 4 dynamic gestures (HELLO, YES, NO & LIKE). Each participant provided 30–35 samples per gesture, which, when combined the number of collected samples for a single gesture was approximately 500–550 across participants. In total, the data set consisted of around 8000 to 9500 samples. The recorded data were saved in CSV files, each line of which contained feature values and class labels of a gesture sample.

3.3. Preprocessing

To make sure the dataset was clean, uniform, and suitable for training different machine-learning models, several preprocessing steps were performed before classification.

3.3.1. Normalization

Although the dataset was collected from multiple participants, individual samples may still reflect execution-style variations. To mitigate this effect and encourage the models to learn gesture-related patterns rather than participant-specific characteristics, all features were normalized. Feature normalization was performed per cross-validation fold: scaler parameters were fitted exclusively on the training partition and then applied unchanged to the corresponding test partition to avoid information leakage.

3.3.2. Noise Reduction

Then the IMU readings, especially the accelerometer and gyroscope contain... little vibrations or sudden spikes due to minor hand movement or sensor noise. These signals were smoothed with a low-pass filter. Once we were able to eliminate this unwanted jitter, the gesture patterns became much more discernible- and as a result easier for system to interpret.

3.3.3. Segmentation of Dynamic Gestures

A static gesture can be described with a single frame, dynamic gestures are distributed over time. To solve this, we applied a sliding window technique. Every dynamic motion was divided into equal-length windows with a predetermined overlap. This method enabled the system to replicate the short movements as well as whole pattern of every gesture and maintain a homogeneous structure for each gesture.

Window segmentation was performed prior to cross-validation, and fold assignment was applied only after window extraction to ensure that no partial window segments were shared between training and testing splits.

When all the preprocessing was done and fixed to have been in one static gesture, they were represented as 13-dimension feature vectors and in dynamic case, it became 110 dimensions with the multiple time frames. The above steps contributed to a well-arranged as well as uniform dataset that can be adopted to train reliable models including XGBoost, Random Forest and MLP.

3.4. Evaluation Protocol and Leakage Control

Due to the absence of explicit subject identifiers in the dataset, model evaluation was conducted using sample-level stratified 5-fold cross-validation. To minimize information leakage, all preprocessing operations, including normalization and feature scaling, were fitted exclusively on the training subset of each fold and subsequently applied to the corresponding test subset.

Performance metrics are reported as mean ± standard deviation across folds, providing an estimate of model stability and variability.

Inference latency was evaluated on a standard CPU, where the trained XGBoost model required approximately 1.8 ms to process 200 dynamic windows, corresponding to about 0.009 ms per window. This inference time is negligible compared to the 2-second acquisition window, confirming that the system operates in near–real-time.

3.5. Model Selection

In this work, three different machine-learning models were tested for gesture recognition: XGBoost, Random Forest (RF), and a Multilayer Perceptron (MLP). Among these, XGBoost gave the most stable and accurate results, so it was used as the main model for detailed evaluation. The other two models served as comparison baselines.

XGBoost is a boosted-tree algorithm that's particularly good for structured datasets like this, where each gesture is summarized by numerical sensor readings. It deals with non-linear feature interactions, trains in a short time and generally doesn't overfit due to its built-in regularization. These advantages qualify for it as a good candidate for real-time low-cost implementations.

The final set of hyperparameters used for XGBoost is listed below. These values were chosen after several small experiments to find a balance between accuracy and model simplicity.

3.5.1. Learning rate (= 0.1)

Controls how quickly the model learns (weights are updated) with each boosting round. A middle value ensured a consistent training and avoided any surprising change.

3.5.2. Maximum depth (6)

This is the limit on how deep each the trees can grow. A depth of 6 was sufficient to learn relevant patterns without overcomplicating the model.

3.5.3. Number of trees (500)

More trees tend to increase the accuracy, but very high will lead to overfitting. The number of trees around 500 was found to be a good choice for both static and dynamic gestures.

3.5.4. Subsample ratio (0.8)

80% of the training samples were used for each tree and avoided overfitting by introducing randomness into the model.

3.5.5. L2 regularization (= 1)

It decreases the effect of very high weights inside the model, making the range of data feasible to work on and helps in stabilizing the model while noise is generated in IMU or flex sensor readings.

All experiments were conducted with a fixed random seed (e.g., 42) for reproducibility. Hyperparameters were selected through preliminary exploratory trials (e.g., 10–20 configurations) to balance accuracy and model simplicity. XGBoost training used a fixed number of trees (500) without early stopping.

XGBoost exhibited good performance with high accuracy during online inference (both gesture types) and could be implemented in real-time on a wearable glove system.

4. Results

4.1. Static Gesture Recognition

The performance of the system on static gestures remained consistently strong across all three tested models, XGBoost, Random Forest (RF), and MLP. Among them, XGBoost delivered the most reliable results, achieving the highest performance under the 5-fold cross-validation protocol, with consistently high accuracy, precision, recall, and F1 scores across folds. These results show that the model was able to capture fine differences in finger bending and IMU readings across the 15 static gestures, despite variations in gesture execution styles present in the dataset.

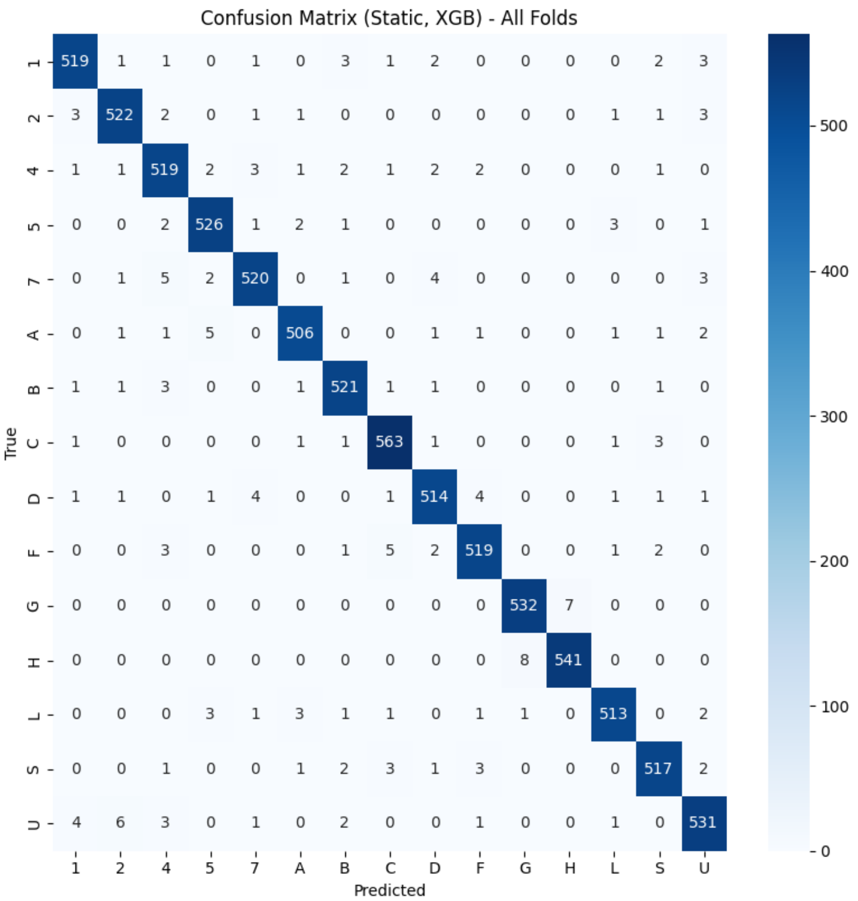

Figure 3. Confusion matrix for static gesture recognition using XGBoost, showing high accuracy with minor errors.

The confusion matrix shown in Fig. 3 illustrates the detailed performance of XGBoost, aggregated across all cross-validation folds. Most gestures were classified correctly with high confidence, but a few minor confusions were noted. The most common mix-ups occurred between A and S, and between 4 and B pairs of gestures that share very similar finger postures and therefore produce nearly overlapping flex sensor values. This pattern suggests that although flex sensors are effective, distinguishing between gestures with almost identical finger shapes can still be challenging.

To provide a broader perspective, RF and MLP were also tested on the same static dataset. Random Forest (RF) and MLP were also evaluated under the same cross-validation protocol, and both achieved strong performance, though with slightly higher variability across folds compared to XGBoost. Although both models performed well, XGBoost remained the most stable and consistent model across all cross-validation folds for static gesture recognition. Despite a few natural confusions caused by similar handshapes, the overall recognition accuracy remained high, demonstrating the effectiveness of the proposed glove-based system for static ASL recognition.

4.2. Dynamic Gesture Recognition

Dynamic gesture recognition produced even stronger results than static gestures. Using XGBoost under a 5-fold cross-validation protocol, the system achieved consistently high accuracy, macro-precision, macro-recall, and macro-F1 scores across folds. These results show that the combination of temporal segmentation and IMU features works very well for gestures that involve motion rather than fixed handshapes.

The confusion matrix in Fig. 4, aggregated across all cross-validation folds, shows that all four dynamic gestures, HELLO, LIKE, NO, and YES, were recognized with very high accuracy. Only one minor error appeared, where a HELLO sample was mistakenly classified as LIKE. This slight overlap is understandable because both gestures begin with forward hand motion, which can produce similar IMU patterns. Apart from this isolated misclassification, the remaining predictions were correct, indicating that the temporal segmentation strategy effectively captured motion information.

To offer a complete comparison, Random Forest (RF) and MLP were also evaluated on the same dynamic dataset. Random Forest (RF) and MLP were also evaluated using the same cross-validation protocol and demonstrated strong performance on dynamic gestures, although with slightly higher variability across folds compared to XGBoost. Although RF performed perfectly on these four gestures, this outcome likely reflects the limited size and high separability of the four-gesture dynamic subset rather than a general advantage of the model. XGBoost showed the most balanced performance across both static and dynamic tasks, which is why it remains the primary model for the system. Overall, the consistently strong cross-validated performance across all three models confirms that IMU-based temporal features are well suited for dynamic ASL gesture recognition.

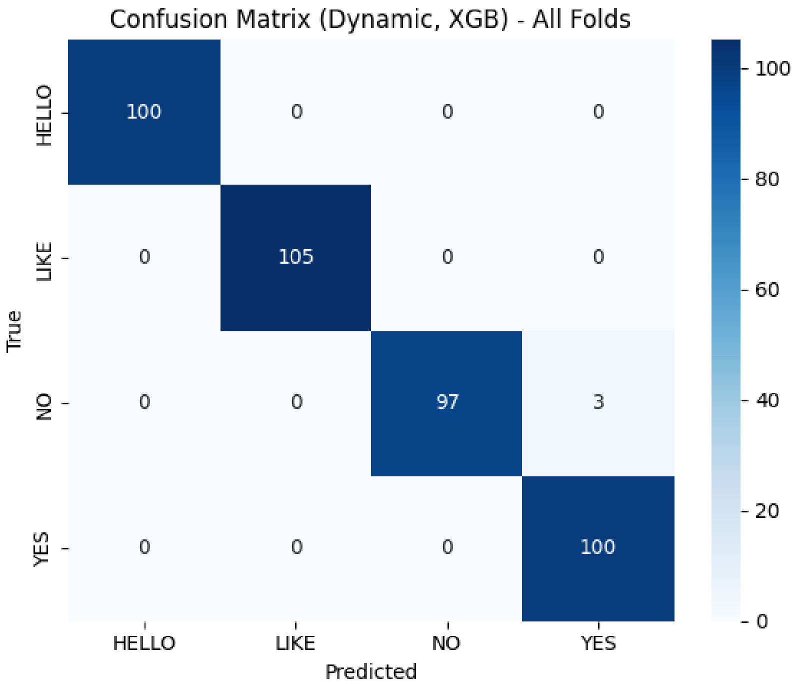

Figure 4. Confusion matrix for dynamic gesture recognition using XGBoost, with minor misclassifications between motion-wise similar gestures.

4.3. Fold-wise Cross-Validation Analysis

To assess the stability and statistical reliability of the proposed system, fold-wise performance metrics were analyzed using a 5-fold cross-validation protocol. For each fold, accuracy, macro-precision, macro-recall, and macro-F1 score were computed independently.

Tables 1 and 2 present the fold-wise results for static and dynamic gesture recognition, respectively, using the best-performing classifier (XGBoost). The results show minimal variation across folds, indicating that the model does not rely on a favorable data partition and exhibits consistent performance across different train–test splits.

The low standard deviation observed across folds further confirms the robustness of the proposed approach and supports the reliability of the reported mean performance values. These findings demonstrate that the system exhibits stable and consistent performance under sample-level cross-validation and is not sensitive to specific data splits.

Table 1. Fold-wise cross-validation performance of XGBoost for Static Gesture Recognition.

	Accuracy	Precision	Recall	F1-Score
Fold 1/5	0.9727	0.9733	0.9726	0.9727
Fold 2/5	0.9814	0.9815	0.9813	0.9813
Fold 3/5	0.9801	0.9803	0.9802	0.9802
Fold 4/5	0.9770	0.9774	0.9768	0.9770
Fold 5/5	0.9733	0.9735	0.9733	0.9732

Table 2. Fold-wise cross-validation performance of XGBoost for Dynamic Gesture Recognition.

	Accuracy	Precision	Recall	F1-Score
Fold 1/5	0.9753	0.9773	0.9750	0.9749
Fold 2/5	1.0000	1.0000	1.0000	1.0000
Fold 3/5	1.0000	1.0000	1.0000	1.0000
Fold 4/5	1.0000	1.0000	1.0000	1.0000
Fold 5/5	0.9877	0.9881	0.9875	0.9875

4.4. Comparative Evaluation of Classifiers

To compare the effectiveness of different learning approaches, three classifiers—XGBoost, Random Forest (RF), and a Multilayer Perceptron (MLP)—were evaluated using a 5-fold cross-validation protocol. The comparison aims to assess not only overall performance but also the stability and consistency of each model across folds for both static and dynamic gesture recognition tasks.

Table 3. Cross-validated (5-fold) performance of all models for static gestures (mean ± std).

	Accuracy	Precision	Recall	F1-Score
XGBoost	0.9769±0.0039	0.9772±0.0038	0.9768±0.0039	0.9769±0.0039
RF	0.9758±0.0037	0.9761±0.0037	0.9757±0.0038	0.9758±0.0038
MLP	0.9568±0.0027	0.9576±0.0028	0.9568±0.0026	0.9568±0.0026

Table 4. Cross-validated (5-fold) performance of all models for Dynamic Gestures (mean ± std).

	Accuracy	Precision	Recall	F1-Score
XGBoost	0.9926±0.0110	0.9931±0.0102	0.9925±0.0112	0.9925±0.0112
RF	0.9975±0.0055	0.9976±0.0053	0.9975±0.0056	0.9975±0.0056
MLP	0.9926±0.0068	0.9930±0.0064	0.9925±0.0068	0.9926±0.0068

Due to the fundamentally different characteristics of static and dynamic gestures, the performance results are reported separately for each task. Table 3 presents the cross-validated performance comparison for static gestures, while Table 4 reports the corresponding results for dynamic gestures, with all metrics expressed as mean ± standard deviation across folds.

For static gesture recognition, XGBoost consistently achieved the most balanced performance across folds, exhibiting high accuracy and low variability compared to the other models. Random Forest also demonstrated strong performance but showed slightly higher variation across folds, while the MLP achieved reasonable results with comparatively lower stability.

In the case of dynamic gesture recognition, all three models achieved strong cross-validated performance, reflecting the effectiveness of temporal segmentation and IMU-based features for motion-driven gestures. While Random Forest and MLP performed competitively, XGBoost maintained the most consistent performance across both static and dynamic tasks, as indicated by its lower variability across folds.

Overall, the comparative analysis confirms that XGBoost offers the best trade-off between accuracy, cross-validated stability, and computational efficiency, making it the most suitable classifier for the proposed ASL glove-based recognition system.

4.5. Overall Analysis

Beyond accuracy and F1-scores, additional analyses were conducted to assess robustness to data variability and to understand how sensor inputs contribute to each prediction. Since XGBoost was the main model used for detailed evaluation, the ROC curves and feature-importance results are reported for this classifier, while RF and MLP were included only in the performance comparison.

The ROC curves in Fig. 5 and Fig. 6, aggregated across all cross-validation folds, show AUC values above 0.99 for all gesture classes, indicating excellent class separation. Even when gestures were performed with slight variations in execution speed or hand orientation, the models maintained strong discrimination between categories. ROC–AUC was computed using a one-vs-rest multi-class strategy based on out-of-fold predicted probabilities. These high AUC values reflect strong ranking performance across decision thresholds under consistent sensing conditions and may coexist with isolated misclassifications observed in the corresponding confusion matrices.

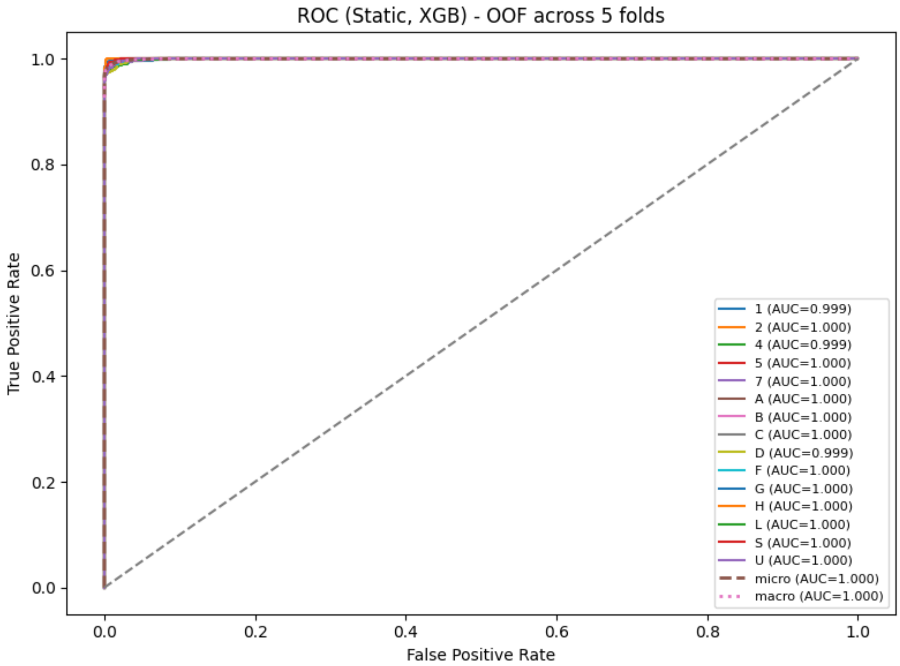

Figure 5. ROC curves for static gestures, with AUC values exceeding 0.99, confirming strong separability across all classes.

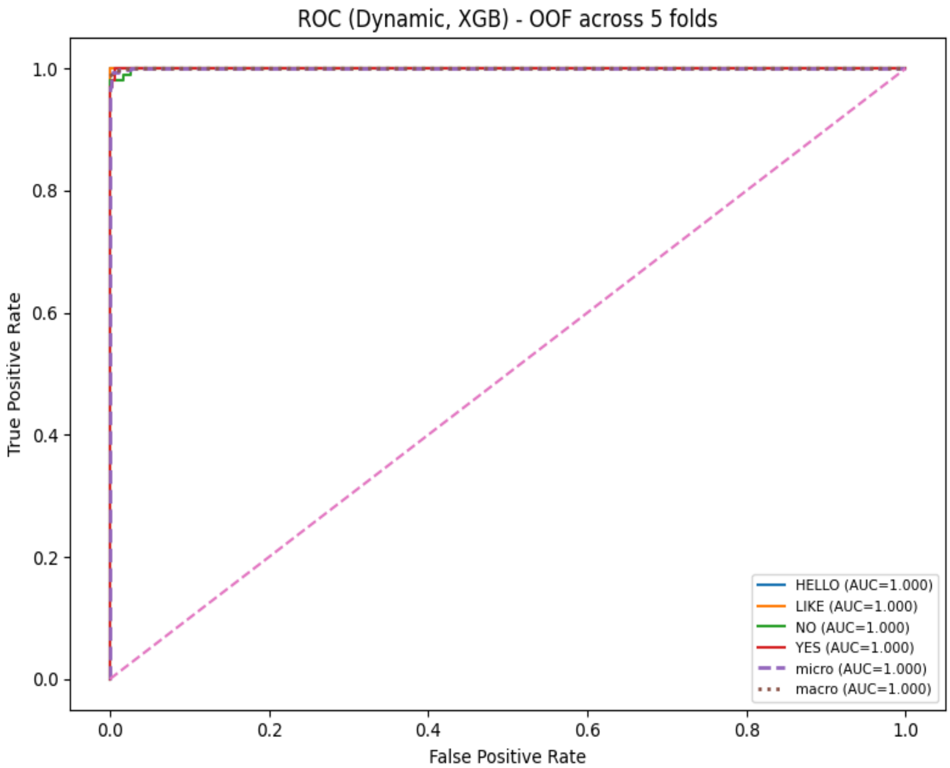

Figure 6. ROC curves for dynamic gestures, with AUC values exceeding 0.99, confirming strong separability across all classes.

The feature-importance plot in Fig. 7 provides a clearer view of how the sensors contributed to the predictions. For static gestures, the XGBoost feature-importance analysis shows that flex sensors dominated the model's decisions, accounting for nearly 87% of the total importance, while IMU readings made up the remaining 13%. This makes sense because static gestures depend mostly on finger posture, which flex sensors capture directly. In the case of dynamic gestures, flex sensors still played the dominant role (around 80%), while IMU signals contributed more than in the static case (approximately 20%). This increase highlights the need for motion information when recognizing temporal gestures such as HELLO or LIKE.

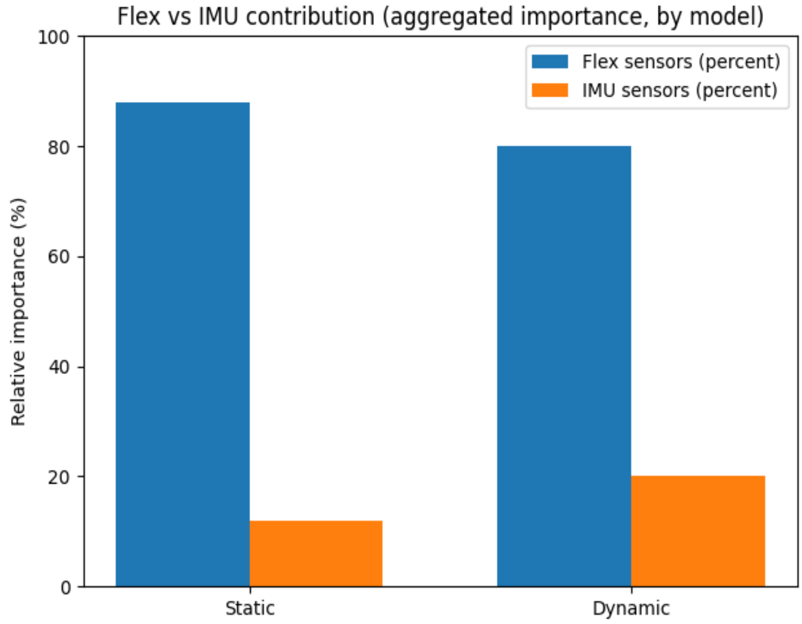

Figure 7. Feature importance graph (flex vs. IMU contributions).

Together, these analyses show how the two-sensing modalities complement each other. Flex sensors pick up detailed finger shapes, while the IMU adds important movement cues. Although RF and MLP achieved strong cross-validated accuracy, XGBoost provides the added advantage of interpretability through ROC and feature-importance analysis, enabling clearer insight into how each sensor modality contributes to the final predictions.

5. Limitations and Future Work

The system performed well for both static and dynamic gestures; however, several limitations should be acknowledged. Firstly, the dataset was collected from a limited number of participants (16), which was sufficient for an initial feasibility study but does not fully capture the diversity encountered in real-world deployment scenarios. Variations in hand size, finger strength, and gesture execution styles may influence system performance. With a larger and more diverse participant pool, the proposed approach is expected to become more robust and representative. In addition, explicit subject identifiers were not retained during dataset aggregation. As a result, strict subject-independent or group-based evaluation could not be performed in the present study. Model evaluation therefore relied on a leakage-aware sample-level cross-validation strategy. This limitation may lead to mildly optimistic performance estimates and will be addressed in future data collection efforts. Moreover, the dynamic gesture set included a limited number of word-level gestures and is therefore intended as a proof-of-concept rather than a comprehensive dynamic ASL vocabulary.

Secondly, minor hardware-related issues were observed during data acquisition. The IMU occasionally produced unstable readings, particularly in the accelerometer and gyroscope channels, which may have introduced low-level noise in some dynamic gesture samples. Although the glove supports Bluetooth communication, data were recorded in wired mode to ensure signal stability; therefore, the wireless capabilities of the system were not fully evaluated in this study.

Future work will focus on expanding the dataset to include 100 or more participants with explicit subject indexing, enabling rigorous subject-independent evaluation. Additionally, more advanced temporal models such as CNN–LSTM and Transformer-based architectures will be explored. Finally, the system will be integrated into a complete ASL-to-text or ASL-to-speech application to enhance its practical applicability in real-world environments.

6. Conclusions

In this work, we developed an ASL gesture dataset collected using an inexpensive somatosensory glove with wireless capability. A total of 9500 samples were obtained from 16 participants for 19 gesture types comprising static handshapes and dynamic motion-based signs. Three machine learning models, XGBoost, Random Forest (RF) and Multi-layer Perceptron (MLP) were evaluated. Among the evaluated models, XGBoost consistently produced the most stable and superior results. Among these, XGBoost demonstrated consistently strong performance, achieving sample-level cross-validated accuracies above 97% for both static and dynamic gestures.

The findings reveal that accurate gesture recognition is feasible with low-cost hardware and a simple sensing configuration. The collected dataset was found to be consistent and informative and may serve as a useful reference baseline for researchers working on sign language gloves and wearable sensing systems.

Overall, this research demonstrates how low-cost sensors combined with machine-learning models can serve to facilitate the development of feasible near-real-time assistive technology. With further improvements, such as increasing the number of participants and exploring deeper sequence-learning models, this approach can be taken one step forward to a practical application like near-real-time ASL-to-text or ASL-to-speech systems that provide a higher level of communication access for Deaf and Hard-of-Hearing.

Data Availability Statement:

The dataset generated and analyzed during the current study is available from the corresponding (1st) author on reasonable request.

References

1. Lu C, Kozakai M, Jing L. Sign language recognition with multimodal sensors and deep learning methods. Electronics. 2023 Nov 29;12(23):4827.
2. Kim HJ, Baek SW. Application of wearable gloves for assisted learning of sign language using artificial neural networks. Processes. 2023 Apr 1;11(4):1065.
3. Umut İ, Kumdereli ÜC. Novel Wearable System to Recognize Sign Language in Real Time. Sensors (Basel, Switzerland). 2024 Jul 16;24(14):4613.
4. Hashi AO, Mohd Hashim SZ, Mirjalili S, Kebande VR, Al-Dhaqm A, Nasser M, A Samah AB. A hybrid CNN-transformer framework optimized by Grey Wolf Algorithm for accurate sign language recognition. Scientific Reports. 2025 Dec 10;15(1):43550.
5. Ibrahim NB, Zayed HH, Selim MM. Advances, challenges and opportunities in continuous sign language recognition. Journal of Engineering and Applied Sciences. 2020;15(5):1205-27.
6. Achenbach P, Laux S, Purdack D, Müller PN, Göbel S. Give me a sign: Using data gloves for static hand-shape recognition. Sensors. 2023 Dec 15;23(24):9847.
7. Amin MS, Rizvi ST, Mazzei A, Anselma L. Assistive data glove for isolated static postures recognition in American sign language using neural network. Electronics. 2023 Apr 18;12(8):1904.
8. Filipowska A, Filipowski W, Raif P, Pieniążek M, Bodak J, Ferst P, Pilarski K, Sieciński S, Doniec RJ, Mieszczanin J, Skwarek E. Machine learning-based gesture recognition glove: design and implementation. Sensors (Basel, Switzerland). 2024 Sep 23;24(18):6157.
9. Kim H, Lee I, Panda S, Hajra S, Jeong B, Seo J, Kaja KR, Belal MA, Vivekananthan V, Kim HJ. Smart gloves-based triboelectric nanogenerator for sign language detection. Micro and Nano Systems Letters. 2025 Jul 22;13(1):11.
10. Bulugu I. Tanzanian sign language recognition system for an assistive communication glove sign tutor based on the inertial sensor fusion control algorithm. Journal of Electrical Systems and Information Technology. 2025 Dec;12(1):1-21.
11. Chuang WC, Hwang WJ, Tai TM, Huang DR, Jhang YJ. Continuous finger gesture recognition based on flex sensors. Sensors. 2019 Sep 15;19(18):3986.
12. Bianco S, Napoletano P, Raimondi A, Rima M. U-wear: User recognition on wearable devices through arm gesture. IEEE transactions on human-machine systems. 2022 May 13;52(4):713-24.
13. Pan M, Tang Y, Li H. State-of-the-art in data gloves: A review of hardware, algorithms, and applications. IEEE Transactions on Instrumentation and Measurement. 2023 Feb 9;72:1-5.