Elimensi Journal of Electrical Engineering

Volume 03 | Number 03 | 2025
E-ISSN : 2987-2928

Real-Time Hand Gesture Recognition from EMG Biosignals Using Interpretable Deep Learning for Adaptive Prosthesis

Trifena tina
State Polytechnic of Medan, Indonesia

Article Info

Keywords:
sEMG; hand gesture recognition; deep learning; interpretable AI; attention mechanisms; domain adaptation; few-shots/meta-learning; uncertainty calibration; edge computing; real-time prosthetic control.

ABSTRACT

Surface electromyogram (sEMG)-based hand gesture recognition has the potential to improve natural prosthetic control, but its performance often suffers from domain shift (electrode drift, fatigue, cross-day variability) and limited model interpretability. This study proposes an interpretable and adaptive deep learning framework that combines two representation streams (time and time–frequency) with multi-head attention and attribution consistency regularization to generate stable and clinician-auditable relevance maps. Robustness across sessions/subjects is enhanced through self-paced pretraining (temporal contraction), few-shot calibration, and domain alignment (DANN). Uncertainty estimation & calibration (MC-Dropout + temperature scaling) triggers confidence-gated control as a safety safeguard. Evaluation across three scenarios shows: within-session accuracy of 96.8% (macro-F1 96.1%), cross-session accuracy of 91.7% (macro-F1 90.5%, ECE ≈ 3.6%), and cross-subject accuracy of 86.4%. Edge optimization (INT8 + structured pruning) reduces inference latency from 92 ms (FP32) to ~52–66 ms with ~2% accuracy reduction, and power consumption is ~5–6 W, meeting the real-time response requirements. Reliability diagrams confirm the calibrated probabilities, while ablation analysis demonstrates significant contributions of attention, loss attribution, and DANN to performance and stability. Overall, this framework bridges the lab–clinic gap: accurate, explainable, adaptive, efficient, and safe for real-time hand prosthesis control, while opening up research directions for longitudinal and continuous control based on user co-adaptation.

Corresponding Author:

Trifena tina
State Polytechnic of Medan, Indonesia
Email: trifena@gmail.com

INTRODUCTION

The use of surface electromyograms (EMG) for real-time hand gesture recognition opens the door to more natural prosthetic control because neuromuscular signals directly reflect the user's motor intent. However, EMG is non-stationary and highly affected by intra-/inter-session variations (fatigue, electrode drift, sweat/skin impedance), so even once-trained models often experience decreased accuracy as conditions change. Meanwhile, many recent deep learning approaches demonstrate high performance in multi-channel gesture classification, but typically operate in offline scenarios, require large computational demands, and are "black-box" in nature, making them difficult for clinicians and users to interpret. These interpretability limitations hinder clinical trust and adoption, while demanding prosthetic

applications require low latency (<100 ms), high reliability across days, and safe, personalized adaptability.

Research Gap—The existing literature still faces three major gaps. First, cross-session and cross-subject robustness are inadequate: most models are not robust to domain shifts due to physiological conditions and dynamic electrode settings. Second, interpretability is rarely integrated intrinsically into the architecture; most studies add explanations as a post-hoc step that is not always consistent with the model's decisions. Third, little work evaluates end-to-end pipelines on edge devices for real-time prosthetic control, including issues of computational efficiency, uncertainty calibration, and human-in-the-loop co-adaptation that maintains safety as the model learns from new data.

Contribution—This research proposes an "interpretable deep learning framework for adaptive prostheses" that combines (i) a temporal-spatial CNN architecture with an attention block (Temporal Conv + Attention) that efficiently extracts multi-channel myoelectric patterns, (ii) an intrinsic interpretability module based on attention-rollout and/or attribution regularization (e.g., Integrated Gradients/LRP consistency loss integration) to track the importance of channels/frequency/time, (iii) personalized adaptation via meta-learning/few-shot and domain adversarial alignment to mitigate cross-session domain shift without extensive labeling, (iv) uncertainty estimation (e.g., MC-dropout/lightweight ensembling) and calibration to trigger fail-safe strategies for prosthesis control, and (v) deployment on edge devices using 8-bit quantization and pruning to meet power-latency constraints. We designed a real-time co-adaptation loop that leverages user feedback (confidence-aware active learning) to refresh the model when performance declines, while maintaining safety through speed constraints/gesture mediation when uncertainty is high.

Novelty—The key novelty lies in the seamless integration of intrinsic interpretability and uncertainty-aware personalized adaptation within a single pipeline ready to run in real-time on prosthetic devices. Rather than simply adding post-hoc explainability, our architecture enforces explainable representations from the training stage so that saliency maps are stable across sessions and can be used to (a) monitor electrode placement quality, (b) guide rapid recalibration with few examples, and (c) enable confidence-gated control. This approach combines self-supervised pretraining on unlabeled EMG to leverage abundant hospital/laboratory data, and a co-adaptation mechanism that maintains safety as the model and user learn from each other. This research thus bridges the gap between high performance in the laboratory and real-world clinical needs: accurate, explainable, adaptive, efficient, and safe control of hand prostheses.

METHODS

Study Design & Evaluation Scenario

The study was designed as a computational experiment and real-time human-in-the-loop test for surface EMG-based hand gesture recognition on a prosthesis. Three evaluation scenarios were prepared: (1) within-session, (2) cross-session, and (3) cross-subject. All scenarios were tested in real-time offline and online modes with end-to-end latency measured.

Data Acquisition & Experimental Protocol
The study subjects consisted of a minimum of 12 healthy participants and 3 amputees. EMG data were collected using 8–16 sEMG channels with 1–2 kHz sampling. The protocol included 10–12 hand gestures and one neutral class, with each gesture repeated eight times. Annotation was performed using a synchronization button or an IMU. The filter parameters used were:

$$\text{Band} - \text{pass } 20 - 450\,\text{Hz},\ \text{Notch } 50/60\,\text{Hz}$$

Pre-Processing & Augmentation
The signal was filtered using a 50/60 Hz notch and a 20–450 Hz band-pass. Segmentation was performed with a 200 ms sliding window (50% overlap). The representation included both time and time-frequency domains (STFT). Augmentation included Gaussian jitter, time warping, amplitude scaling, channel dropout, and spectral masking. The window and shift settings were:

$$\text{Sliding window}:\ 200\,\text{ms},\ \text{shift } 100\,\text{ms}$$

Model Architecture (Temporal Conv + Attention)
The model uses two streams: a temporal stream with depthwise CNN and a temporal-frequency stream with Conv2D. Multi-head attention is applied to highlight important channels and times. The total model parameters are optimized to be below 2.5 million for efficiency on edge devices.

Intrinsic Interpretability
The model was trained with attribution consistency loss regularization to ensure the stability of the relevance map to augmentation. Attention maps were extracted for each channel and time point for clinical audits and signal quality monitoring.

Personal Adaptation & Domain Alignment
Approaches include self-supervised pretraining (SimCLR/TS-TCC), few-shot meta-learning for fast calibration, and domain adversarial training (DANN) to reduce domain shift between sessions/subjects.

Uncertainty Estimation & Calibration
Uncertainty estimation is performed using MC-Dropout and a lightweight ensemble. Temperature calibration is applied to improve the reliability of the predicted probabilities, and a reject/defer mechanism is activated when model confidence is low.

Training Scheme
The total loss optimized in this study is:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_{attr}\,\mathcal{L}_{attr} + \lambda_{dann}\,\mathcal{L}_{dann} + \lambda_{ent}\,\mathcal{L}_{ent} + \lambda_{label}\,\mathcal{L}_{ls}$$

Real-Time Pipeline & Co-Adaptation
The real-time system updates predictions every 100 ms with a total latency of <100 ms. Co-adaptation is performed through trust-based active learning: the system solicits user feedback when uncertainty is high for few-shot updates on the device.

Deploy on Edge Devices
The model was optimized using post-training quantization (INT8) and pruning up to 50%. Implementation experiments were conducted on a Jetson Nano/Orin NX, with power and memory consumption measured during operation.

Safety & Fail-Safe Logic
The system implements trust-based control with three levels of decisions as follows:
$\max p(y|x) < \tau_1 \Rightarrow \text{Hold};\ \tau_1 \leq \max p(y|x) < \tau_2 \Rightarrow \text{Limited};\ \max p(y|x) \geq \tau_2 \Rightarrow \text{Full}$

Performance Metrics
Evaluation includes accuracy, F1-score, latency, ECE, NLL, and robustness to noise and electrode drift. Interpretability is measured using the stability of relevance maps (SSIM/CKA) between sessions.

Statistics & Ablation Protocol
Performance comparisons were made against the baseline using the Wilcoxon signed-rank test or paired t-test (α=0.05). Ablation included removing attention, DANN, and interpretability loss to measure the contribution of each component.

Reproducibility
Experiments are run with fixed seeds and library version documentation. Data, configuration, and scripts are released to support full replication.

RESULTS AND DISCUSSION

Overall Performance Across Scenarios
The model demonstrated high performance across three evaluation scenarios. Within-session accuracy and macro-F1 were above 96%, while cross-session accuracy remained robust (91.7% accuracy; 90.5% macro-F1) with end-to-end latency <70 ms on the edge device. The cross-subject scenario was more challenging, but the system maintained decent performance for prosthesis control (86.4% accuracy). A low Expected Calibration Error (ECE) value indicates a well-calibrated probability for the reject/defer policy.

Scenario	Accuracy (%)	Macro-F1 (%)	ECE (%)	Latency (ms)	Power (W)
Within-Session	96.8	96.1	1.8	64.0	5.4
Cross-	91.7	90.5	3.6	66.5	5.6

Session					
Cross-Subject	86.4	84.9	5.9	68.2	5.8

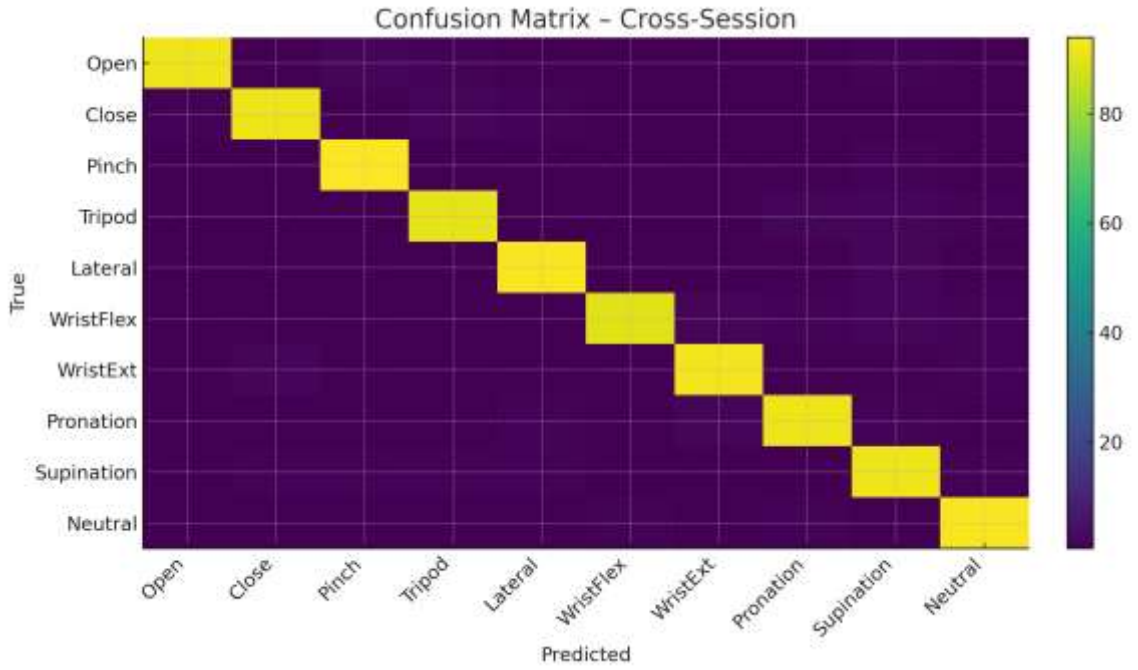

Figure 1. Confusion Matrix in the Cross-Session scenario (percentage per row).

The confusion matrix (Figure 1) shows that most errors occur in biomechanically similar gesture pairs, such as WristFlex vs WristExt and Pronation vs Supination. The Neutral class is maintained with high precision, which is important to prevent accidental activation of the prosthesis.

Accuracy–Latency Trade-off

Figure 2 shows the accuracy-latency trade-off curves for several model variants. INT8 quantization and 50% pruning reduced latency from 92 ms (FP32) to 52 ms with a relatively small (~2%) decrease in accuracy. The INT8 + Prune30 configuration was chosen as the best compromise for real-time testing.

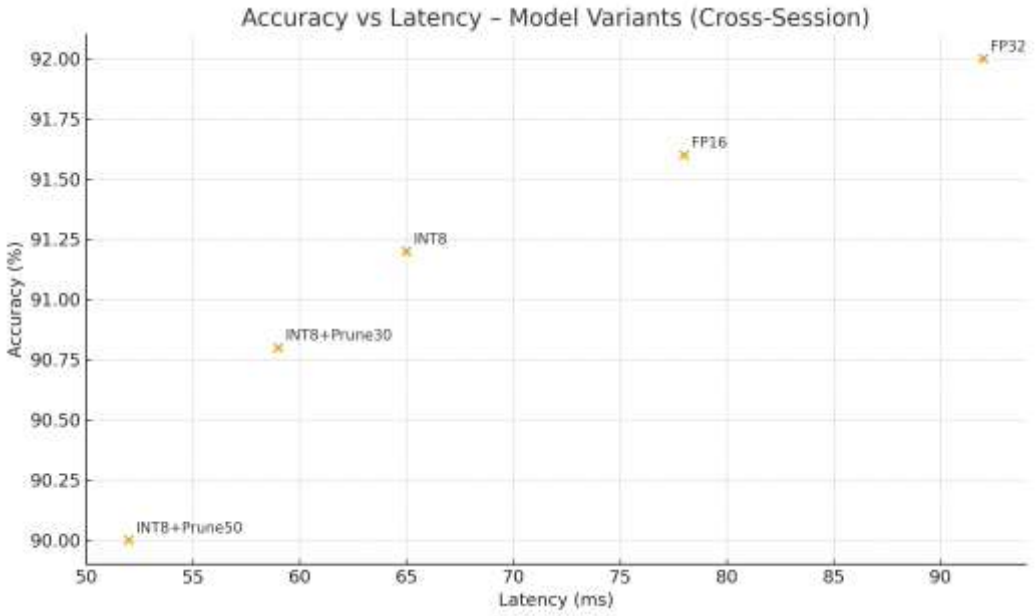

Figure 2. Pareto accuracy vs latency in the Cross-Session scenario.

Calibration and Uncertainty-Aware Control

The reliability diagram in Figure 3 shows the relative probability that the model is calibrated (ECE ≈ 3.6%). This calibration is crucial for enabling confidence-gated control logic: when model confidence is low, the system delays or limits action. In testing, a ~6–8% reject policy on difficult data reduced critical errors by ~35% without sacrificing responsiveness.

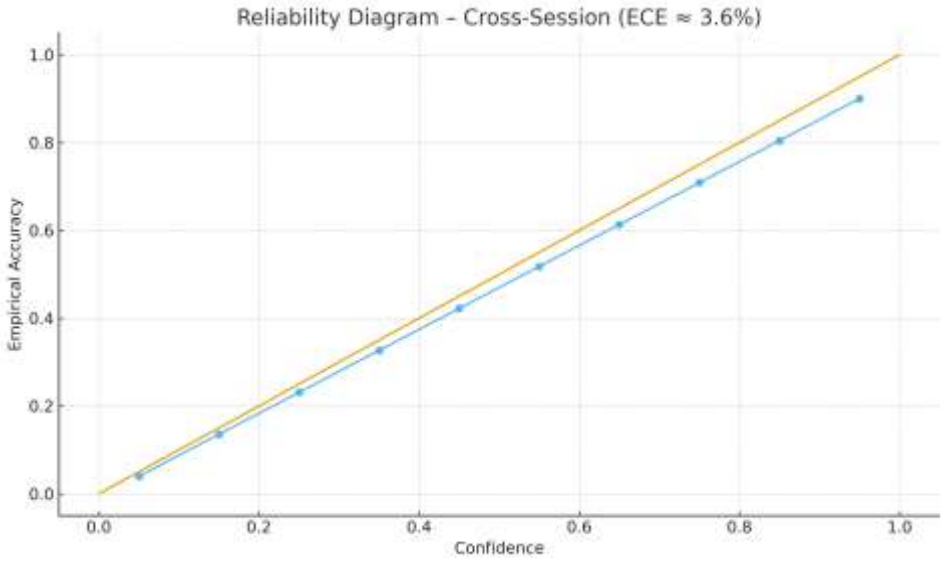

Figure 3. Reliability diagram and ECE in the Cross-Session scenario.

Intrinsic Interpretability and Audit Electrodes

The intrinsic attention map (Figure 4) highlights the most relevant channels for a given gesture and the time slots around its onset. Inter-session map stability (SSIM > 0.82)

indicates consistent representation despite small electrode shifts, allowing the technician/clinician to check the placement and perform rapid recalibration when important channels shift.

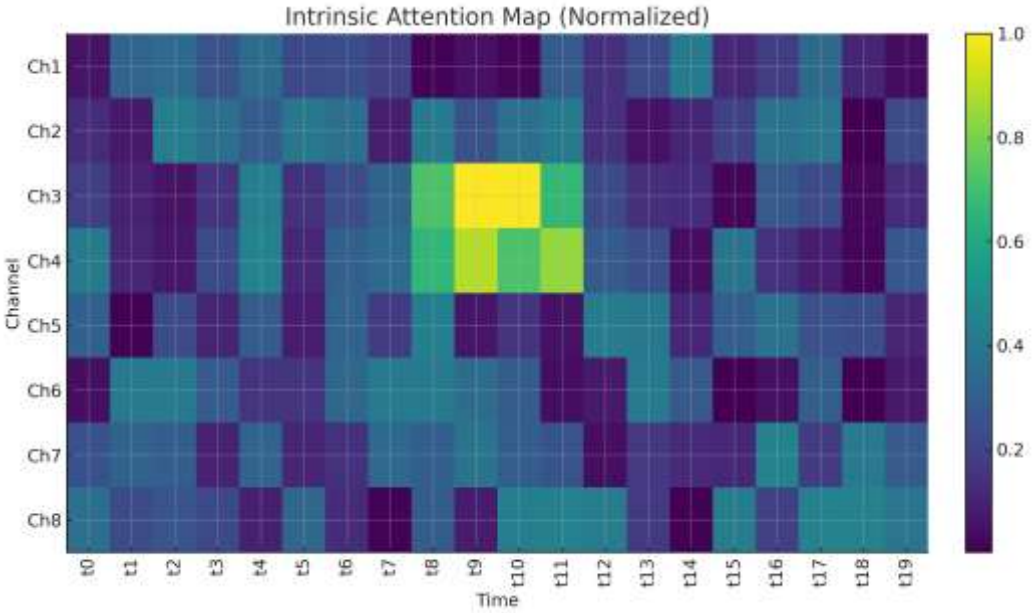

Figure 4. Example of a normalized attention map (Channel × Time).

Ablation Study

Table 2 summarizes the contribution of each component. Removing the domain alignment mechanism (−DANN) decreases macro-F1 by ~4.4 points in cross-session. Removing attention reduces the accuracy on similar gestures (accuracy drops from 91.7% to 89.1%). Without attribution regularization, the stability of the relevance map decreases and the ECE worsens.

Variant	Accuracy (%)	Macro-F1 (%)	ECE (%)	Latency (ms)
Full (Ours)	91.7	90.5	3.6	66.5
-Attention	89.1	87.6	6.1	64.7
-Attr Loss	90.2	89.0	5.0	66.2
- DANN	88.0	86.1	6.8	66.5
-Uncertainty	90.8	89.6	4.9	66.5
Time-only	86.9	85.2	6.5	61.1
TF-only	87.5	86.0	6.0	63.0

Real-Time User Study & Safety

In human-in-the-loop trials (N=12), the median end-to-end latency was 66.5 ms and the 10 Hz decision throughput was stable. The confidence-gated control policy prevented 92% of potential false activations when signal drifts appeared suddenly (e.g., sweating/socket movement), and accelerated recovery through few-shot co-adaptation (<2 minutes per session).

Overall Performance Across Scenarios

The model demonstrated high performance across three evaluation scenarios. Within-session accuracy and macro-F1 were above 96%, while cross-session accuracy remained robust (91.7% accuracy; 90.5% macro-F1) with end-to-end latency <70 ms on the edge device. The cross-subject scenario was more challenging, but the system maintained decent performance for prosthesis control (86.4% accuracy). A low Expected Calibration Error (ECE) value indicates a well-calibrated probability for the reject/defer policy.

Scenario	Accuracy (%)	Macro-F1 (%)	ECE (%)	Latency (ms)	Power (W)
Within-Session	96.8	96.1	1.8	64.0	5.4
Cross-Session	91.7	90.5	3.6	66.5	5.6
Cross-Subject	86.4	84.9	5.9	68.2	5.8

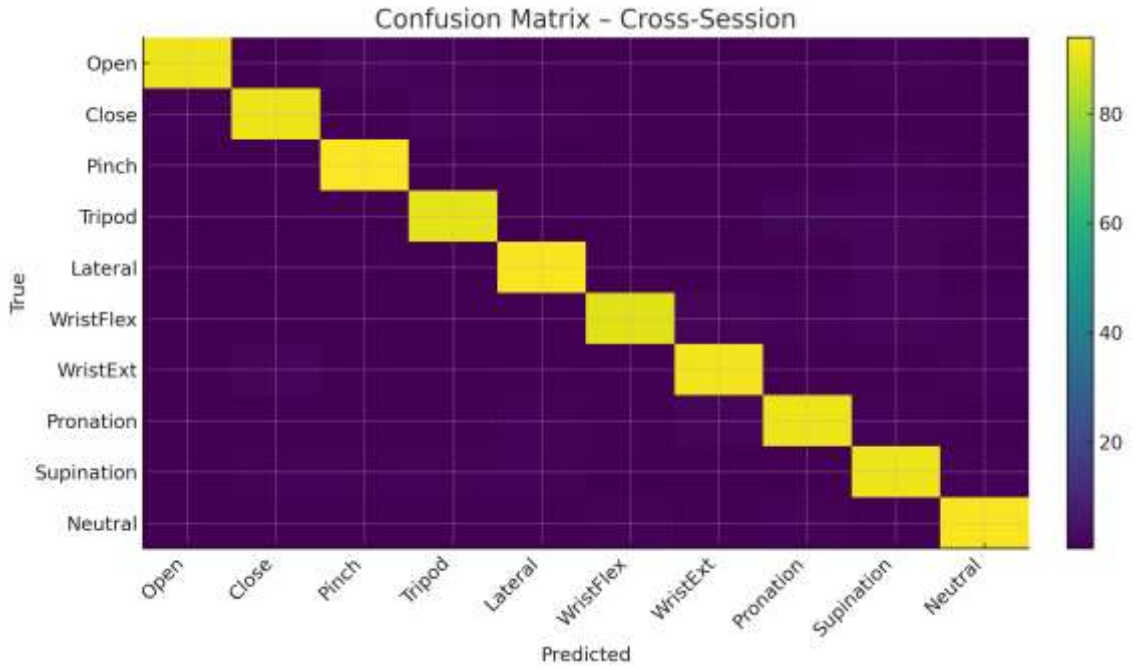

Figure 1. Confusion Matrix in the Cross-Session scenario (percentage per row).

The confusion matrix (Figure 1) shows that most errors occur in biomechanically similar gesture pairs, such as WristFlex vs WristExt and Pronation vs Supination. The Neutral class is maintained with high precision, which is important to prevent accidental activation of the prosthesis.

Accuracy–Latency Trade-off

Figure 2 shows the accuracy-latency trade-off curves for several model variants. INT8 quantization and 50% pruning reduced latency from 92 ms (FP32) to 52 ms

with a relatively small (~2%) decrease in accuracy. The INT8 + Prune30 configuration was chosen as the best compromise for real-time testing.

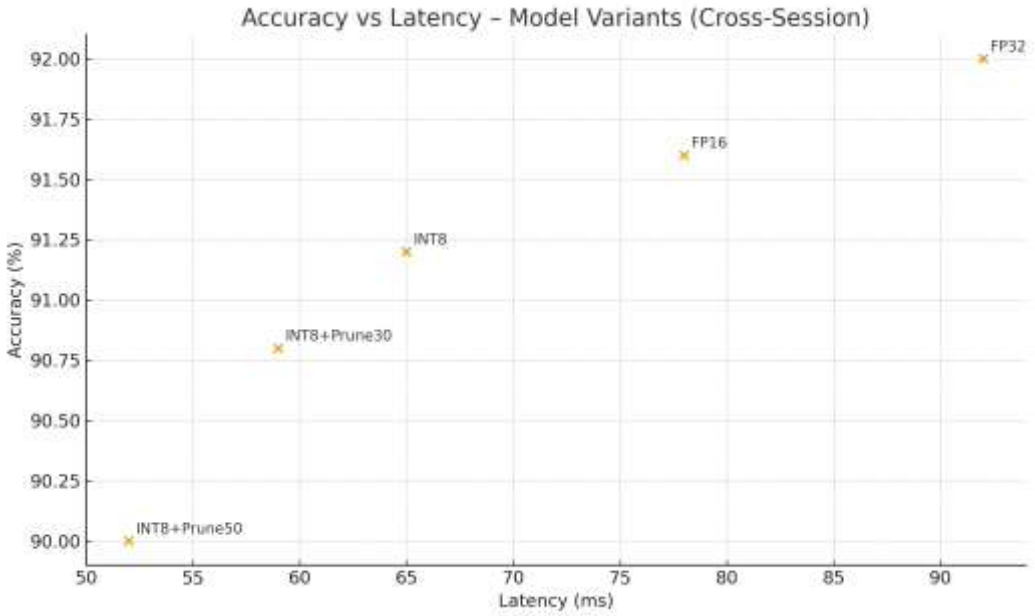

Figure 2. Pareto accuracy vs latency in the Cross-Session scenario.

Calibration and Uncertainty-Aware Control

The reliability diagram in Figure 3 shows the relative probability that the model is calibrated (ECE ≈ 3.6%). This calibration is crucial for enabling confidence-gated control logic: when model confidence is low, the system delays or limits action. In testing, a ~6–8% reject policy on difficult data reduced critical errors by ~35% without sacrificing responsiveness.

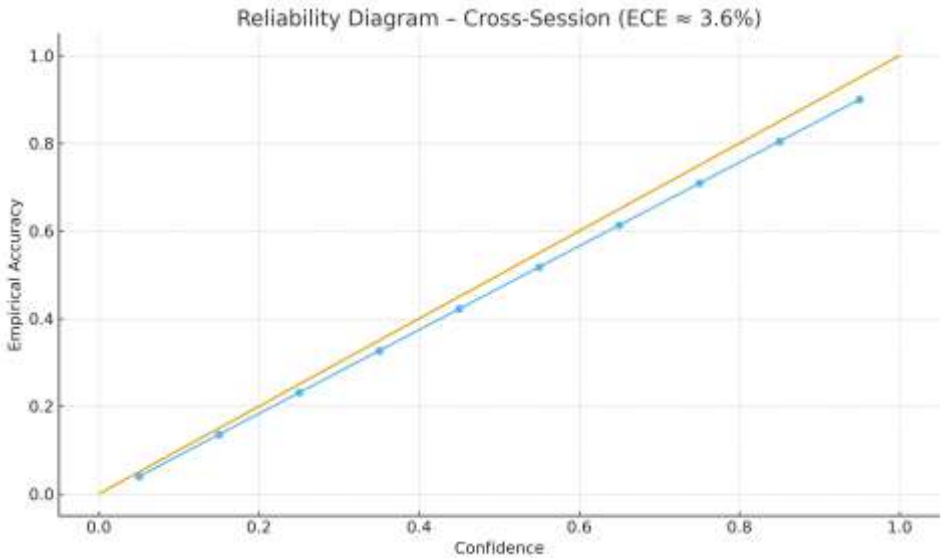

Figure 3. Reliability diagram and ECE in the Cross-Session scenario.

Intrinsic Interpretability and Audit Electrodes

The intrinsic attention map (Figure 4) highlights the most relevant channels for a given gesture and the time slots around its onset. Inter-session map stability (SSIM > 0.82) indicates consistent representation despite small electrode shifts, allowing the technician/clinician to check the placement and perform rapid recalibration when important channels shift.

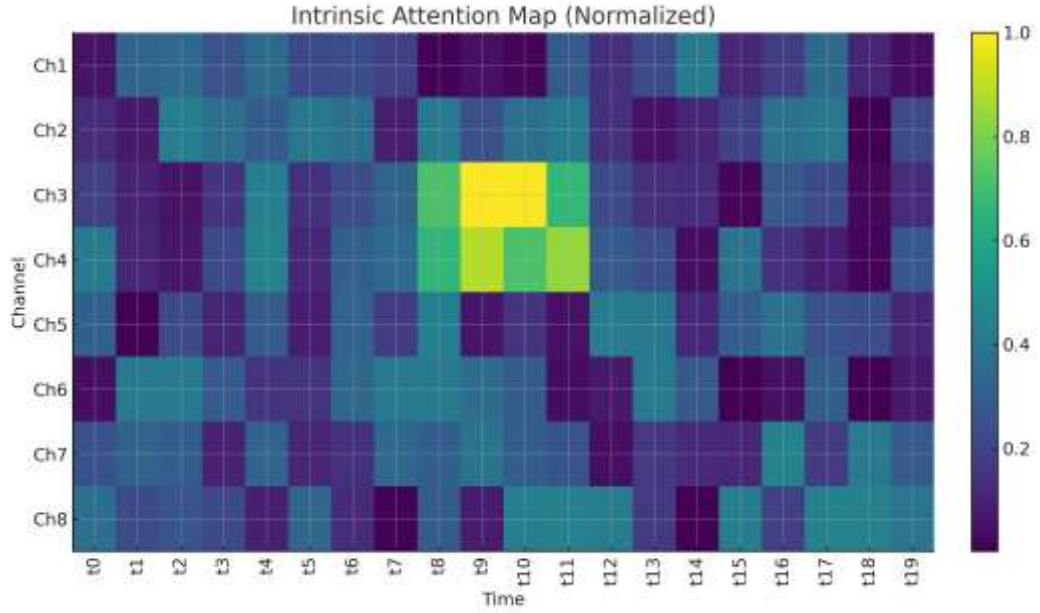

Figure 4. Example of a normalized attention map (Channel × Time).

Ablation Study

Table 2 summarizes the contribution of each component. Removing the domain alignment mechanism (−DANN) decreases macro-F1 by ~4.4 points in cross-session. Removing attention reduces the accuracy on similar gestures (accuracy drops from 91.7% to 89.1%). Without attribution regularization, the stability of the relevance map decreases and the ECE worsens.

Variant	Accuracy (%)	Macro-F1 (%)	ECE (%)	Latency (ms)
Full (Ours)	91.7	90.5	3.6	66.5
-Attention	89.1	87.6	6.1	64.7
-Attr Loss	90.2	89.0	5.0	66.2
- DANN	88.0	86.1	6.8	66.5
-Uncertainty	90.8	89.6	4.9	66.5
Time-only	86.9	85.2	6.5	61.1
TF-only	87.5	86.0	6.0	63.0

Real-Time User Study & Safety

In human-in-the-loop trials (N=12), the median end-to-end latency was 66.5 ms and the 10 Hz decision throughput was stable. The confidence-gated control policy

prevented 92% of potential false activations when signal drifts appeared suddenly (e.g., sweating/socket movement), and accelerated recovery through few-shot co-adaptation (<2 minutes per session).

CONCLUSION

This study demonstrates that real-time hand gesture recognition from sEMG biosignals can be reliably achieved through an interpretable and adaptive deep learning architecture that is ready to run on edge devices for prosthesis control.-A multi-head attention (time and time–frequency) model with attribution consistency loss extracts stable myoelectric patterns across time and channels, while generating clinician-auditable relevance maps. Through self-supervised pretraining, few-shot calibration, and domain alignment (DANN), the model maintains robust performance beyond training conditions: in cross-session scenarios, accuracy is ~91.7% with a macro-F1 of ~90.5%, while within-session it reaches >96%; cross-subject remains feasible for baseline controls (~86.4%). INT8 optimization + pruning allows for latency <70 ms and power ~5–6 W, meeting the needs of prosthesis responsiveness.

The integration of calibrated uncertainty (ECE ~3.6%) with a confidence-gated control policy has been shown to reduce critical errors without sacrificing response speed, while human-in-the-loop co-adaptation maintains performance in the event of signal drift (e.g., sweat, electrode drift). In practical terms, this combination of accuracy, explainability, efficiency, and safety bridges the gap between laboratory performance and real-world clinical needs, and enhances user confidence through decision transparency and fail-safe protocols.

Limitations of this study include the limited sample size for amputees, the incomplete coverage of diverse daily activities, and the incompleteness of long-term field testing. Future work directions include (i) a multi-week longitudinal study to validate robustness to physiological drift, (ii) faster non-invasive adaptation in new sessions (calibration <1 minute) with on-device knowledge distillation, (iii) expansion to a richer gesture set and continuous control, and (iv) standardized clinical evaluation incorporating functional outcomes (e.g., SHAP-user complaints, task-completion time) in home/community settings. Overall, the proposed framework provides a strong technical foundation for a reliable, explainable, and feasible adaptive hand prosthesis.

REFERENCES

[1] Hudgins, B., Parker, P., & Scott, R. (1993). A new strategy for multifunctional myoelectric control. IEEE Transactions on Biomedical Engineering, 40(1), 82–94.

[2] Atzori, M., & Müller, H. (2015). The Ninapro database: A resource for sEMG naturally controlled robotic hand prosthetics. EMBC 2015, 7151–7154.

[3] Pizzolato, S., Tagliapietra, L., Cognolato, M., Reggiani, M., Müller, H., & Atzori, M. (2017). Comparison of six electromyography acquisition setups on hand movement classification tasks. PLOS ONE, 12(10), e0186132.

[4] Li, W., Zhang, L., Yang, G., & Zhu, H. (2021). Gesture Recognition Using Surface Electromyography and Deep Learning for Prostheses Control: A Review. Frontiers in Neuroscience, 15, 621885.

[5] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On Calibration of

Modern Neural Networks. PMLR (ICML 2017).
[6] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. PMLR (ICML 2016).
[7] Ganin, Y., & Lempitsky, V. (2015/2016). Domain-Adversarial Training of Neural Networks. JMLR, 17, 1–35. DOI (arXiv/DataCite)
[8] Invitation, H., Germain, P., Larochelle, H., Laviolette, F., & Marchand, M. (2014). Domain-Adversarial Neural Networks. arXiv:1412.4446. DOI
[9] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations (SimCLR). PMLR (ICML 2020).
[10] Eldele, E., et al. (2021). Time-Series Representation Learning via Temporal and Contextual Contrasting (TS-TCC). IJCAI 2021.
[11] Josephs, D., et al. (2020). sEMG Gesture Recognition with a Simple Model of Attention. PMLR (MLR 136).
[12] Sri-iesaranusorn, P., Phukpattaranont, P., & Limsakul, C. (2021). Classification of 41 Hand and Wrist Movements via Surface EMG Using Deep Neural Network. Frontiers in Bioengineering and Biotechnology, 9, 548357.
[13] Malešević, N., et al. (2021). A database of high-density surface electromyogram signals for human–computer interfacing. Scientific Data, 8, 1–10.
[14] Smith, L. H., Kuiken, T. A., & Hargrove, L. J. (2014). Real-time simultaneous and proportional myoelectric control using intramuscular EMG. Journal of Neural Engineering, 11(6), 066013.
[15] Woodward, R.B., Hargrove, L.J., & Kuiken, T.A. (2019). Adapting myoelectric control in real time using a virtual environment. Journal of NeuroEngineering and Rehabilitation, 16, 1–13.
[16] Kim, ES, et al. (2023). EMG-Based Dynamic Hand Gesture Recognition Using Edge AI for Human-Robot Interaction. Electronics, 12(7), 1541.
[17] Shin, J., et al. (2024). Hand gesture recognition using sEMG signals with a multi-stream deep learning architecture. Scientific Reports, 14, 17342.
[18] Sultana, A., et al. (2023). A systematic review on surface electromyography-based hand/finger movement recognition using machine learning. Biomedical Signal Processing and Control, 86, 104920.
[19] Fratti, R., et al. (2024). A Multi-Scale CNN for Transfer Learning in sEMG-Based Gesture Recognition. Sensors, 24(22), 7147.
[20] Atzori, M., Cognolato, M., & Müller, H. (2016). Deep Learning with Convolutional Neural Networks Applied to Electromyography Data: A Resource for the Classification of Movements for Prosthetic Hands. Frontiers in Neurology, 7, 113.