

DOI: 10.37943/V)UK7425

Darkhan Zholtayev

PhD, Assistant Professor, School of AI and Data Science

Darkhan.zholtayev@astanait.edu.kz, orcid.org/0000-0001-5347-2595,

Astana IT University, Kazakhstan

Relive Research LLP, Kazakhstan

Temirlan Meiramkhanov

Master's Degree, School of AI and Data Science

242813@astanait.edu.kz, orcid.org/0009-0003-0042-0777

Astana IT University, Kazakhstan

Aiman Ozhikenova

PhD, Associated Professor, Director

a.ozhikenova@satbayev.university, orcid.org/0000-0001-7078-5737

Qazaq Youth Science Hub LLP, Kazakhstan

Satbayev University, Kazakhstan

Zhadyra Alimbayeva

PhD, Associated Professor

zh.alimbayeva@satbayev.university, orcid.org/0000-0002-2628-2515

Kazakh National Women's Teacher Training,

Satbayev University, Kazakhstan

Ozhiken Assylbek

PhD Student. Senior Researcher

a.ozhiken@satbayev.university, orcid.org/0000-0002-2945-8213

U.A. Dzholdasbekova Institute of Mechanics and Engineering RSE, Kazakhstan

Beibit Abdikenov

PhD, Associated Professor, School of AI and Data Science

beibit.abdikenov@astanait.edu.kz, orcid.org/0000-0002-0284-0949

Astana IT University, Kazakhstan,

Relive Research LLP, Kazakhstan

ENSEMBLE-BASED EMG-IMU GESTURE RECOGNITION FOR REAL-TIME PROSTHETIC HAND CONTROL

Abstract: Reliable control of upper-limb prostheses remains challenging because muscle signals can change between sessions and wearable EMG sensors may suffer from poor electrode contact. This study presents a real-time prosthetic hand control platform based on an OYMotion gForce Pro armband and a five-finger Dynamixel-actuated prosthetic hand. The system is designed to work without an external dataset, using only a short user-specific recording session and brief calibration before real-time use — specifically, five main training repetitions plus four supplementary calibration repetitions per gesture — intentionally minimising the calibration burden to enable practical single-session deployment.

Five hand gestures—power grasp, open hand, OK gesture, index pointing, and thumbs-up—were classified using several machine-learning and ensemble-based methods. All models were evaluated under the same live data stream to ensure a fair comparison. Six recording sessions collected across five able-bodied participants on different recording days were included in the final statistical analysis. For every session, the main and supplementary training samples were combined to train session-specific models before live evaluation. Across six independent recording sessions from one able-bodied participant, Linear Discriminant Analysis achieved the best average live accuracy of **85.9%**, followed closely by majority voting and adaptive ensemble methods. In contrast, the LSTM model showed lower performance, reaching **70.2%** accuracy.

The results indicate that, for low-data real-time prosthetic control, ensemble-based methods preserved near-best average accuracy while reducing reliance on any single classifier assumption, making them a more robust deployment choice than individual classifiers. Statistical analysis confirmed that session-to-session variability was the dominant factor in system performance (Kendall's $W = 0.249$) rather than classifier choice ($W = 0.043$), suggesting that practical reliability depends more on robust calibration and signal monitoring than on model selection.

Keywords: electromyography; inertial measurement unit; pattern recognition; prosthetic hand; gesture recognition; ensemble learning; real-time control; session calibration; signal quality; rehabilitation engineering.

Introduction

Upper-limb amputation imposes significant functional, social, and economic burden on affected individuals, and the rejection rate of electric upper-limb prostheses remains high because users still face limitations in functionality, control reliability, comfort, and cost [1-2]. Surface electromyography (sEMG) is widely used as a non-invasive interface for decoding forearm muscle activity into hand or wrist commands, and recent reviews describe pattern-recognition control as one of the main routes toward more natural multi-function prosthetic control [3-4]. The appeal of this approach is practical: instead of requiring the user to operate a small number of antagonist muscle sites through an artificial switching strategy, the controller learns discriminative muscle-activation patterns associated with intended hand postures [1, 4].

In practice, however, laboratory accuracy rarely transfers directly to daily use. Recent studies identify three recurring causes. First, inter-session variability in electrode placement, skin impedance, sweat, fatigue, limb position, and gesture speed changes the feature distribution between training and deployment [2, 5-6]. Second, low-channel wearable armbands make EMG experiments accessible, but channel count and electrode placement still strongly influence online prosthetic-control performance [7-8]. Third, there is a latency-stability trade-off: short windows reduce delay but can amplify transient noise, while longer windows and decision smoothing improve stability at the cost of responsiveness [1, 9]. These constraints are especially important for consumer-grade dry-electrode systems, where signal quality can vary substantially across repetitions and sessions. A substantial body of recent work has explored deep learning and adaptation approaches to address these issues. Contemporary reviews summarize convolutional, recurrent, attention-based, and transfer-learning models for sEMG gesture recognition [3, 10], while recent studies report strong offline performance using spectrogram-based hybrid networks, time-varying feature enhancement, and random convolutional kernel transforms [11-13]. At the same time, calibration-burden reviews and cross-session amputee studies show that robust deployment depends not only on model architecture, but also on how much session-specific data is available and how well the system handles distribution shift [2, 6]. The corresponding question for low-data, single-session deployment with consumer hardware - which is what most independent researchers, clinicians, and amputees actually have access to - is less well answered. In this regime, classical feature-engineered classifiers remain attractive because they train quickly, run without GPU acceleration, and can be evaluated transparently under live conditions [1, 13].

Adding inertial measurement information to surface EMG can improve recognition robustness, particularly for gestures whose differentiating feature is wrist orientation rather than muscle activation alone. Recent reviews and multimodal studies describe sensor fusion as a practical way to compensate for some limitations of sparse sEMG, including limb-position and motion-context effects [1, 10, 14-15]. For consumer wearables that integrate both modalities - as the OYMotion gForce Pro does - the marginal cost of using IMU data is essentially zero [16].

This paper describes a complete experimental platform that integrates eight-channel sEMG and IMU acquisition, a feature-engineered classification pipeline with six distinct base models, several ensemble-based decision strategies, a short session-specific supplementary training phase,

and a real-time gating layer that drives a five-degree-of-freedom Dynamixel-actuated prosthetic hand. Crucially, the system requires no external pre-collected dataset: a short, user-driven session of five training repetitions plus four supplementary repetitions per gesture is sufficient for the platform to become operational. Six independent live evaluation sessions were conducted under a shared-target protocol in which all models received the same data stream simultaneously, eliminating cross-trial confounds.

The main novelty of this work lies in the integration and live evaluation of conventional machine-learning, ensemble-learning, and deep-learning approaches within an OYMotion gForce Pro armband-based prosthetic control system. To the authors' knowledge, ensemble-learning strategies have not been systematically evaluated for real-time control of a physical prosthetic hand using this low-cost wearable EMG bracelet. Unlike studies limited to offline gesture classification or benchmark datasets, the present work directly connects model predictions to a real Dynamixel-actuated prosthetic hand and evaluates performance under repeated live sessions. This enables a practical assessment of whether ensemble methods provide meaningful advantages over individual classifiers when sensor contact quality, session variability, and limited calibration data are present.

Novelty and Contributions

Conventional sEMG gesture-recognition studies often emphasise offline classification accuracy, which may not represent performance during continuous, real-time prosthetic operation. This study addresses that limitation through a deployment-centred experimental framework rather than by proposing another classifier. Its principal novelty is a controlled, end-to-end comparison in which the calibration budget, signal preprocessing, incoming live sEMG stream, gesture set, and prosthetic-hand interface are held constant while only the decision model is changed. This design makes it possible to separate the influence of classifier selection from the effects of session variability, calibration quality, and signal degradation.

Within this framework, six base classifiers - Linear Discriminant Analysis, Support Vector Machine, Random Forest, Multilayer Perceptron, XGBoost, and Long Short-Term Memory - and multiple ensemble decision strategies are evaluated under identical conditions. Repetition-grouped validation prevents overlapping windows from the same gesture repetition from appearing in both training and validation sets, thereby reducing information leakage. Each live model is trained using a fixed, low-data, user-specific calibration protocol and evaluated concurrently on the same continuous sEMG stream. Signal quality monitoring and confidence-aware ensemble decisions are incorporated into the real-time pipeline, linking algorithmic predictions directly to the operation of a five-finger prosthetic hand.

Accordingly, the contributions of this study are: (1) a reproducible shared-stream protocol for comparing heterogeneous classifiers under identical real-time conditions; (2) an evaluation of classical, deep-learning, and ensemble methods under a deliberately restricted user-specific calibration budget; (3) an end-to-end assessment that connects sEMG acquisition, gesture recognition, decision fusion, signal-quality monitoring, and physical hand actuation; and (4) statistical evidence that recording-session conditions and calibration reliability can influence deployable performance more strongly than classifier complexity. The resulting contribution is therefore not a new classification architecture, but experimentally supported guidance for selecting and deploying gesture-recognition models in low-data, wearable-sEMG prosthetic-control systems.

Methods and Materials

Hardware Configuration

Surface EMG and inertial signals were acquired with the OYMotion gForce Pro armband (model gForce100 EMG+IMU) [16]. The device provides eight channels of dry-electrode surface EMG, configurable to 500 hertz sampling rate at 8-bit amplitude resolution, alongside a quaternion stream at approximately 50 hertz. Communication is over Bluetooth Low Energy (BLE). For the experiments reported here, the channel mask was set to 0xFF (all eight channels active) and the data length to 128 samples per BLE notification. On platforms where BLE scanning was

unreliable, direct connection by hardware MAC address was used to bypass scanning latency. The gForce Pro was the only sensor used in this study; no additional EMG, force, or motion sensors were employed.



a) b)

Figure 1: OYMotion gForce Pro armband sensor, Tendon driven prosthetic hand

Actuation was provided by five Dynamixel AX-12A-compatible servos (Robotis, [17]), one per finger, communicating over a shared half-duplex serial bus at 1 Mbit/s using Protocol 2.0. Each gesture was mapped to a five-tuple of goal positions in the 0 - 300° range (encoded internally as 0–1023 raw position units), validated by physical inspection of the prototype hand prior to data collection. Smooth transitions between gestures were generated by linear interpolation across eight intermediate set-points with a 35 ms inter-step delay, yielding an approximately 280 ms transition time. Position verification was performed by reading back the present-position register and retrying up to three times if the residual error exceeded 10°.

Prosthetic Hand Platform

The prosthetic hand used in this study is a third-party device employed solely as the actuation platform for the recognized gestures. Its mechanical design and fabrication are not part of the contribution of this paper; the contribution is the recognition algorithm and the real-time control pipeline that drives any compatible Dynamixel-actuated hand. Actuation is provided by five Dynamixel AX-12A-compatible servos (Robotis, [17]), one per finger, communicating over a shared half-duplex serial bus at 1 megabit per second using Dynamixel Protocol 1.0.

Each finger is mechanically underactuated, with one servo driving a tendon routed through three rotational joints. The angular configuration of each finger is represented by $q = [q_1, q_2, q_3]^T$, where q_1 , q_2 , and q_3 denote the flexion angles of the first, second, and third finger joints. Following the three-link tendon-driven finger model described by Suleimenov et al [18], the finger dynamics can be represented as:

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + g(q) + B\dot{q} + K_s(q - q_0) = \tau T - \tau_{ext}. \quad (1)$$

Here, $M(q)$ is the inertia matrix of the three links, $C(q, \dot{q})$ represents Coriolis and centrifugal effects, $g(q)$ is the gravitational-torque vector, and $B\dot{q}$ represents viscous joint friction. The term $K_s(q - q_0)$ represents the torque generated by the joint return springs, where K_s is the spring-stiffness matrix and q_0 is the equilibrium or extended finger configuration. The vectors τT and τ_{ext} represent the tendon-generated torque and external contact torque, respectively.

The tendon-generated joint-torque vector is approximated as:

$$\tau T = [r_1 T_1, r_2 T_2, r_3 T_3]^T, \quad (2)$$

where r_1 , r_2 , and r_3 are the effective pulley radii, while T_1 , T_2 , and T_3 are the tendon tensions acting at the three joints. The tendon tensions are related to the motor torque τ_m by:

$$T_1 = \tau_m / [r_1 \exp(\mu q_1)], \quad (3)$$

$$T_2 = \tau_m / [r_1 \exp(\mu(q_1 + q_2))], \quad (4)$$

$$T_3 = \tau_m / [r_1 \exp(\mu(q_1 + q_2 + q_3))], \quad (5)$$

where μ is the effective tendon–pulley friction coefficient. These relationships describe how one servo produces coupled flexion torques at all three finger joints. When the tendon is released, the joint springs return the finger toward its extended configuration. Because the present study considers relatively slow gesture transitions rather than high-speed finger dynamics, the mechanical behaviour may also be approximated by the quasi-static relationship:

$$g(q) + Ks(q - q_0) \approx \tau T - \tau_{ext}. \quad (6)$$

Each recognized gesture is mapped to a five-tuple of servo goal positions, with one position assigned to each finger. The goal positions cover the servo's 0–300-degree operating range and are encoded internally as values from 0 to 1023. All gesture configurations were validated by physical inspection before data collection. Smooth transitions between gestures are generated by linear interpolation across eight intermediate set-points with a 35-millisecond inter-step delay, resulting in an approximate transition time of 280 milliseconds. Position verification is performed by reading the present-position register from each servo and retrying the position command up to three times when the residual error exceeds 10 degrees.

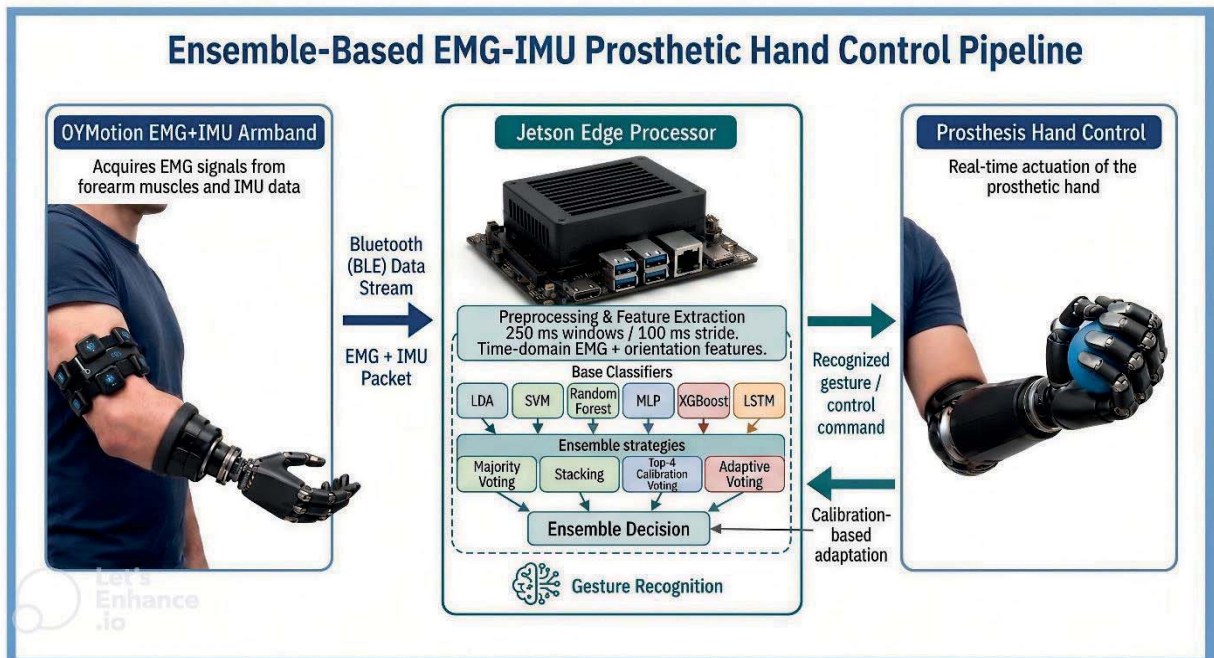


Figure 2: Block scheme of an EMG-based prosthetic hand control system using an OYMotion

EMG bracelet, Jetson edge processor, and an ensemble of machine-learning algorithms for real-time gesture recognition and prosthetic hand actuation

Gesture Set

Five functional hand postures were selected to cover common grasping, pointing, precision, opening, and communicative patterns relevant to activities of daily living. Recent prosthetic-hand and grasp-taxonomy work continues to emphasise that practical gesture sets should reflect daily tasks while remaining small enough for reliable training and real-time control [10, 19-21]. The gesture set is deliberately small. Increasing the number of classes is straightforward in software but increases the per-class data demand at training time, which is the binding constraint in single-session deployment. Each gesture is summarized below, including its kinematic definition and its representative everyday use case.

Power grasp. All four fingers flexed around the palm with the thumb opposed across the fingers. This is the dominant grasp for cylindrical and spherical objects — holding a bottle, a cup, a hammer handle, a doorknob, or a ball — and is the highest-amplitude EMG class in the set, since it engages the entire flexor digitorum group simultaneously.

Open hand. All five digits are extended, with the palm open and the thumb abducted from the hand. This posture represents hand opening and object release, which are essential actions in prosthetic-hand control. In daily use, it corresponds to releasing a grasped object, preparing the hand before grasping, or placing the prosthetic hand in a neutral open configuration. It also provides a kinematically distinct class from the flexion-dominant power grasp.

OK gesture (opposition pinch). The thumb tip and index fingertip touch or nearly touch to form an “OK” posture, while the remaining fingers are extended or relaxed according to the implemented prosthetic-hand pose. This gesture approximates a precision/opposition pinch and represents fine manipulation tasks such as picking up a coin, a small button, a screw, or a pill. In the implementation, this class is labeled as ok.

Index point. The index finger is fully extended while the remaining three fingers and the thumb are flexed into the palm. This is the gesture used for pressing buttons (elevator, keypad, touchscreen), pointing to objects or directions, and operating single-key controls. As a discriminator it leans heavily on the differential activation of the extensor indicis against a flexed-finger background.

Thumbs-up. The thumb is fully extended while the four fingers are flexed into the palm. Beyond its communicative role, this gesture was deliberately included because its differentiating feature is largely the wrist and thumb orientation rather than the muscle activation pattern, which makes it an effective test case for the contribution of the IMU channels to overall classification.

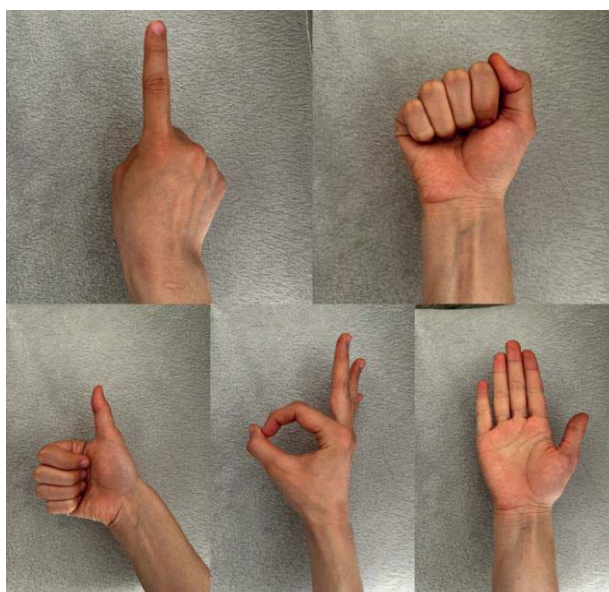


Figure 3: Experimental gesture set used in the study: index point, power grasp, thumbs-up, OK gesture, open hand.

On-Device Data Collection Protocol

Training data were collected across multiple on-device recording sessions involving different subjects and different recording days. Within each session, data were collected in interleaved repetitions. For each of five repetitions per gesture, all five gestures were presented in sequence rather than blocked by gesture. Interleaving was used to distribute time-of-session effects, such as fatigue, electrode drying, and attention drift, approximately uniformly across classes. Each repetition followed a fixed schedule consisting of an eight-second warmup, during which the subject positioned the arm and read the visual cue, a seven-second active capture window, during which EMG and IMU signals were recorded, and a five-second rest interval. At a 500 Hz sampling rate, the seven-second capture window yielded an expected 3500 samples per repetition. Observed sample ratios remained close to the expected value across sessions, indicating that BLE throughput was not a limiting factor during data acquisition.

After the main training phase, a short supplementary training phase was collected using the same gesture set. This phase followed a shorter protocol consisting of a three-second warmup, a four-second active capture window, and a two-second rest interval, repeated four times per gesture. These supplementary samples were not treated as an independent validation or test set. Instead, they were concatenated with the main training data, and the final live model was trained on the combined session-specific dataset before live evaluation. The shorter supplementary phase was included as a practical way to add additional labeled windows from the same recording session while limiting the extra data-collection burden before live control.

Preprocessing and Windowing

Preprocessing was deliberately minimal. Both EMG and IMU buffers underwent direct-current removal by subtracting the per-channel mean of each window immediately before feature extraction. No bandpass filtering, notch filtering, or rectification was applied. This choice reflects a practical constraint: at 8-bit amplitude resolution and 500-hertz sampling, causal infinite-impulse-response filters introduce ringing and quantization artefacts that, in our pilot experiments, did not improve discrimination over simple mean subtraction.

Sliding analysis windows of 250 ms were extracted at a 100 ms stride. At the nominal 500 Hz EMG sampling rate, each EMG window contained 125 samples. The quaternion stream was sampled at an estimated rate of approximately 50 Hz, so the same window duration corresponded to approximately 12 IMU frames. For each repetition, EMG and IMU buffers were adjusted to the expected capture duration, windowed separately, and then paired by window index. If the two modalities produced different numbers of windows, only the common paired window count was used for feature extraction.

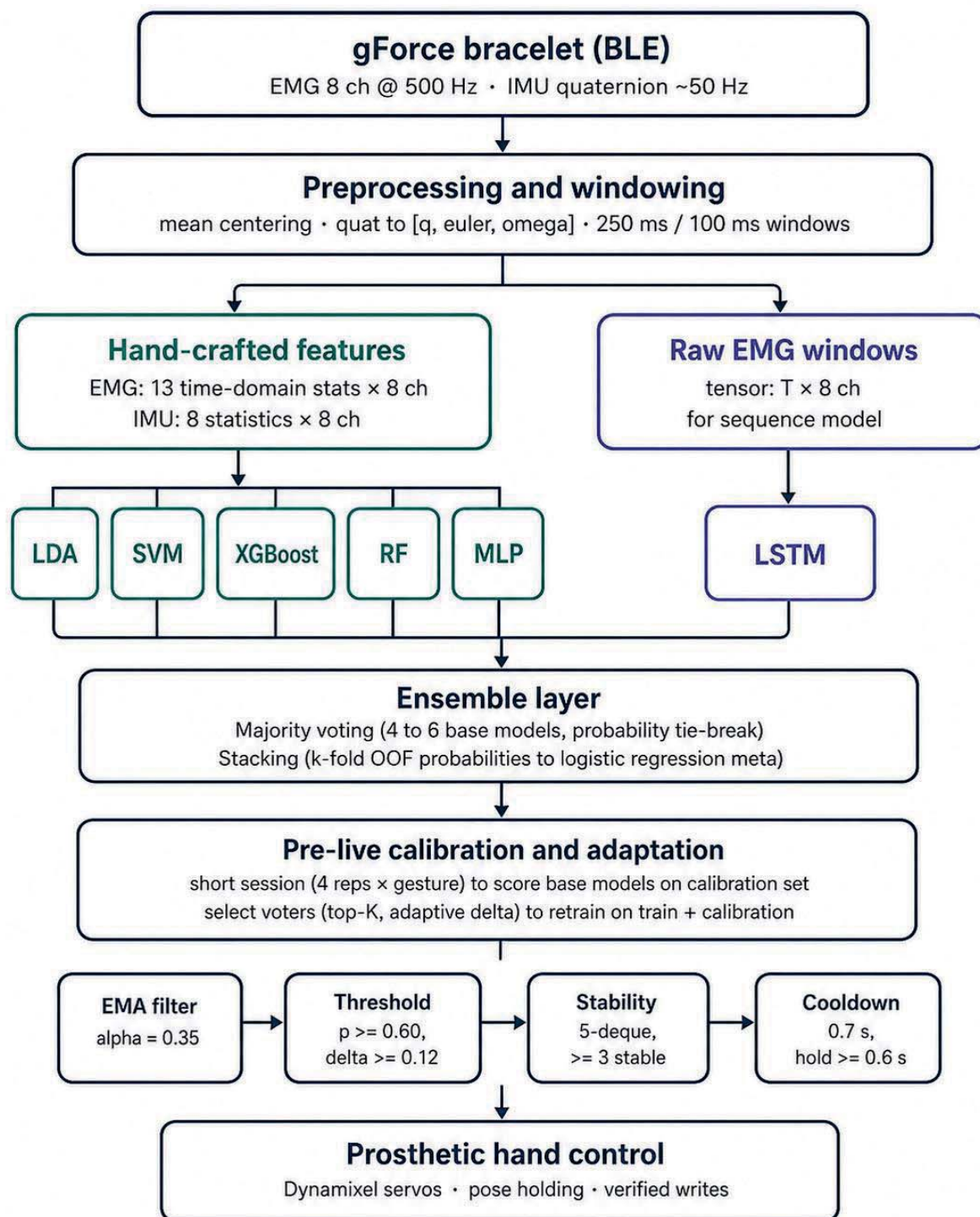


Figure 4: Ensemble ML model block scheme

Feature Engineering

Thirteen handcrafted time-domain features were computed per EMG channel: mean absolute value, root mean square, variance, waveform length, zero-crossing count with a 3.0 ADC-count deadband, slope sign change count with the same deadband, minimum, maximum, peak-to-peak amplitude, mean, standard deviation, log-variance, and integrated absolute EMG. With eight active EMG channels, this produced a 104-dimensional EMG feature vector. This compact feature

set reflects the emphasis in recent applied studies on interpretable handcrafted descriptors, feature evaluation, and electrode-site selection for real-time myoelectric pattern recognition [22-23].

For IMU data, the quaternion stream was converted into an eight-dimensional representation consisting of four quaternion components, three Euler angles, and angular speed. Eight statistical features were then computed for each IMU-derived channel: mean, standard deviation, minimum, maximum, peak-to-peak amplitude, root mean square, energy, and waveform length. This produced a 64-dimensional IMU feature vector. The concatenated feature vector used by the classical classifiers therefore had dimensionality 168.

No frequency-domain or wavelet features were used. Given the short window duration and the low amplitude resolution of the EMG stream, the study focused on compact time-domain descriptors suitable for real-time inference.

Classifiers

Six base classifiers and four ensemble variants were evaluated. The base classifiers were Linear Discriminant Analysis with default scikit-learn settings; a Support Vector Machine with a radial-basis-function kernel, regularization parameter $C = 3.0$, and probability estimates enabled; a Random Forest with 150 trees, no maximum depth, a minimum-samples-per-leaf value of 2, and balanced class weights; a Multilayer Perceptron with two hidden layers of 64 and 32 units, rectified linear activation, Adam optimization, learning rate $1e-3$, L2 regularization coefficient $1e-3$, and early stopping with 25-iteration patience on a 20% validation fraction; XGBoost with 160 estimators, maximum tree depth 3, learning rate 0.05, subsampling 0.9, and column subsampling 0.9; and a single-layer LSTM network with hidden size 64 and a two-layer classification head with dropout 0.20, trained for 24 epochs with Adam and weight decay $1e-4$. Unlike the classical models, the LSTM consumed the raw windowed EMG sequence rather than the handcrafted EMG-IMU feature vector and did not receive IMU input.

The four ensemble variants were full majority voting across all six base classifiers (voting_all), probability stacking with a logistic-regression meta-learner trained on out-of-fold predictions (stacking_all), a top-four supplementary-block voting ensemble (voting_top4_calib), and an adaptive supplementary-block voting ensemble (voting_adaptive_calib). For the two supplementary-block ensembles, candidate base classifiers were first trained on the main training data and ranked using the short supplementary pre-live block before that block was merged with the main training set for final live-model training. The adaptive variant included any base classifier within five percentage points of the top supplementary-block accuracy, subject to a minimum of three members.

Group-Aware Validation

Offline validation used GroupShuffleSplit with ten splits and a 0.25 validation fraction, with the repetition index as the grouping key. This means that all overlapping windows extracted from the same repetition are constrained to land exclusively in either the training or the validation partition, never both. Without this constraint, the overlap between adjacent windows can produce near-duplicate samples that, when randomly distributed, inflate validation accuracy and do not represent performance on new repetitions. This concern is consistent with recent work showing that cross-day use, limb position, electrode placement, channel configuration, and other confounding factors can degrade real-world myoelectric control even when offline performance is high [2, 5-6, 8]. In pilot runs, window-level random splits produced higher validation scores than group-aware splits and were therefore not used for model selection.

Real-Time Decision Layer

The live prediction loop ran at 5 Hz. At each tick, the most recent 800 ms of EMG and IMU buffer content was extracted, segmented into six overlapping analysis windows, classified, and aggregated into a per-tick probability distribution. Raw per-tick predictions were then passed through a five-stage gating cascade before any motor command was issued. First, class probability vectors were smoothed with an exponential moving average with a smoothing factor of 0.35. Second, the smoothed top-class probability had to exceed 0.60 and the margin between the top and

second-best classes had to exceed 0.12; otherwise, the tick was treated as uncertain and the current commanded gesture was held. Third, the same accepted symbol had to appear in three consecutive confident ticks, corresponding to 0.6 s at 5 Hz. Fourth, accepted symbols populated a five-element vote queue, and a stable candidate was declared when one symbol occupied at least three of the five queue positions and matched the current consecutive symbol. Fifth, once a gesture was committed, a 0.60 s minimum hold time had to elapse before any state change, and a 0.70 s cooldown after each commit suppressed command chatter.

Once committed, the hand held the corresponding pose and did not return to a neutral position between gestures. A 220 ms pose-refresh cycle re-issued the goal-position commands to all five servos to compensate for mechanical drift or torque-induced position error. The trade-off relative to a return-to-neutral paradigm is that misclassified transitions persist until corrected by a new stable prediction, but the user is spared the disruptive behavior of the hand dropping to neutral between intended gestures.

Evaluation Sessions

Six independent recording sessions were included in the final analysis: 23-04-1, 24-04-03, 24-04-1, 24-04-2, 25-04-1, and 25-04-2. These sessions were collected across **five participants** (able-bodied), spanning different recording days, using the on-device data-collection protocol described above. For each session, the main training and supplementary training data were combined to train session-specific live models before evaluation.

Live evaluation was performed using a shared-target protocol. During this protocol, a target gesture was displayed on the monitor, and the participant reproduced the instructed gesture in real time. Each target gesture was shown for five seconds. The first one second of each trial was excluded from scoring to allow the participant to form the gesture and to reduce transition effects. During the remaining four seconds, all ten classifiers simultaneously processed the same incoming EMG and IMU data stream and produced predictions. The predicted gesture labels were compared with the target command displayed on the monitor. A prediction was counted as correct when the model output matched the displayed target gesture. Each gesture was presented five times in randomized order, producing 25 target trials per session. Only target-aligned prediction ticks marked as scored were used for accuracy calculation. This resulted in 124 or 125 scored predictions per model per session and 745 scored predictions per model across the six sessions.

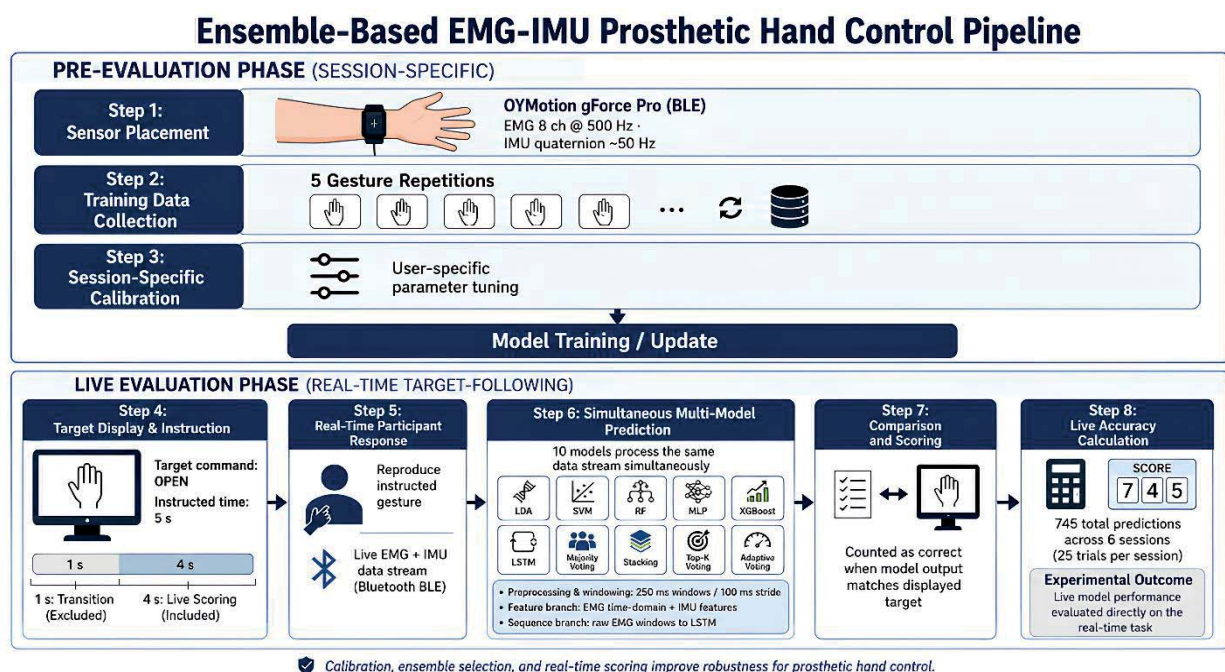


Figure 5. Experimental protocol used for each recording session.

Results

Cross-Session Live Accuracy

Although LDA achieved the highest mean live accuracy, full majority voting provided a more balanced deployment profile. Its mean accuracy was only 0.4 percentage points lower than LDA (85.5% versus 85.9%), while its minimum session accuracy was slightly higher (70.2% versus 69.3%). This distinction is important for real-time prosthetic control, where robustness across sessions may be more valuable than the highest average score alone. Reliance on a single classifier family is risky in this setting because the EMG/IMU feature distribution can change across recording sessions due to armband placement, electrode contact, muscle activation strength, fatigue, and subject-specific movement patterns. A single classifier encodes one dominant inductive bias, whereas majority voting combines heterogeneous decision rules from multiple model families. Therefore, full majority voting was considered the preferred practical deployment choice: it preserved near-best average performance, slightly improved the worst-session result relative to the best single classifier, and reduced dependence on any one classifier assumption.

For statistical comparison, live predictions were aggregated at the target-trial level rather than at the individual prediction-tick level. Each target trial within each session formed one repeated-measures block, and the dependent variable was the proportion of correct raw predictions within that block. Classifier differences were tested with the Friedman test across 150 target-trial blocks (6 sessions $\times$ 25 trials per session), followed by post-hoc pairwise Wilcoxon signed-rank tests with Bonferroni correction

Table 1. Cross-session live raw accuracy across six recording sessions. Each model was evaluated under the shared-target protocol on the same incoming EMG and IMU stream. SD is the standard deviation across six sessions. N scored is the total number of scored prediction ticks aggregated across sessions. Group letters are based on target-trial-level Friedman testing ($\chi^2(9) = 58.19$, $p < 0.001$, Kendall's $W = 0.043$) followed by post-hoc pairwise Wilcoxon signed-rank tests with Bonferroni correction; models sharing a letter do not differ significantly at $\alpha = 0.05$.

Model	Sessions	Mean	SD	Median	Min	Max	N scored	Group
LDA	6	85.9 %	10.3 %	87.1 %	69.3 %	99.2 %	745	a
Voting (all)	6	85.5 %	10.6 %	85.9 %	70.2 %	99.2 %	745	a
Voting adaptive	6	85.1 %	9.8 %	85.1 %	69.3 %	98.4 %	745	a
Stacking (all)	6	84.2 %	10.8 %	83.1 %	72.6 %	99.2 %	745	a
Voting top-4	6	83.6 %	10.4 %	83.5 %	70.2 %	98.4 %	745	a
Random Forest	6	83.1 %	10.4 %	80.7 %	71.8 %	98.4 %	745	a
MLP	6	83.1 %	14.5 %	85.5 %	58.9 %	99.2 %	745	a
SVM	6	82.8 %	12.2 %	82.3 %	67.7 %	100.0 %	745	a

XGBoost	6	81.6 %	12.1 %	79.1 %	65.3 %	98.4 %	745	a
LSTM	6	70.2 %	18.0 %	70.2 %	44.4 %	95.2 %	745	b

The Friedman test indicated a significant classifier effect ($\chi^2(9) = 58.19$, $p < 0.001$, Kendall's $W = 0.043$). Post-hoc Wilcoxon tests with Bonferroni correction showed that LSTM was significantly lower than each feature-engineered classifier and ensemble variant, whereas differences among the feature-engineered models and ensemble methods were not significant after correction. The low Kendall's W for classifier (0.043) indicates that, while the overall effect was statistically significant, the magnitude of classifier-related variance was small.

Multimodal LSTM Comparison

In the main evaluation, the LSTM received raw EMG sequences while the feature-engineered classifiers received combined EMG and IMU features. To verify that the lower LSTM performance was not an artefact of this input asymmetry, we conducted an additional comparison in which the LSTM was provided with the identical multimodal input (16 channels: 8 EMG + 8 IMU, `seq_mode = emg_imu`). Two further live sessions were recorded under this matched-input configuration, contributed by two participants who had also taken part in the main six-session evaluation. In both sessions, the LSTM remained below every feature-engineered classifier and ensemble variant despite receiving equivalent multimodal information: in the first session it reached 89.3% live accuracy, against 91.3–96.0% for the classical and ensemble models (best: `voting_all`, 96.0%); in the second session it reached 72.5%, against 79.9–85.2% (best: `stacking`, 85.2%).

This result indicates that the lower LSTM performance is not caused by input asymmetry, since providing the same multimodal features did not close the gap, which persisted across both participants. The finding is consistent with a data-quantity limitation rather than a feature limitation, as each session provides only 45–81 labelled windows — sufficient for feature-engineered classifiers but limited for a sequence model that must learn temporal dependencies from raw input.

An offline comparison on held-out repetitions further supports this interpretation. When the LSTM was trained with and without IMU features, the effect of adding IMU was inconsistent across sessions — improving accuracy in one session (77.2% to 85.1%) while reducing it in the other (77.4% to 60.1%). This instability is itself characteristic of a data-limited regime: a model with sufficient training data would be expected to benefit consistently from additional informative features, whereas here the same features help in one session and hurt in another, indicating that the LSTM operates below its data requirement.

Per-Session Variability

Gesture-Level and Session-Level Effects

A Friedman test across gesture classes revealed a significant effect of gesture on accuracy ($\chi^2(4) = 23.17$, $p < 0.001$, Kendall's $W = 0.193$). Table 2 summarises mean accuracy per gesture class averaged across all models and sessions. Power grasp and lateral pinch were the most reliably recognised gestures (94.2% and 90.8%, respectively), while index pointing showed the lowest accuracy (74.1%). Post-hoc Wilcoxon tests with Bonferroni correction identified three significant pairwise differences: power vs. index (adjusted $p = 0.017$), lateral vs. OK (adjusted $p = 0.035$), and power vs. thumbs-up (adjusted $p = 0.043$).

Table 2. Mean live accuracy per gesture class averaged across all models and sessions. Friedman $\chi^2(4) = 23.17$, $p < 0.001$, Kendall's $W = 0.193$. Adjusted p -values are from post-hoc Wilcoxon signed-rank tests with Bonferroni correction.

Gesture	Mean accuracy	Significant comparisons
Power grasp	94.2%	> index ($p = 0.017$), > thumbs-up ($p = 0.043$)
Open hand	90.8%	> ok ($p = 0.035$)
Thumbs-up	77.6%	None after Bonferroni correction
OK gesture	76.3%	None after Bonferroni correction
Index pointing	74.1%	Lowest mean accuracy

A Friedman test across sessions also revealed a significant session effect ($\chi^2(5) = 31.10$, $p < 0.001$, Kendall's $W = 0.249$). Table 3 summarises session-level results including mean accuracy, best model per session, and Friedman-based significance. The Kendall's W for session (0.249) was substantially larger than that for classifier (0.043), confirming that session-dependent factors — including armband placement, electrode contact, and signal quality — were the dominant source of performance variation. Two sessions differed significantly from the best-performing session after Bonferroni correction: 24-04-03 vs. 24-04-1 (adjusted $p = 0.001$) and 24-04-03 vs. 25-04-1 (adjusted $p = 0.006$).

Table 3. Per-session live evaluation summary. Friedman $\chi^2(5) = 31.10$, $p < 0.001$, Kendall's $W = 0.249$.

Session	Scored per model	Mean accuracy	Best model	Sig. vs 24-04-03
24-04-03	124	98.6%	SVM 100.0%	—
23-04-1	124	88.6%	Voting all 93.5%	n.s.
24-04-2	125	85.9%	MLP 92.8%	n.s.
25-04-2	124	78.3%	RF 85.5%	n.s.
25-04-1	124	77.4%	Voting all 83.1%	$p = 0.006$
24-04-1	124	66.8%	SVM 73.4%	$p = 0.001$

End-to-End

Latency End-to-end latency was measured by timestamp logging from muscle-activation onset to servo-movement completion across two dedicated timing sessions. Servo actuation was consistent at a median of 168 ms (163–178 ms). Total end-to-end latency was typically 480–530 ms under clean signal conditions, with a longer tail of up to approximately 2–3 s when inconsistent initial predictions required additional gating cycles before commitment. This bimodal profile reflects the intended stability-over-speed design of the decision-gating cascade.



Figure 5: Experimental results: Gestures in corresponding order: ok, power, thumbs up, index and open hand gestures

Discussion

Session Variability and Practical Model Selection

The live evaluation results indicate that the main challenge is not simply selecting a more complex classifier. The strongest feature-engineered methods showed similar average performance, whereas differences between recording sessions were larger. This was supported by the higher Kendall's W for session effects (0.249) compared with classifier effects (0.043), suggesting that session-dependent factors such as electrode placement, contact quality, cross-day variation, and gesture execution were a larger source of performance variation among the tested factors [2, 5-6].

This finding affects model selection. Although LDA achieved the highest mean accuracy, full majority voting remains a practical deployment choice because it preserves near-best performance while reducing reliance on a single classifier. In a new session, the feature distribution may change depending on sensor placement and gesture execution; combining heterogeneous decision rules can therefore provide a more balanced response when the best single classifier is not known in advance.

The lower LSTM accuracy should not be interpreted as evidence that recurrent models are unsuitable for EMG control. In this implementation, the LSTM used only raw EMG windows, whereas the classical models used handcrafted EMG features combined with IMU-derived features. Under the present data scale and sensor resolution, input representation and session stability likely mattered more than sequence-model capacity.

Implications for Practical Prosthetic Deployment

For practical EMG-based prosthetic control with consumer wearable sensors, these results suggest that improving session consistency may be more important than increasing classifier complexity. Engineering effort should therefore prioritise stable armband placement, reliable electrode-skin contact, signal-quality monitoring, and short session-specific adaptation procedures [2, 6-7].

The achieved live accuracy of 85–86% compares favourably with real-time benchmarks in the myoelectric pattern-recognition literature, although direct comparisons should be interpreted cautiously because studies differ in participants, gesture sets, sensors, and evaluation protocols. For example, Abbaspour et al. [24] reported a gap of approximately 28 percentage points between the best offline result (LDA, 97.9%) and the best real-time result (MLP, 69.8%), illustrating the common degradation from offline validation to real-time control.

However, accuracy alone is not sufficient to establish safety for daily prosthetic use. In the present implementation, a single misclassified prediction does not immediately trigger a new committed pose because the controller holds the last accepted gesture until a subsequent stable prediction is accepted. The multi-stage gating cascade — including a 0.60 probability threshold, a three-tick confirmation window, a five-element vote queue, and a 0.70 s cooldown — suppresses spurious activations. Further validation with amputee participants, longer sessions, and task-based safety metrics is still required before drawing conclusions about unsupervised real-world deployment.

Limitations

Several limitations should be acknowledged. First, the participant sample consisted of five able-bodied users across six recording sessions. The limited number of participants is a deliberate consequence of the system's per-user calibration design: because each session produces a user-specific model trained from scratch, the approach does not rely on cross-subject generalisation and is not intended to. Nevertheless, cross-subject robustness — and especially generalisation to amputees with altered residual-limb anatomy and different muscle activation patterns — was not formally evaluated and remains an important direction for future work.

Second, the protocol evaluated session-specific live performance rather than long-term cross-day deployment with repeated removal and re-wearing of the armband by the same user. Such re-donning scenarios are challenging because electrode placement and skin contact can change substantially between days, and recent studies identify these distribution shifts as a major barrier to robust myoelectric control [2, 5-6]. Therefore, the reported results should be interpreted as session-specific live control performance rather than evidence of calibration-free long-term robustness.

Third, the gesture set was limited to five classes. While the software pipeline can support additional gestures, increasing the number of classes would raise the per-class data requirement and may change the relative ranking of classifiers. Finally, the LSTM baseline was implemented as an EMG-only raw-sequence model, whereas the classical classifiers used handcrafted EMG/IMU features; therefore, the LSTM result should not be interpreted as a general conclusion about recurrent models for multimodal EMG control.

Conclusion

This paper presented a complete experimental platform for real-time five-gesture prosthetic hand control using consumer-grade EMG and IMU hardware. The platform integrates on-device data collection, group-aware offline validation, multiple classical and ensemble classifiers, multi-stage real-time decision gating, and direct Dynamixel actuator control. No external EMG dataset is required: each live model is trained from a short session-specific protocol consisting of five main training repetitions and four supplementary training repetitions per gesture.

Across six independent recording sessions and ten classifiers evaluated under a shared-target live protocol, the strongest feature-engineered models achieved similar live accuracy. LDA obtained the highest mean accuracy at 85.9%, while full majority voting remained close at 85.5% and provided a slightly higher minimum session accuracy. Statistical analysis confirmed that all feature-engineered classifiers and ensemble variants formed a statistically homogeneous group, whereas the LSTM baseline differed significantly from all other models (Friedman $\chi^2(9) = 58.19$, $p < 0.001$, Kendall's $W = 0.043$). Because no single classifier was consistently optimal across sessions, and because the effect of session on accuracy (Kendall's $W = 0.249$) exceeded the effect of classifier choice ($W = 0.043$), majority voting is a practical deployment choice when the optimal

model cannot be known in advance for a new recording session. The lower LSTM accuracy reflects the specific EMG-only raw-sequence implementation used in this study and should not be interpreted as a general limitation of recurrent models for EMG control.

Future work should extend the platform to larger multi-subject and amputee studies, repeated cross-day re-donning scenarios, and larger gesture sets. Adaptation methods such as electrode-quality monitoring, automated retraining triggers, and feature-distribution alignment may help translate session-specific live performance into more robust daily-use prosthetic control.

References and notes: This work was supported by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan under Grant BR24992820.

References

- [1] Ni, S., Al-qaness, M. A. A., Hawbani, A., Al-Alimi, D., Abd Elaziz, M. E., & Ewees, A. A. (2024). A survey on hand gesture recognition based on surface electromyography: Fundamentals, methods, applications, challenges and future trends. *Applied Soft Computing*, 166, 112235. <https://doi.org/10.1016/j.asoc.2024.112235>
- [2] Wang, X., Ao, D., & Li, L. (2024). Robust myoelectric pattern recognition methods for reducing users' calibration burden: Challenges and future. *Frontiers in Bioengineering and Biotechnology*, 12, 1329209. <https://doi.org/10.3389/fbioe.2024.1329209>
- [3] Xiong, D., Zhang, D., Zhao, X., & Zhao, Y. (2021). Deep learning for EMG-based human-machine interaction: A review. *IEEE/CAA Journal of Automatica Sinica*, 8(3), 512-533. <https://doi.org/10.1109/JAS.2021.1003865>
- [4] Zheng, M., Crouch, M. S., & Eggleston, M. S. (2022). Surface electromyography as a natural human-machine interface: A review. *IEEE Sensors Journal*, 22(10), 9198-9214. <https://doi.org/10.1109/JSEN.2022.3165988>
- [5] Eddy, E., Campbell, E., Bateman, S., & Scheme, E. (2024). Understanding the influence of confounding factors in myoelectric control for discrete gesture recognition. *Journal of Neural Engineering*, 21(3). <https://doi.org/10.1088/1741-2552/ad4915>
- [6] Wang, H., Li, N., Gao, X., Jiang, N., & He, J. (2024). Analysis of electrode locations on limb condition effect for myoelectric pattern recognition. *Journal of NeuroEngineering and Rehabilitation*, 21, 177. <https://doi.org/10.1186/s12984-024-01466-y>
- [7] Yao, P., Wang, K., Xia, W., Guo, Y., Liu, T., Han, M., Gou, G., Liu, C., & Xue, N. (2024). Effects of training and calibration data on surface electromyogram-based recognition for upper limb amputees. *Sensors*, 24(3), 920. <https://doi.org/10.3390/s24030920>
- [8] Simon, A. M., Newkirk, K., Miller, L. A., Turner, K. L., Brenner, K., Stephens, M., & Hargrove, L. J. (2024). Implications of EMG channel count: Enhancing pattern recognition online prosthetic testing. *Frontiers in Rehabilitation Sciences*, 5, 1345364. <https://doi.org/10.3389/fresc.2024.1345364>
- [9] Xu, H., & Xiong, A. (2021). Advances and disturbances in sEMG-based intentions and movements recognition: A review. *IEEE Sensors Journal*, 21(12), 13019-13028. <https://doi.org/10.1109/JSEN.2021.3068521>
- [10] Li, W., Shi, P., & Yu, H. (2021). Gesture recognition using surface electromyography and deep learning for prostheses hand: State-of-the-art, challenges, and future. *Frontiers in Neuroscience*, 15, 621885. <https://doi.org/10.3389/fnins.2021.621885>
- [11] Rezaee, K., Khavari, S. F., Ansari, M., Zare, F., & Alizadeh Roknabadi, M. H. (2024). Hand gestures classification of sEMG signals based on BiLSTM-metaheuristic optimization and hybrid U-Net-MobileNetV2 encoder architecture. *Scientific Reports*, 14, 31257. <https://doi.org/10.1038/s41598-024-82676-1>
- [12] Shin, J., Miah, A. S. M., Konnai, S., Takahashi, I., & Hirooka, K. (2024). Hand gesture recognition using sEMG signals with a multi-stream time-varying feature enhancement approach. *Scientific Reports*, 14, 22061. <https://doi.org/10.1038/s41598-024-72996-7>

- [13] Ovadia, D., Segal, A., & Rabin, N. (2024). Classification of hand and wrist movements via surface electromyogram using the random convolutional kernels transform. *Scientific Reports*, 14, 4134. <https://doi.org/10.1038/s41598-024-54677-7>
- [14] Tacca, N., Dunlap, C., Donegan, S. P., Hardin, J. O., Meyers, E., Darrow, M. J., Colachis, S. IV, Gillman, A., & Friedenberg, D. A. (2024). Wearable high-density EMG sleeve for complex hand gesture classification and continuous joint angle estimation. *Scientific Reports*, 14, 18564. <https://doi.org/10.1038/s41598-024-64458-x>
- [15] Zhang, S., Zhou, H., Tchantchane, R., & Alici, G. (2024). A wearable human-machine-interface system based on colocated EMG-pFMG sensing for hand gesture recognition. *IEEE/ASME Transactions on Mechatronics*. <https://doi.org/10.1109/TMECH.2024.3386929>
- [16] OYMotion Technologies Co., Ltd. (2026). gForcePro+ EMG armband: Technical specifications and SDK documentation. Retrieved May 18, 2026, from <https://www.oymotion.com/product32/149>
- [17] ROBOTIS Co., Ltd. (2026). DYNAMIXEL AX-12A e-Manual. Retrieved May 18, 2026, from <https://emanual.robotis.com/docs/en/dxl/ax/ax-12a/>
- [18] Suleimenov, K., Kapsalyamov, A., Abdikenov, B., Ozhikenova, A., Igembay, Y., & Ozhikenov, K. (2025). Comparative Analysis of Model-Based and Data-Driven Control for Tendon-Driven Robotic Fingers. *Mathematics*, 13(22), 3669, <https://doi.org/10.3390/math13223669/>
- [19] Thornton, C. J., Chappell, M. J., & Evans, N. D. (2024). Portable markerless hand motion capture system for determining the grasps performed in activities of daily living. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11, 2296625. <https://doi.org/10.1080/21681163.2023.2296625>
- [20] Tchimini, J., Hansen, R. L., Jorgensen, P. H., Dideriksen, J., & Dosen, S. (2024). Application of EMG feedback for hand prosthesis control in high-level amputation: A case study. *Scientific Reports*, 14, 31676. <https://doi.org/10.1038/s41598-024-80828-x>
- [21] Pawar, B., & Munjala, M. (2022). A systematic review on available technologies and selection for prosthetic arm restoration. *Technology and Disability*, 34(2), 83-102. <https://doi.org/10.3233/TAD-210353>
- [22] Davarinia, F., & Maleki, A. (2024). Feature evaluation for myoelectric pattern recognition of multiple nearby reaching targets. *Medical Engineering & Physics*, 130, 104198. <https://doi.org/10.1016/j.medengphy.2024.104198>
- [23] Greene, R. J., Hunt, C., Kumar, S., Betthausen, J., Routkevitch, D., Kaliki, R. R., & Thakor, N. V. (2023). Functionally adaptive myosite selection using high-density sEMG for upper limb myoelectric prostheses. *IEEE Transactions on Biomedical Engineering*, 70(10), 2980-2990. <https://doi.org/10.1109/TBME.2023.3274053>
- [24] Abbaspour, S., Naber, A., Ortiz-Catalan, M., GholamHosseini, H., & Lindén, M. (2021). Real-time and offline evaluation of myoelectric pattern recognition for the decoding of hand movements. *Sensors*, 21(16), 5677. <https://doi.org/10.3390/s21165677>