A Robust Real-Time Multimodal Polynomial Fusion Framework for Sensor-Based Sign Language Recognition Using Flex–IMU Smart Gloves

Dadang Iskandar Mulyana, Edi Noersasongko, Guruh Fajar Shidik, Pujiono
Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia

Abstract—**Sign language recognition is a critical component of assistive technologies for individuals with hearing and speech impairments. While vision-based approaches have shown promising performance, their reliability is often affected by illumination variations, occlusions, and background complexity. Wearable sensor–based solutions, particularly smart gloves integrating flex sensors and inertial measurement units (IMUs), provide a more stable alternative by directly capturing hand articulation and motion patterns. However, existing sensor-based methods frequently suffer from temporal instability, noise sensitivity, and limited discrimination among structurally similar gestures, which is especially challenging in Hijaiyah sign language, where many letters differ only by subtle finger configurations. This study proposes a robust real-time Multimodal Polynomial Fusion (MPF) framework for sensor-based sign language recognition using a flex–IMU smart glove, with a specific focus on Hijaiyah gestures as the application domain. The proposed framework applies nonlinear polynomial temporal smoothing within a sliding window to stabilize raw flex–IMU trajectories, followed by multimodal fusion to enhance gesture separability and temporal consistency. A large-scale multimodal dataset comprising 231,000 samples collected from 33 users performing 28 Hijaiyah gesture classes was constructed to enable rigorous subject-independent evaluation. Experimental results obtained from offline testing, session-aware analysis, and real-time streaming scenarios demonstrate that the proposed MPF framework consistently outperforms a baseline approach based on raw normalized signals. The proposed method improves recognition accuracy from 92.42% to 96.32%, while also achieving higher macro-level precision, recall, and F1-score. Furthermore, MPF significantly reduces misclassification rates and improves temporal stability, particularly for fine-grained Hijaiyah gestures with similar structural patterns. These results confirm that the proposed framework provides a robust and reliable solution for real-time wearable sign language recognition and offers practical benefits for Hijaiyah-based assistive communication systems.**

Keywords—Sign language recognition; Hijaiyah sign language; wearable sensors; smart glove; multimodal fusion; polynomial temporal smoothing; real-time recognition

I. Introduction

Sign language recognition has been extensively investigated as a core technology for assistive communication systems that support individuals with hearing and speech impairments. Automatic translation of hand gestures into linguistic symbols enables improved accessibility in education, healthcare, and everyday interactions. Early studies in this area predominantly relied on vision-based approaches, utilizing RGB cameras, depth sensors, or skeletal tracking to model hand shape and motion patterns [1], [3], [11]. With the rapid advancement of machine learning and deep learning techniques, vision-based systems have achieved promising recognition accuracy, particularly for well-studied sign languages such as American Sign Language (ASL) and British Sign Language (BSL) [12], [19], [31].

Despite these advances, vision-based sign language recognition systems remain constrained by several inherent limitations. Their performance is highly sensitive to environmental factors, including illumination variations, background clutter, occlusion, and motion blur, which frequently occur in real-world settings [3], [12], [19]. In addition, camera-based systems typically require fixed viewpoints and considerable computational resources, limiting their suitability for real-time, mobile, and embedded assistive applications [11], [31]. These challenges have motivated an increasing interest in alternative sensing modalities that are less dependent on external visual conditions.

Wearable sensor–based approaches have therefore emerged as a robust alternative for sign language recognition. Smart gloves [2] equipped with flex sensors enable continuous measurement of finger bending, while inertial measurement units (IMUs) capture hand orientation and dynamic motion cues [9], [10], [13], [14]. By directly sensing hand articulation at the source, wearable systems significantly reduce the influence of environmental disturbances and provide more stable gesture representations across diverse operating conditions [4], [18], [22], [32]. Numerous studies have demonstrated the feasibility of glove-based systems for gesture recognition and sign language translation using combinations of flex sensors, accelerometers, and gyroscopes [6], [7], [15], [20], [21], [26].

However, existing wearable sensor–based sign language recognition systems still face several unresolved challenges. Many prior studies rely on relatively small datasets involving limited numbers of participants, typically ranging from three to ten users, which restricts cross-user generalization and robustness [7], [15], [24], [28], [29]. In addition, gesture vocabularies are often limited, with many systems evaluating only partial sets of letters or gestures rather than complete sign alphabets [5], [8], [17], [28]. Furthermore, most preprocessing pipelines employ simple linear filtering techniques, such as moving average or low-pass filters, which are insufficient to model the nonlinear temporal characteristics inherent in flex–IMU sensor signals. As a result, micro-tremors, sensor drift, and temporal instability frequently persist, leading to misclassification, especially for gestures with subtle structural differences [12], [19], [25].

These challenges are particularly evident in Hijaiyah sign language, which represents the Arabic alphabet and consists of 28 gesture classes. Many Hijaiyah letters differ only by minor variations in finger configuration or hand orientation, making them especially sensitive to noise and temporal inconsistencies in sensor measurements [5], [17], [28]. Existing Arabic and Hijaiyah sign language recognition systems are often limited to partial datasets, handcrafted feature extraction, or small-scale user evaluations, leaving large-scale multimodal modeling and real-time robustness insufficiently explored [8], [17], [28].

To address these limitations, more advanced temporal modeling techniques are required to stabilize sensor trajectories while preserving meaningful motion dynamics. Polynomial regression has been shown to be effective for motion smoothing and trajectory stabilization, as it is capable of modeling nonlinear temporal behavior more accurately than conventional linear filters [3], [12], [23]. Nevertheless, the integration of polynomial temporal modeling with multimodal flex–IMU fusion for sign language recognition remains largely unexplored, particularly in the context of real-time wearable systems.

In this study, a robust real-time Multimodal Polynomial Fusion (MPF) framework is proposed for sensor-based sign language recognition using a flex–IMU smart glove, with Hijaiyah gestures serving as the application domain. The proposed framework applies nonlinear polynomial temporal smoothing within a sliding window to stabilize raw flex and IMU signals, followed by multimodal fusion to enhance gesture separability and temporal consistency. Unlike conventional linear preprocessing approaches, the proposed method explicitly models nonlinear sensor dynamics, thereby reducing intra-class variance and improving discrimination among structurally similar gestures.

The effectiveness of the proposed framework is validated using a large-scale multimodal dataset comprising 231,000 samples collected from 33 users, covering all 28 Hijaiyah gesture classes. A subject-independent evaluation protocol is adopted to rigorously assess cross-user generalization. Comprehensive experiments conducted under offline, session-aware, and real-time streaming conditions demonstrate that the proposed MPF framework consistently outperforms a baseline approach based on raw normalized signals in terms of recognition accuracy, robustness, and temporal stability.

The remainder of this study is organized as follows: Section II reviews related work on vision-based and wearable sensor–based sign language recognition. Section III describes the proposed Multimodal Polynomial Fusion framework, including system architecture, preprocessing, and real-time implementation. Experimental results and analysis are presented in Section IV. The discussion is presented in Section V. Finally, Section VI concludes the study and outlines directions for future research.

II. Related Work

Research on sign language recognition has evolved along two dominant paradigms, namely vision-based approaches and wearable sensor–based systems. Each paradigm offers distinct advantages while also exhibiting inherent limitations, particularly when deployed in real-time assistive scenarios.

Vision-based sign language recognition has been extensively explored using RGB cameras, depth sensors, and skeletal tracking models. Early studies employed handcrafted features extracted from hand shape and motion trajectories, while more recent works leverage convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based architectures to improve recognition accuracy [1], [3], [11], [12], [19], [31]. These approaches have demonstrated strong performance for widely studied sign languages, especially ASL, under controlled conditions. However, several studies report that vision-based systems remain highly sensitive to illumination variations, occlusions, background clutter, and camera viewpoints, which significantly degrade performance in unconstrained environments [3], [12], [19]. Moreover, the computational complexity of deep vision models limits their feasibility for real-time embedded deployment, particularly in wearable or mobile assistive systems [11], [31].

To overcome the environmental dependency of camera-based solutions, wearable sensor–based sign language recognition has attracted increasing attention. Smart gloves equipped with flex sensors, accelerometers, gyroscopes, and inertial measurement units (IMUs) enable direct measurement of finger articulation and hand motion, providing more stable gesture acquisition independent of external visual conditions [9], [10], [13], [14]. Early glove-based systems demonstrated the feasibility of translating hand gestures into symbolic representations using flex sensors and basic machine learning classifiers [6], [7], [15]. Subsequent studies integrated IMUs to capture three-dimensional motion cues, improving recognition robustness for dynamic gestures [4], [18], [20], [22], [26], [30], [32].

Despite these advances, many wearable sensor–based systems remain limited by dataset scale and evaluation scope. Numerous studies evaluate their models using small participant cohorts, often involving fewer than ten users, which restricts the assessment of cross-user generalization and robustness [7], [15], [24], [28], [29]. In addition, gesture vocabularies are frequently constrained to partial alphabets or small gesture sets, limiting their applicability to complete sign language systems [5], [8], [17], [28]. These limitations are particularly critical for alphabets with subtle inter-class differences, where robust temporal modeling and extensive user diversity are essential.

Another important challenge lies in the preprocessing and temporal modeling of wearable sensor signals. Most existing works rely on simple linear filtering techniques, such as moving average smoothing or low-pass filters, to reduce sensor noise and jitter [12], [19], [25]. While these methods offer basic noise suppression, they are insufficient for modeling the nonlinear temporal dynamics inherent in flex–IMU signals, including micro-tremors, drift, and execution variability across users. As a result, gesture trajectories often remain unstable, leading to increased intra-class variance and misclassification among structurally similar gestures.

Multimodal fusion has been proposed as a strategy to improve robustness by integrating complementary sensor modalities. Several studies combine flex sensors with IMUs or tactile sensors using feature-level or decision-level fusion, demonstrating improved recognition accuracy compared with single-modality approaches [4], [18], [22], [27], [32]. Classical classifiers such as k-nearest neighbors (k-NN), support vector machines (SVMs), dynamic time warping (DTW), and random forests have been widely used in these systems. Recent deep learning-based glove systems [16] demonstrate higher accuracy but still rely on linear preprocessing. More recent works explore deep learning-based fusion models to improve multimodal integration [18], [19], [27]. However, most fusion strategies still rely on linear preprocessing pipelines and do not explicitly address nonlinear temporal instability in sensor trajectories.

Polynomial regression and related nonlinear modeling techniques have been investigated in motion smoothing and trajectory stabilization tasks, showing superior performance over linear filters in capturing complex temporal patterns [3], [12], [23]. In the context of hand motion and gesture analysis, polynomial modeling has been reported to improve signal stability and reduce high-frequency noise. Nevertheless, the application of polynomial temporal smoothing within a multimodal flex–IMU fusion framework for real-time sign language recognition remains largely unexplored in the existing literature.

Research on Arabic and Hijaiyah sign language recognition is comparatively limited when contrasted with ASL and other widely studied sign languages. Existing works often focus on partial Hijaiyah or Arabic gesture sets, rely on handcrafted features, or evaluate performance using small-scale datasets [5], [8], [17], [28]. Several studies highlight the difficulty of distinguishing Hijaiyah letters that differ only by subtle finger configurations, emphasizing the need for more robust temporal modeling and multimodal integration [5], [17], [28]. To date, no study has reported a large-scale multimodal flex–IMU dataset for Hijaiyah sign language combined with advanced nonlinear temporal fusion evaluated under real-time conditions.

Based on this review, a clear research gap can be identified. Existing vision-based approaches struggle with environmental sensitivity, while wearable sensor–based systems remain constrained by limited datasets, linear temporal modeling, and insufficient real-time validation. Although multimodal fusion improves robustness, the lack of nonlinear temporal stabilization continues to limit performance, particularly for fine-grained alphabets such as Hijaiyah. These limitations motivate the development of a robust real-time framework that integrates nonlinear temporal modeling with multimodal sensor fusion, supported by large-scale subject-independent evaluation. The proposed Multimodal Polynomial Fusion (MPF) framework directly addresses these gaps by stabilizing flex–IMU trajectories through polynomial temporal smoothing and enhancing gesture separability via multimodal integration.

III. Methodology

This section describes the overall framework for real-time Hijaiyah sign recognition using the proposed Multimodal Polynomial Fusion (MPF) approach. The methodology includes hardware design, data acquisition, preprocessing, polynomial temporal smoothing, multimodal fusion, feature extraction, classification, and real-time implementation.

A. Smart-Glove Architecture

The smart glove integrates five flex sensors and a three-axis IMU to capture finger bending and hand motion, producing an eight-channel multimodal signal. As shown in Fig. 1, the pipeline applies preprocessing, polynomial smoothing, multimodal fusion, and classification, with a real-time sliding window ensuring temporal consistency and robustness during continuous gesture recognition.

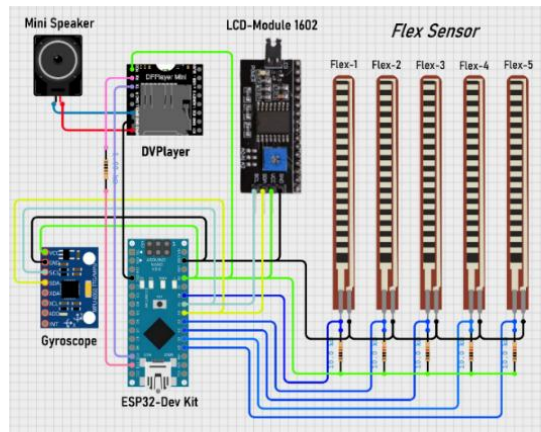


Fig. 1. Proposed multimodal Flex–IMU.

The smart glove integrates five flex sensors and a three-axis IMU to capture finger bending and hand orientation, enabling reliable discrimination of all 28 Hijaiyah gestures, including those with similar postures, as shown in Fig. 2.

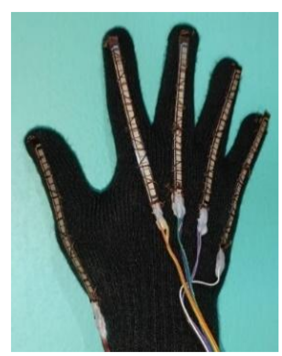


Fig. 2. Sensor placement on the hand.

The assembled smart-glove prototype integrates flex sensors, an IMU, a microcontroller, and a power unit into a wearable form factor, enabling stable signal acquisition during natural hand movements while maintaining user comfort, as shown in Fig. 3.

Fig. 3. Fully assembled smart-glove hardware prototype.

Hijaiyah gesture classes used in this study, representing the complete Hijaiyah alphabet and following established conventions in sensor-based sign language recognition using flex–IMU smart gloves [13], [17], [26], [33], in Fig. 4.

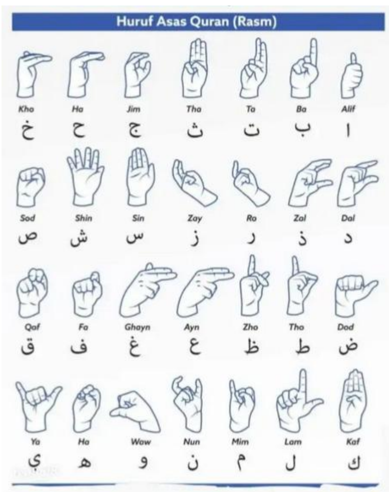


Fig. 4. The 28 Hijaiyah gesture.

B. Data Acquisition

Data collection involved 33 participants, each performing all 28 Hijaiyah gesture classes in 250 repetitions, yielding a total of 231,000 multimodal samples. This large-scale dataset captures natural variations in hand shape, execution speed, and sensor alignment across users. To enable rigorous subject-independent evaluation, the dataset was partitioned into 26 users for training and 7 unseen users for testing, ensuring that model performance reflects true cross-user generalization rather than memorization of individual motion patterns. Each raw observation is represented as:

1) Feature vector

$$S(t) = [f_1(t), f_2(t), f_3(t), f_4(t), f_5(t), x(t), y(t), z(t)]$$

where, $f_1(t)$ – $f_5(t)$ denote flex-sensor values and $a_x(t), a_y(t), a_z(t)$ denote IMU signals.

C. Preprocessing

To mitigate differences arising from hand size, glove fitting, and sensor drift, each sensor channel undergoes z-score normalization:

1) Z-score normalization

$$S'(t) = \frac{S(t) - \mu}{\sigma}$$

where, μ and σ represent the mean and standard deviation computed from the training dataset.

D. Polynomial Temporal Smoothing

Raw flex–IMU signals naturally contain micro-tremors, noise, and nonlinear temporal variations that can degrade gesture separability. To mitigate these effects, polynomial regression is applied within a sliding window of length W, producing smoothed temporal trajectories that better capture the underlying motion patterns while suppressing high-frequency fluctuations in the channel. The smoothed value is computed as:

1) Polynomial smoothing

$$\hat{s}(t) = \sum_{k=0}^{d} a_k t^k$$

where, d denotes the polynomial order and α_k are coefficients estimated using least-squares fitting.

E. Multimodal Polynomial Fusion

Following temporal smoothing, all flex and IMU channels are integrated into a unified multimodal representation:

1) MPF vector

$$F = [\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4, \hat{f}_5, \hat{x}, \hat{y}, \hat{z}]$$

where, $\tilde{f}_i(t)$ and $\tilde{a}_j(t)$ represent the polynomial-smoothed flex and IMU signals, respectively.

The MPF approach strengthens cross-channel modeling and improves discrimination among similar gestures by producing more stable, noise-resistant sensor representations. Fig. 5 highlights these enhancements compared with the baseline pipeline.

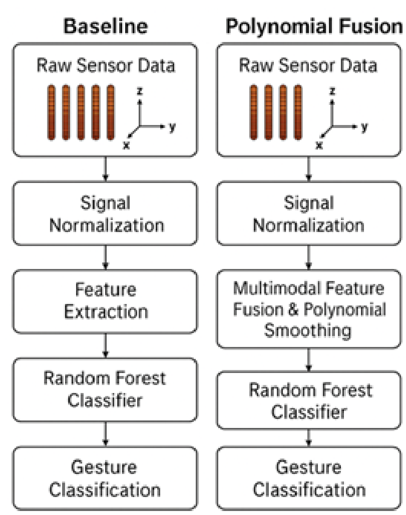


Fig. 5. Baseline vs. Polynomial fusion framework.

F. Feature Extraction, Classification, and Implementation

After multimodal polynomial fusion, compact features are extracted from each fused window, including basic statistical descriptors and temporal gradients, to represent gesture dynamics. A Random Forest classifier is employed due to its robustness to noise and nonlinear feature interactions, with identical settings applied to both the baseline and MPF models to ensure fair comparison. The entire pipeline is implemented in a real-time streaming framework using a sliding-window mechanism, enabling stable predictions and smooth gesture transitions with low computational overhead. Model performance is evaluated using accuracy, macro precision, recall, F1-score, and confusion matrix analysis under subject-independent and real-time conditions.

IV. Results

This section presents the experimental results of the proposed Multimodal Polynomial Fusion (MPF) framework. The evaluation is conducted under subject-independent, session-aware, and real-time streaming conditions to assess recognition accuracy, robustness across users, and temporal stability. Performance is compared against a baseline system using raw normalized flex–IMU features.

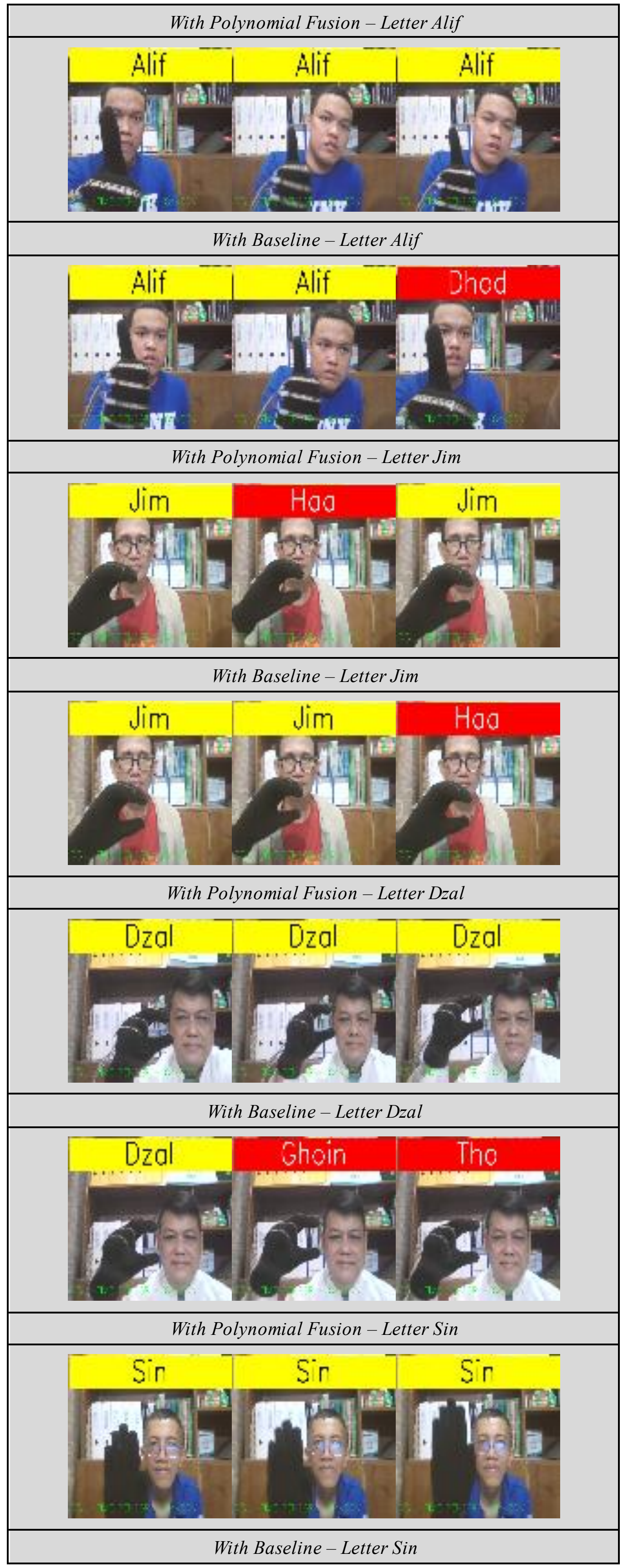
A. Overall Performance Evaluation

To provide a concise quantitative comparison between the baseline system and the proposed Multimodal Polynomial Fusion (MPF) framework, overall recognition performance is summarized using standard evaluation metrics. Accuracy and macro-averaged precision, recall, and F1-score are reported to ensure balanced assessment across all gesture classes under subject-independent and real-time conditions. The results presented in Table I highlight the consistent performance improvements achieved by the proposed framework over the baseline approach.

TABLE I. Overall Performance Comparison

Metric	Baseline Model	MPF Model
Accuracy (%)	92.42	96.32
Precision (Macro)	92.56	96.51
Recall (Macro)	92.44	96.32
F1-Score (Macro)	92.44	96.30

The real-time evaluation shows that MPF consistently outperforms the baseline across all metrics, providing more stable and reliable predictions during continuous gesture execution. Higher precision, recall, and F1-score indicate reduced misclassification and improved consistency across gesture classes, confirming MPF's robustness for real-world assistive use. Fig. 6 further illustrates the real-time behavior of the proposed framework under subject-independent conditions; representative gesture detection results from seven unseen users are presented. Each example corresponds to a different test subject and shows a direct comparison between the baseline model and the proposed Multimodal Polynomial Fusion (MPF) approach for the same Hijaiyah gesture, highlighting user-specific variability and temporal prediction stability.



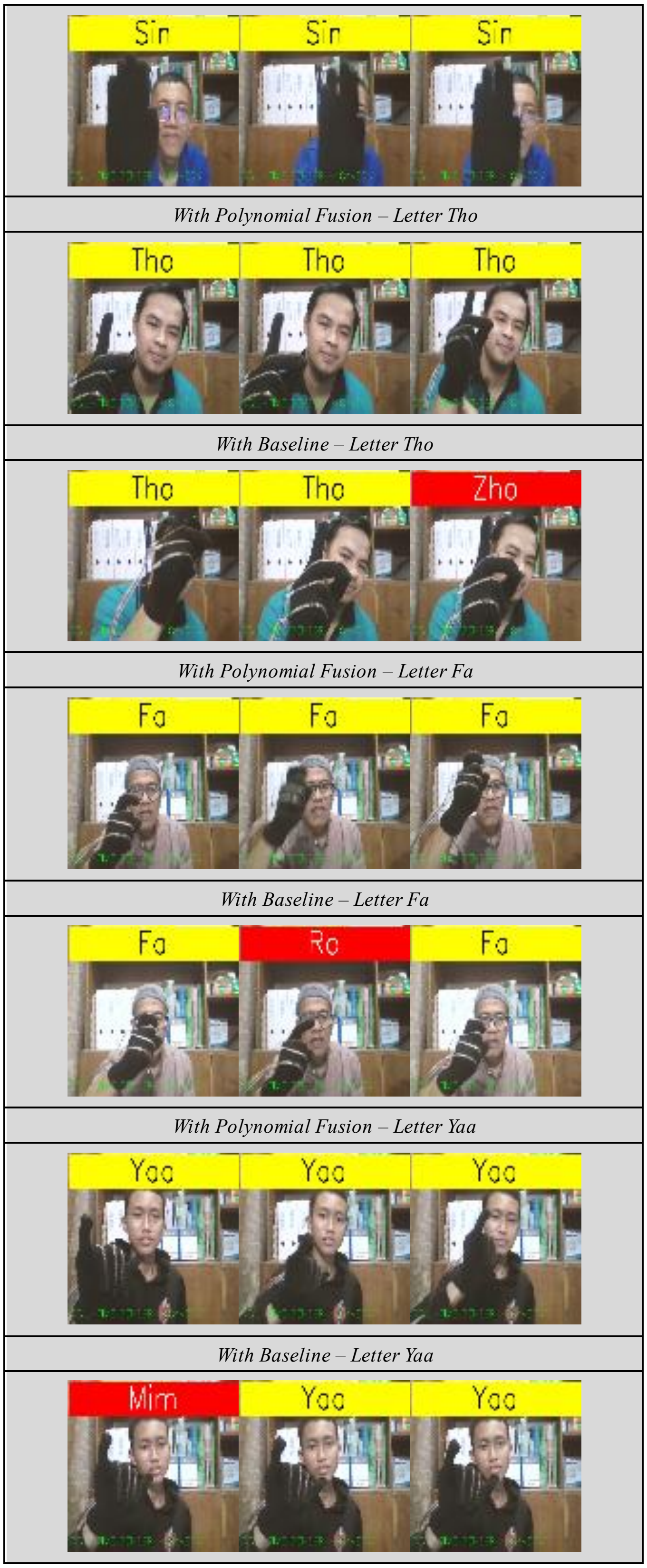


Fig. 6. Representative real-time gesture detection results for selected Hijaiyah gestures.

1) Heatmap of per-gesture real-time accuracy: Fig. 7 presents the per-gesture accuracy heatmap for the proposed MPF model, showing consistently high recognition accuracy across most Hijaiyah gesture classes under real-time streaming conditions.

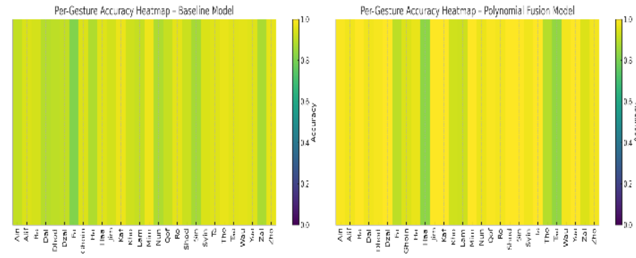


Fig. 7. Presents the per-gesture accuracy heatmap for the MPF model.

2) Real-time accuracy comparison: Compares the real-time accuracy of the baseline and MPF models, as shown in Fig. 8. The MPF model achieves noticeably higher performance, confirming the effectiveness of polynomial smoothing and multimodal fusion in stabilizing gesture predictions during continuous execution.

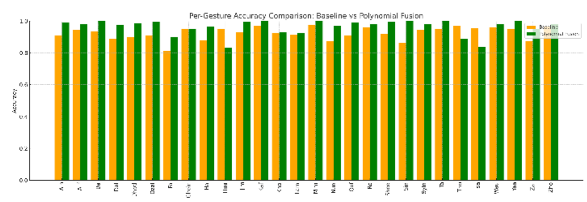


Fig. 8. Real-time accuracy comparison.

B. Misclassification and Confusion Analysis

Misclassification patterns were examined to understand how often specific gestures were confused with others during real-time evaluation. This analysis provides insight into gesture similarities that challenge both models and reveals how Polynomial Fusion (MPF) reduces these errors. In Fig. 9, the baseline model exhibits frequent misclassification in gestures with subtle structural similarities such as Fa, Sin, Zai, Nun, and Ha, resulting in darker regions on the heatmap and lower per-gesture accuracy. In contrast, the MPF model shows far fewer errors, with a more uniform and brighter heatmap that reflects improved stability and separability. Only minor confusion remains in closely related gestures like Haa, Tsa, Kho, Fa, Lam.

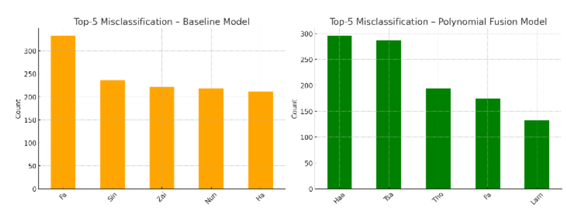


Fig. 9. Top misclassified gestures - polynomial fusion model.

C. Variance Reduction Analysis

To further assess model stability across users, variance analysis was performed using per-user accuracy distributions. Lower variance indicates more consistent performance when encountering natural differences in gesture execution, hand size, glove fit, and movement style.

D. Temporal Stability and Per-Letter Robustness

The per-letter robustness plot shows that the MPF model consistently maintains higher accuracy and smoother performance across all 28 gestures, whereas the baseline model exhibits greater fluctuation and instability. MPF's smoother accuracy curve reflects stronger temporal stability and improved discrimination between similar gestures in Fig. 10, confirming its suitability for real-time recognition.

Fig. 10. Temporal stability and per-letter robustness.

V. Discussion

A. Interpretation of Performance Improvements

Across all offline, session-aware, and real-time evaluations, the MPF framework consistently outperforms the baseline. Its improvements stem from nonlinear temporal smoothing, multimodal fusion, reduced intra-class variance, and enhanced gesture discriminability. These results clearly demonstrate MPF's effectiveness for stable and reliable real-time Hijaiyah sign recognition.

B. Temporal Stability and Gesture Separability

Temporal stability plays a critical role in real-time sign language recognition, where fluctuating predictions can degrade recognition reliability. As observed in the temporal robustness analysis, the proposed MPF framework produces smoother and more consistent predictions over time compared with the baseline. This behavior indicates effective suppression of sensor jitter and short-term fluctuations in flex–IMU signals.

Improved temporal stability contributes directly to enhanced gesture separability by maintaining consistent feature trajectories across consecutive frames. As a result, class boundaries remain more distinct during continuous gesture execution, supporting stable real-time recognition without relying on frame-level corrections.

C. Comparison with Prior Wearable Sensor-Based Methods

Compared with prior wearable sensor–based sign language recognition systems that rely on linear preprocessing and limited user evaluations [6], [7], [15], [24], the proposed MPF framework demonstrates superior robustness and scalability, supported by large-scale subject-independent evaluation and nonlinear temporal modeling.

Table II highlights that the most existing sensor-based sign language recognition systems with reported accuracies below 96% are evaluated using limited datasets and a small or unspecified number of users, which restricts their generalizability. Several approaches rely primarily on flex sensors with predefined gesture mappings, leading to moderate recognition performance and reduced robustness to inter-user variability [6], [30], while others incorporate additional sensing modalities but still operate on relatively small datasets [9], [11], [18]. Moreover, some studies report results under offline evaluation settings [28], [33], which do not adequately reflect stability and consistency during continuous real-time operation. In contrast, the proposed Multimodal Polynomial Fusion (MPF) framework is evaluated on a substantially larger dataset comprising 231,000 samples from 33 users and operates fully in real time, enabling improved robustness and higher recognition accuracy (96.32%) in continuous sign language recognition scenarios.

TABLE II. Literature-Based Comparison of Representative State-of-the-Art Sign Language Recognition Methods

Method / Approach	*Modality*	*Dataset Scale*	*#Users*	*Reported Accuracy (%)*	*Real-Time*	*Ref*
Smart Glove-Based SLR System	Flex Sensors	Limited dataset	–	~85–90	Yes	[6]
Assistive Communication Glove (Combined Sensors)	Flex + Contact + 3D Accel	28 ASL gestures	7	77.9	Yes	[9]
Smart Hand Sign Glove with ESP32	MPU6050 + ESP32	Limited	–	92–95	Yes	[11]
Intelligent Glove with kNN + CHC/DROP3	5 Flex Sensors	~5,000 samples	10	85	Yes	[18]
Indonesian Sign Language Recognition (ANN)	Flex + Accelerometer	1,000 samples	–	91.60 (offline)	No	[28]
Sign Language Detection using Flex Sensor	Flex Sensors	Limited	–	88–92	Yes	[30]
Flex Sensor Dataset for SLR (ML Comparison)	Flex + Accelerometer	~180 samples	–	94 (offline)	No	[33]
Proposed MPF (This Work)	Flex–IMU	231,000 samples	33	96.32	Yes	This study

D. Limitations and Future Directions

Despite its advantages, the proposed wearable sensor-based framework has inherent limitations. The reliance on specialized smart-glove hardware introduces additional cost and usability considerations compared to vision-only approaches. Long-term wearability and user comfort may also affect practical deployment. Future work will explore lightweight hardware designs and cross-modal integration with vision-based cues to further enhance robustness

VI. Conclusion and Future Work

A. Conclusion

This study introduces a Multimodal Polynomial Fusion (MPF) framework for real-time Hijaiyah sign recognition using flex–IMU smart gloves. MPF enhances signal stability, reduces intra-class variability, and improves gesture separability through nonlinear smoothing and multimodal fusion. Across all evaluations, MPF consistently surpasses the baseline, demonstrating a robust and reliable foundation for fine-grained assistive communication systems. These findings directly address the research objective of improving temporal stability and recognition accuracy for fine-grained Hijaiyah gestures in real-time wearable sign language systems.

B. Future Work

Future work may explore adaptive polynomial smoothing, fusion with vision-based modalities, and integration into lightweight neural models for edge deployment. Expanding user diversity, improving glove ergonomics, and extending recognition to continuous Hijaiyah sequences will further enhance the practicality and generalization of sensor-based sign language systems. Despite its advantages, the proposed wearable sensor–based approach has certain limitations. The use of smart gloves introduces hardware dependency and potential cost constraints, which may affect large-scale adoption. In addition, user comfort and long-term wearability remain important considerations. Nevertheless, these limitations are balanced by the system's robustness to environmental conditions and its suitability for real-time assistive applications.

REFERENCES

[1] J. Gałka, M. Ma, M. Zaborski, and K. Barczewska, "Inertial Motion Sensing Glove for Sign Language Gesture Acquisition and Recognition," vol. 16, no. 16, pp. 6310–6316, 2016.

[2] H. P. Athreya, G. Mamatha, R. Manasa, S. Raj, and R. Yashwanth, "Smart Glove for the Disabled : A Survey," vol. 13, no. 2, pp. 21–27, 2021.

[3] C. U. R. F. Signals, Y. Zhang, Y. Wang, and Y. Jiang, "Sign Language Recognition Based on," IEEE Access, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3517417.

[4] G. Chitranshe, "A Framework for Hand Gesture Recognition based on fusion of Flex, Contact and Accelerometer Sensor," pp. 312–319, 2015.

[5] A. Rehman and A. Shoufan, "International Journal of Current Science Research and Review A Linguistic Communication Interpretation Wearable Device for Deaf and Mute User," vol. 05, no. 07, pp. 2714–2720, 2022, doi: 10.47191/ijcsrr/V5-i7-56.

[6] R. Naveen, S. A. Sivakumar, R. Kiruthika, K. Akshayyha, A. R. J, and B. K. M, "Smart Wearable Hand Device for Sign Language," vol. 7, no. 4, pp. 86–92.

[7] P. R. Vidur and M. K. Jambagi, "Wearable Based Sign Language Recognition," vol. 4, no. 12, pp. 2020–2022, 2021.

[8] A. J. Abougarair and W. A. Arebi, "Smart glove for sign language translation," vol. 8, no. 3, pp. 109–117, 2022, doi: 10.15406/iratj.2022.08.00253.

[9] N. Tanyawiwat and S. Thiemjarus, "Design of an Assistive Communication Glove using Combined Sensory Channels," 2012, doi: 10.1109/BSN.2012.17.

[10] L. Dipietro, A. M. Sabatini, S. Member, and P. Dario, "A Survey of Glove-Based Systems and Their Applications," vol. 38, no. 4, pp. 461–482, 2008.

[11] D. Science, "International Journal of Research Publication and Reviews Smart Hand Sign Glove for Deaf and Dumb with Software," vol. 6, no. 11, pp. 6935–6946, 2025.

[12] F. Sensors, "Continuous Finger Gesture Recognition Based on," pp. 1–21, 2019, doi: 10.3390/s19183986.

[13] N. Gautam et al., "Flex sensors and MPU6050 sensors responses on smart glove for sign language translation Flex sensors and MPU6050 sensors responses on smart glove for sign language translation", doi: 10.1088/1757-899X/403/1/012032.

[14] S. E. M. G. Sensors, "A Wearable System for Recognizing American Sign Language in Real - time Using IMU and," pp. 1–10, 2016.

[15] A. Shivaram, V. Sawant, A. Samayam, and N. Mehendale, "Revolutionizing Communication Accessibility : A Sign Language Translator Glove Study," vol. 14, no. 8, pp. 1–6, 2015.

[16] H. Sümbül, "A Novel Mems and Flex Sensor-Based Hand Gesture Recognition and Regenerating System Using Deep Learning Model," IEEE Access, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3448232.

[17] M. I. Sadek, M. N. Mikhael, and H. A. Mansour, "2017 , 34 rd NATIONAL RADIO SCIENCE CONFERENCE Arab Academy for Science , Technology & Maritime Transport A New Approach for Designing a Smart Glove for Arabic Sign Language Recognition System Based on the Statistical Analysis of the Sign Language A . The Most Popular Types of Smart Glove 2017 , 34 rd NATIONAL RADIO SCIENCE CONFERENCE B . Literature review," no. Nrsc, 2017.

[18] P. D. Rosero-montalvo et al., "Sign Language Recognition Based on Intelligent Glove Using Machine Learning Techniques," pp. 5–9, 2018.

[19] G. Saggio, P. Cavallo, M. Ricci, V. Errico, J. Zea, and M. E. Benalc, "Electronics : Implementing k-Nearest Neighbors with Dynamic Time Warping and Convolutional Neural Network Algorithms," 2020.

[20] M. Burns et al., "Design and Implementation of an Instrumented Data Glove that measures Kinematics and Dynamics of Human Hand," pp. 7229–7232, 2021.

[21] X. Huang et al., "Tracing the Motion of Finger Joints for Gesture Recognition via Sewing RGO-Coated Fibers onto a Textile Glove," pp. 1–7.

[22] S. Küçükdermenci, "3 rd International Conference on Frontiers in Academic Multifunctional Smart Glove : An Innovative Solution for Sign Language Interpretation and Wireless Wheelchair Control," pp. 1142–1149, 2024.

[23] M. Borghetti, E. Sardini, and M. Serpelloni, "Sensorized Glove for Measuring Hand Finger Flexion for Rehabilitation Purposes," vol. 62, no. 12, pp. 3308–3314, 2013.

[24] S. M. Biju, O. Al-khatib, H. Z. Sheikh, and F. Oroumchian, "Glove based wearable devices for sign language-GloSign," vol. 12, no. 4, pp. 1666–1676, 2023, doi: 10.11591/ijai.v12.i4.pp1666-1676.

[25] H. Sumbul, "A Novel Fuzzy Logic Based Hand Gesture Recognition System," vol. 13, no. 1, pp. 76–83, 2025, doi: 10.17694/bajece.1519693.

[26] R. Ambar, S. Salim, M. Helmy, A. Wahab, M. M. Abdul, and T. C. Phing, "Development of a Wearable Sensor Glove for Real-Time Sign Language Translation," vol. 7, no. 5, pp. 25–38, 2023, doi: 10.33166/AETiC.2023.05.003.

[27] J. Delpreto, J. Hughes, M. D. Aria, M. De Fazio, and D. Rus, "A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification," IEEE Robot. Autom. Lett., vol. PP, pp. 1–8, 2022, doi: 10.1109/LRA.2022.3191232.

[28] E. Supriyati and M. Iqbal, "Recognition System of Indonesia Sign Language based on Sensor and Artificial Neural Network," vol. 17, no. 1, pp. 25–31, 2013, doi: 10.7454/mst.v17i1.1924.

[29] K. Kadam, S. Telange, K. Yadav, A. Vishwakarma, E. Engineering, and N. Mumbai, "' HELPING HAND : A GLOVE FOR MUTE PEOPLE ,'" vol. 8, no. 4, pp. 505–509, 2023.

[30] G. Sensor, "SIGN LANGUAGE DETECTION FOR DEAF AND DUMB USING FLEX SENSOR," vol. 11, no. 5, 2023.

[31] B. A. Sahm, H. Al-fahaam, and A. A. Jasim, "Journal of Engineering and Technological Sciences American Sign Language Translation from Deaf-Mute People Based on Novel System," vol. 56, no. 2, pp. 193–204, 2024, doi: 10.5614/j.eng.technol.sci.2024.56.2.2.

[32] B. G. Lee and S. M. Lee, "Smart Wearable Hand Device for Sign Language Interpretation System With Sensors Fusion," vol. 18, no. 3, pp. 1224–1232, 2018.

[33] N. Arun, R. Vignesh, B. Madhav, A. K. S, and S. Sasikala, "Flex Sensor Dataset : Towards Enhancing the Performance of Sign Language detection System," 2022.