sEMGest: An Integrated Platform for Hand Gesture Classification from sEMG with Real-Time Digital-Twin Validation

Bruno Malinoski, Gustavo H. dos Santos Schaefer, Carmen Caroline Rasera,
José Jair A. Mendes Junior and Heitor Silvério Lopes
Federal University of Technology of Paraná (UTFPR)
Curitiba, Paraná, Brazil
{malinoski, gschaefer}@alunos.utfpr.edu.br
{rasera, jjjunior, hslopes}@utfpr.edu.br

***Abstract*—Hand gesture recognition from surface electromyography (sEMG) is a cornerstone of intuitive human–machine interfaces and myoelectric prosthesis control. This work presents *sEMGest*, an integrated and reproducible platform that spans custom acquisition hardware, a signal-processing and machine-learning pipeline, and a real-time three-dimensional digital twin of a prosthetic hand used for validation. A four-channel analog front-end based on INA333 instrumentation amplifiers and Sallen–Key active filters feeds an ESP32-S3 microcontroller sampling at 1 kHz with 12-bit resolution; the board was experimentally characterized, showing an approximate 25–540 Hz passband and a 49.5 µV RMS input-referred noise floor. Signals are band-pass filtered (20–500 Hz), segmented into 256 ms windows with 75% overlap, and described by the four time-domain features of the TD4 set. Three classifiers (linear discriminant analysis (LDA), a support vector machine with a radial basis function kernel (SVM-RBF), and random forest (RF)) were compared through grid search and cross-validation, first on the public NinaPro DB5 dataset and then on a proprietary two-participant dataset acquired with the developed hardware. Using repetition-grouped cross-validation, which prevents leakage between overlapping windows of the same repetition, SVM-RBF and RF are statistically tied intra-subject (macro-F1 0.69 on DB5; 0.53–0.60 on the proprietary data), whereas leave-one-subject-out evaluation collapses to 0.29–0.37, evidencing covariate shift and the need for per-user calibration; RF proves the most robust across subjects. The processing latency stays at 2.55 ms (95th percentile), so that the perceived response time is dominated by the 256 ms window-filling delay and remains within the 300 ms myoelectric-control budget. The digital twin renders the recognized gesture in real time, providing a hardware-independent tool for pipeline validation and, prospectively, user training.**

***Index Terms*—surface electromyography, hand gesture recognition, myoelectric prosthesis, digital twin, embedded systems, machine learning, real-time systems.**

I. Introduction

The human hand is among the most complex and versatile anatomical structures of the body, and the loss of an upper limb (by trauma, vascular disease or congenital causes) compromises daily activities and functional autonomy [1]. Myoelectric prostheses use surface electromyography (sEMG) as a control interface: electrodes placed over the skin capture the electrical potentials associated with muscle contraction and drive motorized actuators more directly than purely mechanical systems [5], [6]. Conventional myoelectric control relies mostly on signal amplitude and becomes insufficient when several functions must be controlled reliably; in contrast, pattern-recognition control infers the intended movement from features extracted from the sEMG, widening the command set and improving naturalness [3]. Nonetheless, upper-limb prosthesis abandonment rates between 23% and 45% are still reported, associated with insufficient perceived control and lack of naturalness [2], [4].

Two obstacles are particularly relevant. First, latency: delays above 100–125 ms start to degrade user performance in manipulation tasks and delays above 300 ms make control significantly harder [7], imposing a tight budget on acquisition, processing and classification. Second, robustness: classification performance measured under controlled laboratory conditions is hard to sustain in continuous use because of session variability, muscle fatigue and electrode repositioning [8]. Many systems are also validated only with specialized acquisition equipment and desktop processing, while reducing the number of channels (desirable for wearable devices) remains feasible when the pipeline is properly designed [9]. In parallel, the digital-twin concept, a virtual representation that mirrors a physical system in real time, has been explored in biomedical applications [11], [12] and, for prostheses, can validate control algorithms in a virtual environment before integration with hardware, while extended-reality interfaces support embodiment and training [10], [13].

This work proposes *sEMGest*, an integrated platform that combines sEMG acquisition, machine-learning gesture classification and a real-time digital twin, targeting practical applicability to upper-limb prosthesis control.[1] Its main contributions are:

- an **integrated and reproducible platform** joining a four-channel acquisition front-end (INA333, OPA4277 and ESP32-S3), a Python processing/classification pipeline

[1]Source code (processing pipeline, digital twin and trained models) is available at https://github.com/bmalinoski/sEMGest.

and a three-dimensional real-time visualization environment, in contrast with commercial black-box solutions;
- a **two-stage validation methodology**, first on the reference NinaPro DB5 dataset and then on a proprietary dataset acquired with the developed hardware;
- a **complete experimental characterization** of the hardware (per-channel frequency response, input-referred noise floor and end-to-end latency);
- **empirical evidence of covariate shift** across subjects and sessions, supporting a two-phase architecture with dedicated per-user models; and
- the application of a **digital twin** that renders the classified gesture in real time, acting both as a pipeline-validation instrument and as potential support for user training.

II. Related Work

Classical machine learning remains competitive for sEMG-based gesture recognition. Comparative studies report high intra-subject accuracy for LDA, SVM and RF [14]–[16], and time-domain feature sets are favored for their low computational cost and robustness in real-time pattern recognition [8], [17]. Segmentation parameters (window length, overlap and truncation) strongly influence performance and trade information for latency [18]. Public datasets such as NinaPro support reproducible benchmarking and setup comparison [19], [20], and embedded/real-time solutions with reduced acquisition have been demonstrated on prosthetic hands [21].

On the visualization side, extended-reality and virtual-prosthesis environments have been used mainly for training and embodiment, or, more recently, as myoelectric digital twins for fast and realistic modelling in deep-learning pipelines [10], [11], [13], [22]. In the present platform, the digital twin is coupled end-to-end with a custom low-cost front-end and is used online as a validation instrument for the full acquisition-to-decision chain, in addition to its role as a visualization and training aid.

III. Materials and Methods

A. System Overview

The platform is organized in three subsystems (Fig. 1): (i) a four-channel acquisition front-end with embedded firmware; (ii) a host processing pipeline in Python that performs digital filtering, segmentation, feature extraction and classification; and (iii) a digital twin that renders the classified gesture in real time. Evaluation followed two stages: pre-validation on NinaPro DB5 to establish an algorithmic performance ceiling, and validation on a proprietary dataset acquired with the developed hardware.

B. Acquisition Hardware

Each of the four channels implements a five-stage conditioning chain. An INA333 instrumentation amplifier (CJMCU-333 module, gain $2\times$) provides differential acquisition with high common-mode rejection, followed by a second-order Sallen–Key high-pass filter (20 Hz) and a second-order Sallen–Key low-pass filter (≈ 540 Hz) built with OPA4277 precision op-amps [23]. A non-inverting stage adds a fixed gain of $\approx 500\times$ (total chain gain $\approx 1000\times$), and a unity-gain subtractor centers the signal at 1.5 V to match the ADC input range. The board is powered by a battery through low-dropout regulators, with an on-board mid-supply reference biasing the analog chain. Digitization uses the 12-bit SAR ADC of an ESP32-S3 microcontroller at 1 kHz, obtained by $4\times$ oversampling of a 4 kHz timer. Electrode placement followed SENIAM recommendations [24].

The assembled printed circuit board was experimentally characterized (Fig. 2). The four channels showed a uniform passband of approximately 25–540 Hz, and the input-referred background noise, with inputs short-circuited and the system battery-powered, was 49.5 μV RMS. Considering that forearm sEMG amplitudes typically range from 100 to 500 μV [6], this noise floor is compatible with gesture recognition.

C. Embedded Firmware

The firmware runs on the ESP32-S3 under FreeRTOS. A high-resolution timer dispatches acquisition, and a set of second-order IIR biquads in Direct Form II Transposed applies a 20 Hz high-pass, a 480 Hz low-pass and a 60 Hz notch. Samples are packed into 15-byte binary frames with an XOR checksum and streamed to the host over USB-CDC.

D. Datasets and Acquisition Protocol

Pre-validation used the public NinaPro DB5 dataset (8-channel Myo armband, 200 Hz) [19], [20]. Seven classes were considered: rest and six functional gestures (closed fist, open hand, pinch, pointing, thumbs up and finger abduction).

The proprietary dataset was collected with the developed four-channel hardware from two participants: P1 (right-handed) and P2 (left-handed). To preserve an identical protocol, acquisition was always performed on the right forearm, which is the non-dominant limb for P2. Each gesture followed a cue schedule of 5 s of contraction and 3 s of rest, with 15 repetitions per gesture and participant. Windows were labeled by activity-envelope detection, which locates each contraction directly from the signal amplitude and thereby avoids the misalignment between the manually timed 5/3 s cue and the true contraction onset that a purely schedule-based labeling would introduce.

All experimental procedures involving human participants were conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee of the Federal University of Technology of Paraná (protocol no. 7.218.395). Both participants were members of the research team and provided written informed consent prior to data collection.

E. Signal Processing Pipeline

The host pipeline applies a 20–500 Hz band-pass filter and segments each channel into 256 ms windows with 75% overlap, parameters validated in the literature for real-time applications [18]. Each window is described by the four time-domain features of the TD4 set (MFL, MSR, WAMP and

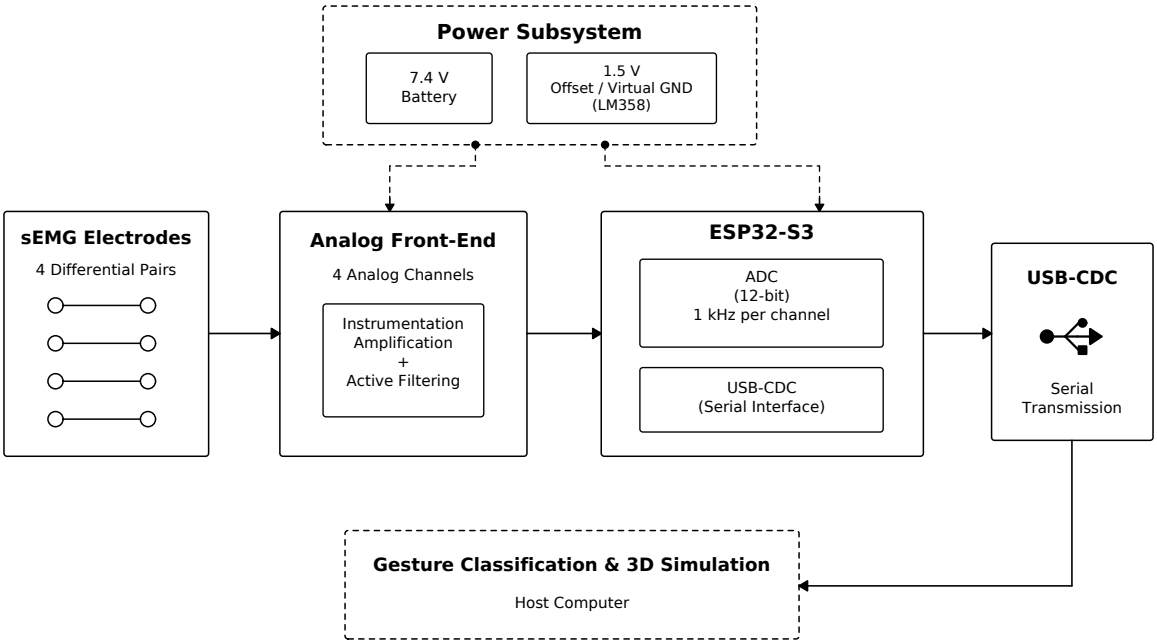
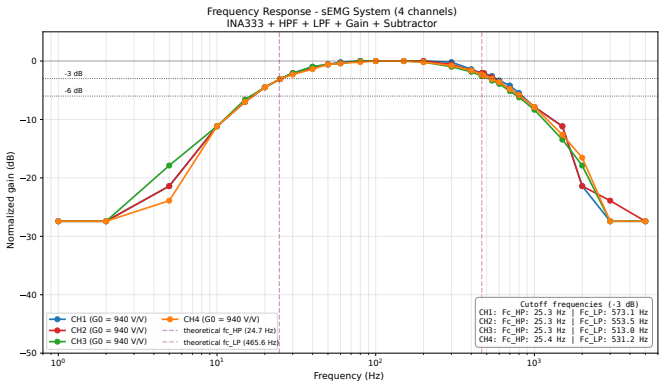
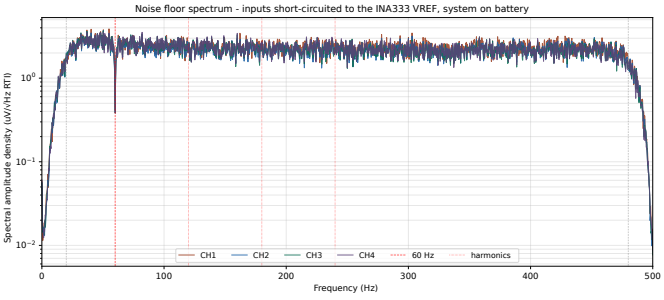


Fig. 1: Overview of the sEMGest platform: four-channel analog acquisition and ESP32-S3 firmware, host processing/classification pipeline, and real-time digital twin of the prosthetic hand.



(a) Magnitude frequency response of the four channels.



(b) Power spectral density of the input-referred background noise.

Fig. 2: Experimental hardware characterization: (a) channel frequency response and (b) input-referred noise floor.

LS) [17], yielding 16 features per window (4 features $\times$ 4 channels). Features are normalized and class imbalance is mitigated by random undersampling of the majority (rest) class to the median of the minority classes.

F. Classifiers and Validation

Three classifiers were implemented with scikit-learn [25] and compared: LDA, SVM-RBF and RF. Hyperparameters were selected by nested grid search, and every cross-validation split, both the outer evaluation loop and the inner grid-search loop, was grouped by *repetition*, so that all windows originating from the same contraction stay in the same fold. This is essential because the 75% window overlap makes neighboring windows of one repetition nearly identical; a random split would place such near-duplicates in both training and test folds and inflate the scores. The intra-subject setting used repetition-grouped cross-validation computed *per subject* (each of the ten NinaPro DB5 subjects individually, and dedicated per-participant models for P1 and P2), reporting the mean and standard deviation across subjects. Cross-subject generalization used full Leave-One-Subject-Out (LOSO) over the ten DB5 subjects, rotating the held-out subject, and the two-direction LOSO between P1 and P2 on the proprietary dataset. For every protocol we report macro-F1, balanced accuracy, and a gesture-only macro-F1 (excluding the rest class), together with the exact per-class window counts and the proportion of rest in the test set.

G. Digital Twin

The digital twin materializes a virtual prosthetic hand that mirrors, in real time, the gesture predicted by the pipeline (Fig. 3). It was implemented in Python with a pygame event loop and legacy OpenGL (PyOpenGL) rendering, in a 1280×720 window split into a 3D viewport in the upper region and a signal panel showing the filtered EMG waveforms and per-channel RMS activation bars in the lower region.

The hand is a serial anthropomorphic kinematic chain of five fingers, with two joints for the thumb (MCP, IP) and three for each remaining finger (MCP, PIP, DIP), dimensioned from the proportions of the open-source OpenBionics model. Each gesture class is encoded as a vector of flexion and abduction joint angles, and phalanx positions are obtained by forward kinematics through the composition of homogeneous 4×4 transforms:

$$T_n = T_{\text{base}} \cdot \prod_{i=1}^{n} R_x(\theta_i)\, T_z(l_i), \tag{1}$$

where θ_i is the flexion angle of joint i, l_i the length of phalanx i, R_x a rotation about the x axis and T_z a translation along z. Bones are rendered as cylinders and joints as spheres, one color per finger. Transitions between successive poses are smoothed by a 200 ms *smoothstep* interpolation,

$$s(t) = t^2\,(3 - 2t), \quad t \in [0, 1], \tag{2}$$

applied to the joint angles to avoid abrupt motion of the 3D model.

The input layer follows a strategy pattern: a common interface abstracts three interchangeable sources feeding the same pipeline and renderer—a *replay* source that streams NinaPro DB5 recordings sample by sample, a *hardware* source fed by the four-channel USB-CDC stream of the ESP32-S3, and a *demo* source for manual pose inspection. A heads-up display overlays the classified gesture, its confidence, the classification latency and the frame rate, and the model can be rotated interactively for inspection. To suppress visual flicker when classifier confidence is low, the pose is updated only when the prediction exceeds a 90% confidence threshold—the previous pose is held otherwise, and the hand reverts to rest after a prolonged low-confidence interval. This acts as a low-cost persistence scheme in place of the heavier sliding-voting post-processing used elsewhere for output smoothing.

IV. Results and Discussion

A. Intra-Subject Classification on NinaPro DB5

Intra-subject performance was evaluated per subject and then averaged across the ten DB5 subjects. Crucially, the cross-validation folds are grouped by *repetition*: because the 256 ms windows overlap by 75%, neighboring windows of the same repetition are nearly identical, and a random split would place such near-duplicates in both the training and the test folds, measuring interpolation within a repetition rather than generalization to unseen repetitions. Grouping every window of a given repetition into the same fold (in both the outer loop and the inner grid-search loop) removes this optimistic bias. Under this protocol, SVM-RBF and RF are statistically tied, with macro-F1 of 0.691 ± 0.050 and 0.692 ± 0.051 respectively, both clearly ahead of LDA (0.656 ± 0.057); balanced accuracy follows the same ordering (Table I). The dispersion across subjects ($\sigma \approx 0.05$) reflects the genuine inter-individual variability of forearm sEMG.

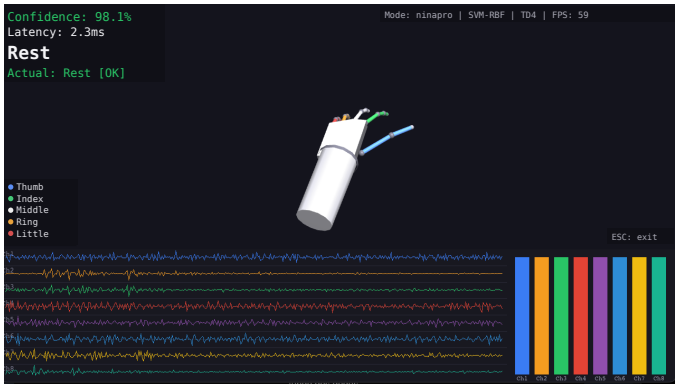


Fig. 3: Digital twin in operation, with the 3D viewport, the EMG signal panel with per-channel activation bars, and the informational HUD (classified gesture, confidence, latency and frame rate).

These figures are markedly lower than those obtained with a subject-aggregated, randomly split validation (macro-F1 up to 0.910), and the gap is precisely the inflation introduced by inter-window leakage between overlapping segments of the same repetition. The repetition-grouped values are therefore the ones that reflect deployment, where the model must classify contractions it has not seen during training. Across subjects, the residual errors concentrate between the lower-amplitude gestures and the rest class.

B. Results on the Proprietary Dataset

On the four-channel proprietary hardware, dedicated per-participant models were trained with the same repetition-grouped cross-validation. The classifier ordering seen on DB5 is preserved (Table I): RF and SVM-RBF are close, with RF marginally ahead (macro-F1 0.601 ± 0.033 vs. 0.590 ± 0.029 for P1; 0.528 ± 0.037 vs. 0.526 ± 0.054 for P2), and both clearly outperform LDA. Absolute values are lower than on DB5, consistent with a four-channel montage and a single acquisition session.

Fig. 4 shows the RF confusion matrix for P1. The dominant error is confusion between the gestures and the rest class in both directions: rest is recovered in only 44% of its windows, and pinch, the weakest gesture, in 51%, with about 31% of its windows assigned to rest. The higher-amplitude, topographically distinct gestures are recognized reliably (thumbs up 83%, finger abduction 78%, pointing 77%). This rest-boundary confusion, rather than confusion between biomechanically similar gesture pairs, is the main error mode on the four-channel montage.

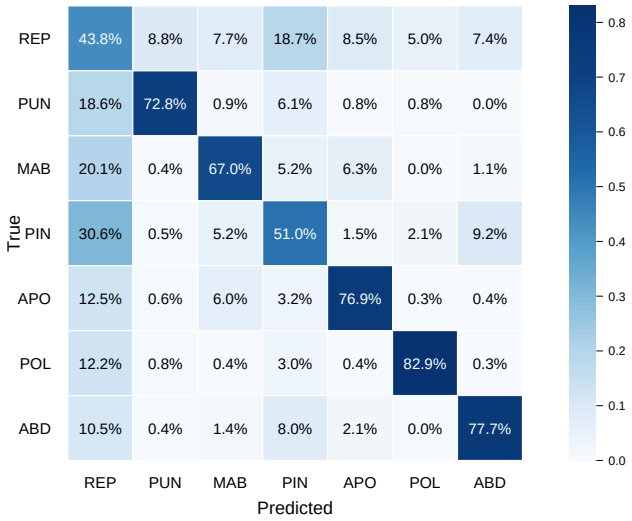


Fig. 4: Confusion matrix (true-class normalized) of the RF classifier for participant P1, from repetition-grouped cross-validation on the proprietary four-channel hardware. Classes: REP (rest), PUN (closed fist), MAB (open hand), PIN (pinch), APO (pointing), POL (thumbs up), ABD (finger abduction).

TABLE I: Repetition-Grouped Cross-Validation Results (Mean $\pm$ Std). Macro-F1 and Bal. Acc. are over all seven classes; F1 (gest.) is the macro-F1 over the six gestures, excluding rest.

Protocol	Clf.	Macro-F1	Bal. Acc.	F1 (gest.)
Intra-subject				
	LDA	0.656 ± 0.057	0.801 ± 0.066	0.611 ± 0.064
DB5	SVM-RBF	0.691 ± 0.050	0.829 ± 0.063	0.649 ± 0.058
	RF	0.692 ± 0.051	0.826 ± 0.061	0.650 ± 0.058
	LDA	0.495 ± 0.033	0.552 ± 0.045	0.510 ± 0.044
P1	SVM-RBF	0.590 ± 0.029	0.643 ± 0.041	0.603 ± 0.036
	RF	0.601 ± 0.033	0.657 ± 0.045	0.613 ± 0.042
	LDA	0.437 ± 0.031	0.492 ± 0.030	0.454 ± 0.032
P2	SVM-RBF	0.526 ± 0.054	0.565 ± 0.051	0.525 ± 0.059
	RF	0.528 ± 0.037	0.579 ± 0.037	0.529 ± 0.042
Inter-subject / inter-participant				
	LDA	0.360 ± 0.085	0.475 ± 0.119	0.275 ± 0.085
DB5 LOSO	SVM-RBF	0.288 ± 0.092	0.316 ± 0.118	0.180 ± 0.103
	RF	0.370 ± 0.106	0.474 ± 0.114	0.279 ± 0.113
	LDA	0.135	0.214	0.106
Own LOSO	SVM-RBF	0.168	0.210	0.132
	RF	0.228	0.278	0.191

The ≈ 6 percentage-point gap between P1 and P2, under identical hardware, pipeline and classifiers, is attributed to acquisition conditions rather than to the methodology. Two factors are consistent with the data: P2 is left-handed but was acquired on the right (non-dominant) forearm to preserve an identical protocol, which reduces the activation uniformity of the target muscles [26]; and a thicker layer of subcutaneous adipose tissue hindered palpation of the muscle bellies, suggesting slightly less precise electrode placement. The small participant count and single session place these figures at the level of a proof of concept rather than a population estimate.

C. Inter-Subject and Inter-Session Generalization

Generalization to unseen subjects was assessed by leave-one-subject-out (LOSO) over all ten DB5 subjects, rotating the held-out subject (Table I). Performance collapses relative to the intra-subject case: the best classifier reaches only 0.370 ± 0.106 macro-F1 (RF), against 0.692 intra-subject. Balanced accuracy tells the same story, and on the gesture-only subset (excluding rest) it falls to 0.18–0.28. Notably, the classifier ranking *inverts* with respect to the intra-subject setting: RF (0.370) and LDA (0.360) now lead, while SVM-RBF drops to 0.288. Raw accuracy is uninformative in this regime: with $\approx 92\%$ of the test windows being rest, SVM-RBF's 0.865 is in fact below the 0.923 of a trivial rest-only predictor, while its balanced accuracy of 0.316 exposes the poor gesture discrimination. The inter-participant LOSO on the proprietary dataset reproduces both the collapse and the inversion (RF 0.228 vs. SVM-RBF 0.168 macro-F1).

Two conclusions follow. First, the collapse is the expected signature of covariate shift (systematic mean and scale changes of the TD4 space across subjects, electrode placement and skin-electrode impedance) which violates the assumption that training and test share the same distribution [4], [8], and it motivates a two-phase architecture with an individualized calibration stage. Second, the flexible RBF boundaries that are competitive within a subject are the least robust across subjects, whereas the simpler decision surfaces of RF and LDA transfer slightly better, an argument for preferring RF as the deployed classifier, since it is both the best (tied) intra-subject and the most robust across subjects.

D. Real-Time Latency

The end-to-end responsiveness of the system is the sum of two contributions of different natures, summarized in Table II: a fixed *windowing* delay imposed by the segmentation scheme, and the *computation* time spent turning a window into a decision. The computation time (feature extraction, normalization and inference) was measured on the proprietary hardware over more than 1,000 consecutive classifications, yielding a mean of 1.79 ms and a 95th percentile of 2.55 ms. This term is negligible next to the windowing delay: because a decision can only be issued once a full 256 ms window is available, the first decision is inherently delayed by 256 ms, and subsequent decisions follow every 64 ms (the 25% step of a 256 ms window with 75% overlap). The serial transfer of each 15-byte frame over the native USB-CDC link adds a sub-millisecond term (one frame per millisecond at 1 kHz) that is likewise negligible. The 200 ms *smoothstep* pose interpolation is a visual animation applied after the decision and does not delay the classification.

The perceived latency is therefore dominated by the 256 ms window-filling delay rather than by computation, and the relevant figure for real-time control is this windowing term plus the few milliseconds of processing, within the 300 ms budget established for myoelectric prostheses by Farrell and Weir [7], although with a smaller margin than the 2.55 ms processing time alone would suggest.

TABLE II: End-to-End Latency Budget of the Real-Time Loop.

Stage	Contribution	Nature
Window filling (first decision)	256 ms	Structural
Inter-decision step (steady state)	64 ms	Structural
USB-CDC frame transfer	< 1 ms (est.)	Estimated
Feature extraction + inference (mean)	1.79 ms	Measured
Feature extraction + inference (p95)	2.55 ms	Measured
Pose interpolation (smoothstep)	200 ms (visual)	Animation

E. Digital-Twin Operation

The digital twin is the component that turns the pipeline into an interactive system, and it played two complementary roles in the evaluation. As a *paired visual-validation instrument*, fed in replay mode with restimulus-labeled DB5 windows from held-out subjects, its online accuracy matched the corresponding offline result within measurement margin, confirming that the streaming implementation faithfully reproduces the algorithmic behavior and ruling out buffer-synchronization or run-time feature-extraction bugs; it thereby complements the confusion matrices with a qualitative, temporal view of the classifier output. As an *online twin of the prosthetic hand*, fed by the four-channel USB-CDC stream, it anticipates the behavior expected once the trained model is embedded in a physical device, with transitions stabilized by the 200 ms smoothstep interpolation and the 90% confidence threshold.

Online operation with the proprietary hardware was evaluated qualitatively: because the target gestures were not logged synchronously, accuracy, macro-F1 and confusion matrices cannot be computed for this phase. Under visual inspection of the rendered pose, predictions oscillated among neighboring gestures (chiefly rest, pointing and open hand) more often than the offline cross-validation suggested, mainly because the exact electrode placement of the training session was not preserved between sessions: the same covariate-shift mechanism, now manifested within the same user across sessions. The confidence threshold and pose smoothing attenuated the visual flicker but did not remove the underlying degradation, reinforcing the need for a brief per-session recalibration. Beyond validation, this hardware-independent embodiment of the recognized gesture also provides a basis for user training and prosthesis-control prototyping before hardware integration, in line with virtual-prosthesis and extended-reality approaches reported in the literature [10], [13], [22].[2]

It should be made explicit what the digital twin mirrors in the present work: it reflects, in real time, the *intended gesture* inferred from the user's muscular activity, mapped onto the kinematic posture of a prosthetic-hand model. It does not yet model the mechanics or dynamics of a physical prosthesis (actuator response, contact forces or joint compliance) which would require a bidirectional coupling to an instrumented device. Establishing that physical, bidirectionally synchronized twin is left as future work; the current system is thus best understood as a real-time virtual hand serving as a validation and visualization instrument, and the user-training role is a prospect its architecture enables rather than a claim evaluated here.

[2]Demonstration videos of the system in real-time operation are available at: https://youtube.com/shorts/pkRm3E10lTw, https://youtube.com/shorts/_bxdrktQDjc and https://youtube.com/shorts/W6CeXGm7YlE.

F. Comparison with the Literature

Table III positions the results against representative sEMG gesture-classification works. Direct comparison requires caution, and not only for the usual differences in channels, dataset, number of gestures and protocol: many reported accuracies rely on randomly split windows, which, as shown above, inflate intra-subject scores when segments overlap. The repetition-grouped 0.69 macro-F1 reported here is therefore conservative by construction, and the more relevant message is the large intra-to-inter-subject gap, whose LOSO collapse reproduces the widely documented behavior of sEMG models without individual calibration [4], [8].

G. Limitations

Results must be read in light of the study's limitations. Pre-validation relied on NinaPro DB5, which differs from the target scenario in sampling rate (200 Hz vs. 1 kHz), number of channels (8 vs. 4), ADC resolution (8 vs. 12 bits) and sensor type; the reported numbers set an algorithmic ceiling, but effective performance still depends on acquisition quality. The proprietary collection comprises only two participants (both members of the research team) in a single session, without contractile-force control, professional anatomical assistance or preservation of electrode placement between sessions. The seven-class set also falls short of the ten or more grasps required by modern multifunction prostheses [19], and online performance was assessed only qualitatively.

H. Practical Implications and Clinical Applicability

Two findings carry the clearest practical weight. First, the intra-subject scores reported under repetition-grouped validation (≈ 0.69 on DB5) are a more realistic estimate of deployed behavior than the higher numbers common in the literature, and they set expectations accordingly: a low-cost four-channel platform is viable as an assistive and training tool, but not yet as a drop-in multifunction controller. Second, and more decisive for clinical use, the inter-subject collapse means a factory-trained, one-size-fits-all model is not usable; any deployment must include a short per-user (and, given the inter-session drift observed with the digital twin, per-session) calibration step, and should favor the more transfer-robust RF over SVM-RBF. For rehabilitation, the immediate value is less as a final prosthesis controller and more as an accessible platform for gesture-based training and biofeedback: the real-time virtual hand lets a user or therapist see the decoded intention without a physical device, and the open, reproducible hardware lowers the barrier for clinical research groups. Reaching routine clinical use would still require larger and more diverse cohorts, amputee participants, a richer grasp set, and the fast on-device calibration outlined above.

TABLE III: Comparison of the Results of this Work with the Literature.

Work	Channels	Gestures	Classifier	Result
Englehart & Hudgins [3]	4	4	LDA	$> 95\%$ acc.
Pizzolato *et al.* [19]	16	52	RF/SVM	70–92% acc. (DB5)
Dellacasa Bellingegni *et al.* [15]	6	7	SVM/LDA	30 amputees
Xue *et al.* [16]	–	10	RF	91.67% acc.
Yang *et al.* [21]	2	7	–	93% (real time)
Mendes Junior *et al.* [18]	8	6	SVM-RBF	$\approx$99% acc.
This work (DB5, intra, per subj.)	8	7	RF/SVM	0.69 macro-F1
This work (own hw., intra)	4	7	RF/SVM	0.53–0.60 macro-F1
This work (DB5, LOSO)	8	7	RF	0.37 macro-F1

V. Conclusion

This work proposed and implemented *sEMGest*, an integrated platform for hand gesture classification from sEMG comprising a four-channel acquisition front-end, a processing/classification pipeline and a real-time three-dimensional digital twin of a prosthetic hand. The hardware was experimentally characterized (25–540 Hz passband, 49.5 μV RMS input-referred noise). Under repetition-grouped cross-validation, RF and SVM-RBF tied intra-subject (macro-F1 ≈ 0.69 on NinaPro DB5, 0.53–0.60 on the proprietary hardware), with a processing latency of 2.55 ms at the 95th percentile and a perceived response time dominated by the 256 ms window-filling delay, within the 300 ms real-time budget. Inter-subject and inter-session evaluations revealed a collapse caused by covariate shift (with RF emerging as the most robust classifier across subjects) justifying a two-phase architecture with dedicated per-user models preceded by a calibration step. As a proof of concept with two participants and a qualitatively evaluated online stage, the results indicate the technical feasibility of the platform, with the digital twin serving as a hardware-independent instrument for pipeline validation and potential user training.

Future work includes enlarging the participant pool, developing a fast per-session calibration procedure and online incremental adaptation to mitigate covariate shift [27], [29], expanding the gesture set, integrating the classifier with a physical robotic hand, adding proportional force control [28], and instrumenting synchronous labeling to quantify online performance.

References

[1] S. L. Carey, D. J. Lura, and M. J. Highsmith, "Differences in myoelectric and body-powered upper-limb prostheses: Systematic literature review," *J. Rehabil. Res. Dev.*, vol. 52, no. 3, pp. 247–262, 2015.

[2] S. Salminger *et al.*, "Current rates of prosthetic usage in upper-limb amputees—have innovations had an impact on device acceptance?," *Disabil. Rehabil.*, vol. 44, no. 14, pp. 3708–3713, 2022.

[3] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, 2003.

[4] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Dev.*, vol. 48, no. 6, pp. 643–659, 2011.

[5] R. Merletti and D. Farina, *Surface Electromyography: Physiology, Engineering, and Applications*. Hoboken, NJ, USA: Wiley, 2016.

[6] C. J. De Luca, "The use of surface electromyography in biomechanics," *J. Appl. Biomech.*, vol. 13, no. 2, pp. 135–163, 1997.

[7] T. R. Farrell and R. F. Weir, "The optimal controller delay for myoelectric prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 111–118, 2007.

[8] A. Phinyomark *et al.*, "EMG feature evaluation for improving myoelectric pattern recognition robustness," *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4832–4840, 2013.

[9] L. P. Murciego *et al.*, "Reducing the number of EMG electrodes during online hand gesture classification with changing wrist positions," *J. NeuroEng. Rehabil.*, vol. 19, no. 1, art. 78, 2022.

[10] K. A. Rodrigues *et al.*, "Embodiment of a virtual prosthesis through training using an EMG-based human-machine interface: Case series," *Front. Hum. Neurosci.*, vol. 16, art. 870103, 2022.

[11] K. Maksymenko *et al.*, "A myoelectric digital twin for fast and realistic modelling in deep learning," *Nat. Commun.*, vol. 14, art. 1600, 2023.

[12] Z. Johnson and M. J. Saikia, "Digital twins for healthcare using wearables," *Bioengineering*, vol. 11, no. 6, art. 606, 2024.

[13] W. Li *et al.*, "Current status and clinical perspectives of extended reality for myoelectric prostheses: review," *Front. Bioeng. Biotechnol.*, vol. 11, art. 1334771, 2024.

[14] D. C. Toledo-Pérez *et al.*, "Support vector machine-based EMG signal classification techniques: A review," *Appl. Sci.*, vol. 9, no. 20, art. 4402, 2019.

[15] A. Dellacasa Bellingegni *et al.*, "NLR, MLP, SVM, and LDA: a comparative analysis on EMG data from people with trans-radial amputation," *J. NeuroEng. Rehabil.*, vol. 14, no. 1, art. 82, 2017.

[16] Y. Xue *et al.*, "sEMG-based human in-hand motion recognition using nonlinear time series analysis and random forest," *IEEE Access*, vol. 7, pp. 176448–176457, 2019.

[17] A. Phinyomark and E. Scheme, "EMG pattern recognition in the era of big data and deep learning," *Big Data Cogn. Comput.*, vol. 2, no. 3, art. 21, 2018.

[18] J. J. A. Mendes Junior *et al.*, "How do sEMG segmentation parameters influence pattern recognition process? An approach based on wearable sEMG sensor," *Biomed. Signal Process. Control*, vol. 81, art. 104546, 2023.

[19] S. Pizzolato *et al.*, "Comparison of six electromyography acquisition setups on hand movement classification tasks," *PLoS ONE*, vol. 12, no. 10, art. e0186132, 2017.

[20] M. Atzori *et al.*, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, art. 140053, 2014.

[21] Z. Yang *et al.*, "Instinctive real-time sEMG-based control of prosthetic hand with reduced data acquisition and embedded deep learning training," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2022, pp. 5666–5672.

[22] A. Boschmann *et al.*, "Immersive augmented reality system for the training of pattern classification control with a myoelectric prosthesis," *J. NeuroEng. Rehabil.*, vol. 18, no. 1, art. 25, 2021.

[23] R. P. Sallen and E. L. Key, "A practical method of designing RC active filters," *IRE Trans. Circuit Theory*, vol. 2, no. 1, pp. 74–85, 1955.

[24] H. J. Hermens *et al.*, "Development of recommendations for SENIAM surface electromyography sensors and sensor placement procedures," *J. Electromyogr. Kinesiol.*, vol. 10, no. 5, pp. 361–374, 2000.

[25] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[26] A. Nynes *et al.*, "Asymmetry in contralateral muscle excitation in proximal vs. distal muscles in upper extremities," *Front. Psychol.*, vol. 16, 2025.

[27] X. Jiang, C. Ma, and K. Nazarpour, "Posture-invariant myoelectric control with self-calibrating random forests," *Front. Neurorobot.*, vol. 18, art. 1462023, 2024.
[28] F. Leone *et al.*, "Hierarchical strategy for sEMG classification of the hand/wrist gestures and forces of transradial amputees," *Front. Neurorobot.*, vol. 17, art. 1092006, 2023.
[29] M. A. Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, 2007.