2.2 Related Work

Recent work on sEMG-based gesture recognition has largely focused on deep learning models using the NinaPro DB7 dataset. For example, Z. Chen et al. [1] proposed a Multi-Feature Fusion Network (MFF-Net) that combines multiple CNN branches with a BiLSTM and attention mechanism operating on frequency-, time-, and spatial-domain features. Using intrasubject 3-fold cross-validation, MFF-Net achieved 92.47% accuracy on DB7 intact subjects (and 84.93% after transfer to the 2 amputee subjects). In a related hybrid model, Z. Chen et al. (2025) also fused sEMG and accelerometer (ACC) signals via a "HyFusion" framework with parallel intramodal/ intermodal branches, residual CNNs, multiscale spatial attention and a joint softmax+metric-learning loss. This multimodal fusion achieved 96.44% accuracy on DB7 (with 94.73% on DB2 and 89.60% on DB3), substantially outperforming simpler data-fusion baselines.

Other works explore different fusion and model types. S. Duan et al. [2] applied a deep forest (no backpropagation) to fuse sEMG and ACC features. Their Multi-Modality Deep Forest (MMDF) achieved ~91.4% accuracy on intact DB7 subjects (40 gestures) and only 77.8% on amputees, highlighting the challenge of amputee data. Notably, these multimodal models emphasize feature-level fusion of sEMG+ACC (or IMU) inputs. By contrast, many baseline models use only sEMG. For instance, A. Fatayer et al. [3] built a standard CNN for classifying 41 hand movements on DB7 and reported ~91.7% accuracy. P. Sri-iesaranusorn et al. [4] similarly used a CNN to recognize 41 classes and achieved 91.69% (±4.68%) accuracy. These works serve as strong supervised baselines.

More recently, transformer-based and pretraining approaches have emerged. Y. Fang [5] introduced TinyMyo, a compact Transformer-encoder "foundation model" pre-trained in a self-supervised fashion on large EMG datasets. TinyMyo (3.6M parameters) generalized across tasks and achieved 89.4% on NinaPro DB5 and 96.7% on the EPN-612 hand gesture dataset, although it was not directly evaluated on DB7 (it would require re-mapping DB7's 12 channels). M. Fasulo et al. [6] proposed PhysioWave, a self-supervised wavelet-transformer model pre-trained on large EMG corpora. PhysioWave's small variant (5M parameters) set new state-of-the-art accuracy on DB5 by adaptively learning wavelet bases and selective masking, although it has much higher complexity and is not natively designed for DB7. These foundation-model works suggest SSL can improve generalization across signal variabilities.

Another line of work addresses label noise and regression. Y. Chen et al. [7] studied the effect of noisy labels in NinaPro DB7. Their Adaptive Label Refinement CNN (ALR-CNN) iteratively refines suspected mislabeled trials during training, yielding 95.85% classification accuracy on DB7. This is notably higher than most baselines, at the cost of a complex training procedure. Finally, C. Lin et al. [8] proposed a Parallel-Efficient Transformer (PET) for continuous hand-kinematics regression on DB7. PET achieved 88.72% (based on correlation or R^2) with extremely fast inference, demonstrating transformer architectures adapted for low-latency edge use.

In summary, the strongest supervised methods on DB7 rely on deep CNNs and attention (with or without ACC fusion) achieving >90% accuracy, whereas recent SSL-style pre-training (TinyMyo, PhysioWave) shows promise in related datasets. These studies together form the baselines and inspiration for our SSL approach.

Summary Table

Title	***Dataset Name & URL***	***Dataset Description***	***Methods Name***	***Accuracy***	***Pros***	***Cons***	***Ref.***
sEMG-Based Gesture Recognition via Multi-Feature Fusion Network	Ninapro DB7 & DB3	**Subjects:** 22 (20 intact + 2 amputees) for DB7, 11 amputees for DB3 **Channels:** 12 sEMG **Gestures:** 18 (8 finger, 9 wrists, 1 rest) **Repetitions:** 6 per gesture **Split:** 3-fold cross-validation based on repetitions (no data leakage).	Multi-Feature Fusion Network (MFF-Net) using CNN (frequency), Bi-LSTM + attention (time), Conv-LSTM (spatial)	92.47% (DB7 intact) 84.93% (DB7 amputees after transfer) 82.00% (DB3 amputees after transfer)	Combines time, frequency, and spatial features for richer representation; attention improves important signal learning; avoids data leakage; very high recognition accuracy; effective for amputees with transfer learning.	Computationally heavy due to multiple subnetworks; requires high-quality multi-channel EMG data; not suitable for low-channel wearable systems.	Z. Chen et al. (2025)
A Hybrid Multimodal Fusion Framework for sEMG-ACC-Based Hand Gesture Recognition	Ninapro DB2, DB3, DB7	**DB2:** 40 healthy subjects, 50 gestures, 6 trials per gesture, 12 sEMG + 36 ACC channels, 2000 Hz. Training: trials 1,3,4,6; Testing: trials 2,5. **DB3:** 11 transradial amputees (complete cases), 50 gestures, same protocol as DB2. **DB7:** 20 healthy subjects, 41 gestures, 2000 Hz, sEMG + ACC (IMU available but ACC used). Same split protocol.	HyFusion (Hybrid Multimodal Fusion Framework) using parallel intramodal (sEMG, ACC) + intermodal branches, residual CNN blocks, Multiscale Spatial Attention (SA), Softmax + Island Loss (metric learning), decision-level summation fusion.	DB7: 96.44% DB2: 94.73% DB3: 89.60% Inference time: 10.45 ms/sample.	Hybrid fusion captures both intramodal specificity and intermodal association; multiscale attention enhances hierarchical feature learning; island loss improves discrimination of similar gestures; works for intact and amputees; state-of-the-art performance; real-time feasible.	Relies on handcrafted features instead of raw end-to-end learning; channel-wise normalization may affect inter-channel correlation; limited overall gain from metric learning; evaluated only in intrasubject setting.	S. Duan et al. (2023)

Multi-modality Deep Forest for Hand Motion Recognition via Fusing sEMG and Acceleration Signals	Ninapro DB7	20 healthy subjects and 2 transradial amputees, 40 gestures (38 for amputees), 6 trials per gesture, 12 sEMG channels + 36 ACC channels (IMU available but ACC used), 2000 Hz (sEMG) and 128 Hz (ACC). 6-fold cross-validation: five trials used for training and one trial used for testing.	Multi-Modality Deep Forest (MMDF)	91.40 ± 2.02% (healthy subjects); 77.80 ± 9.61% (amputee subjects)	The framework effectively improves recognition accuracy by integrating sEMG and acceleration signals at the feature level, reduces dependency on heavy hyperparameter tuning, and automatically determines cascade depth without backpropagation. It significantly outperforms KNN, SVM, Random Forest, and the original gcForest.	The training process is computationally expensive due to the cascade architecture. The performance on amputee subjects is still considerably lower than on healthy subjects, and robustness under electrode displacement or muscle fatigue conditions requires further validation.	Y. Fang (2023)
TinyMyo: a Tiny Foundation Model for Flexible EMG Signal Processing at the Edge	NinaPro DB5, EPN-612 (Related Domain)	Large-scale pre-training on diverse datasets.	TinyMyo (Transformer Encoder)	89.4% (DB5), 96.7% (EPN)	Ultra-low power (runs on MCU); "Foundation Model" generalization; only 3.6M params.	Not natively trained/tested on DB7 (requires fine-tuning for DB7's 12 channels).	M. Fasulo et al. (2026)
PhysioWave: A Multi-Scale Wavelet-Transformer for Physiological Signal Representation	NinaPro DB5 (Related Domain)	Large-scale physiological pre-training.	PhysioWave (Wavelet + Transformer)	State-of-the-art (DB5)	Handles non-stationarity via learnable wavelets; robust to noise.	High complexity; requires pre-training on massive external datasets.	Y. Chen et al. (2025)
sEMG-Based Gesture Recognition Using Deep Learning From Noisy Labels	NinaPro DB7	Large-scale analysis focusing on label noise.	ALR-CNN (Adaptive Label Refinement)	95.85%	Robust to the "reaction time" lag in DB7 labels; Refines labels during training.	Complex training loop; assumes label noise is the primary error source.	A. Fatayer et al. (2022)
Classification of 41 Hand and Wrist Movements via Surface Electromyogram Using Deep Neural Network	NinaPro DB7	Focuses on the high-class count (40+).	Deep Neural Network (Standard CNN)	91.69 ± 4.68%	Simple, effective baseline; Validates performance on high-density class sets.	Struggles with the specific "Functional" movements in Exercise 2.	P. Sri-iesaranusorn (2021)
A parallel and efficient transformer deep learning network for continuous estimation of hand kinematics from electromyographic signals	NinaPro DB7	Used for continuous kinematic estimation task.	PET (Parallel Efficient Transformer)	88.72%	Extremely fast inference (low latency) suitable for edge devices.	Designed for regression (angles), not discrete classification (40 classes).	C. Lin et al. (2025)
Improving Unimodal sEMG-Based Pattern Recognition Through Multimodal Generative Adversarial Learning	NinaPro DB7	41 gestures; intact subjects; sEMG + ACC	GAM (Generative Auxiliary Modality)	93.53% (Intra: 95.06%; Inter: 82.70%)	Strong intersubject improvement; Scales with more GAN data	Needs AdaBN for acceptable intersubject accuracy; training complexity increased.	W. Wei et al. (2025)
Improved Network and Training Scheme for Cross-Trial Surface Electromyography (sEMG)-	NinaPro DB1-NinaPro DB7		sEMGPoseMIM (Two-Stage Training Scheme) + sEMGXCM (Proposed sEMG Encoder Network)	NinaPro DB7: 91.2%	Effectively improves cross-trial sEMG-based gesture recognition by learning trial-invariant and discriminative representations through mutual information maximization and cross-	The method is limited to cross-trial settings, increases training complexity due to its two-stage design, and may lack robustness to electrode displacement and	Q. Dai et al. (2023)

Based Gesture Recognition					modal knowledge distillation, achieving superior performance on Ninapro DB1–DB7.	inter-subject variability.	
Calibration-free sEMG intention recognition via self-supervised pretraining and adversarial domain alignment for upper-limb rehabilitation	Ninapro DB2 CapgMyo DBa	**Ninapro DB2:** A 12-channel sEMG dataset collected from 40 able-bodied subjects performing 50 hand/wrist gestures (6 repetitions, 5 s each) at 2 kHz using Delsys Trigno electrodes, with timestamped movement labels. (only EMG channels) **CapgMyo DBa:** A 128-channel high-density sEMG dataset recorded from 18 able-bodied subjects performing 8 gestures (10 repetitions) at 1 kHz using a forearm electrode grid with 10 mm inter-electrode spacing.	Self-Supervised Temporal–Spectral Pretraining + Adversarial Domain Alignment	89.4% (DB2) 86.9% (DBa)	Improves cross-subject and cross-dataset sEMG recognition accuracy, generalization, and robustness, validated on Ninapro DB2 and CapgMyo DBa, while reducing inter-subject variability.	The framework is limited by offline evaluation, reliance on healthy-subject datasets, lack of test-time adaptation, and incomplete interpretability despite attention-based saliency analyses.	Y. Yang et al. (2025)
An efficient surface electromyography-based gesture recognition algorithm based on multiscale fusion convolution and channel attention	NinaPro DB1, DB3, and DB4		Residual-Inception-Efficient (RIE)	88.27%	Efficiently enhances multitype sEMG gesture recognition through multiscale Inception fusion and ECA-based channel attention, achieving strong generalization with reduced complexity.	It relies on large supervised datasets, was validated only offline, and struggles to distinguish highly similar gestures.	B. Jiang et al. (2024)

References:

1. Z. Chen, X. Qiao, S. Liang, T. Yan, and Z. Chen, "sEMG-Based Gesture Recognition via Multi-Feature Fusion Network," IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 4, pp. 2570–2580, 2025.
2. S. Duan, L. Wu, B. Xue, A. Liu, R. Qian, and X. Chen, "A Hybrid Multimodal Fusion Framework for sEMG-ACC-Based Hand Gesture Recognition," IEEE Sensors Journal, vol. 23, no. 3, pp. 2773–2782, 2023.
3. A. Fatayer, W. Gao, and Y. Fu, "sEMG-Based Gesture Recognition Using Deep Learning From Noisy Labels," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 9, pp. 4462–4473, 2022.
4. P. Sri-iesaranusorn, A. Chaiyaroj, C. Buekban, et al., "Classification of 41 Hand and Wrist Movements via Surface Electromyogram Using Deep Neural Network," Frontiers in Bioengineering and Biotechnology, vol. 9, 2021.
5. Y. Fang, H. Lu, and H. Liu, "Multi-Modality Deep Forest for Hand Motion Recognition via Fusing sEMG and Acceleration Signals," International Journal of Machine Learning and Cybernetics, vol. 14, no. 4, pp. 1119–1131, 2023.
6. M. Fasulo, G. Spacone, T. M. Ingólfsson, Y. Li, L. Benini, and A. Cossettini, "TinyMyo: A Tiny Foundation Model for Flexible EMG Signal Processing at the Edge," 2026.
7. Y. Chen, M. Orlandi, P. M. Rapa, S. Benatti, L. Benini, and Y. Li, "PhysioWave: A Multi-Scale Wavelet-Transformer for Physiological Signal Representation," 2025.
8. C. Lin, X. Zhang, and C. Zhao, "A Parallel and Efficient Transformer Deep Learning Network for Continuous Estimation of Hand Kinematics from Electromyographic Signals," Scientific Reports, vol. 15, 2025.