Online-Causal Forecasting of Physics-Derived Lower-Limb Torque Proxies from Low-Cost Dual-View Motion Capture: A Subject-Specific Feasibility Study

Ange Tong*

National Research Tomsk State University, 36 Lenin Avenue, Tomsk 634050, Russian Federation

* Correspondence: 2024140315@stu.hit.edu.cn

Highlights

What are the main findings?

- At the primary 100-ms horizon, the selected six-channel causal-context RF-TCN achieved strict-test MAE 0.1326 Nm/kg and five-seed MAE 0.1338 ± 0.0011 Nm/kg, indicating low sensitivity to stochastic training within the analyzed recording.
- A controlled five-seed K7-versus-K9 ablation showed only a small mean MAE advantage for the 497-ms receptive field (0.13380 vs 0.13497 Nm/kg); K9 was better in 3/5 seeds and the seed-2026 paired block-bootstrap interval crossed zero.

What are the implications of the main findings?

- The study demonstrates leakage-controlled forecasting of a model-derived dynamic surrogate 100 ms ahead; it does not validate biological joint moments because the torque proxy and gait-context predictors partly share the same marker source.
- In the final stateful RF-TCN benchmark, GPU complete-update p95/p99 latencies were 21.78/25.03 ms with 0/6069 steady-state 50-ms deadline misses; the benchmark does not include camera tracking, actuator dynamics or hard-real-time certification.

Abstract

Low-cost motion capture may support anticipatory biomechanical surrogates, but causal deployability and target validity must be distinguished. We evaluated subject-specific 100-ms forecasting of physics-derived lower-limb torque proxies from dual-view marker video, with four-channel surface electromyography (sEMG) as an optional modality, using one 60-kg adult recording. Approximate 3D trajectories generated proxy targets through an engineering Lagrangian model; no force-plate or validated inverse-dynamics reference was available. Video-sEMG alignment was estimated on training data only and frozen at 559 frames (9.317 s). A six-channel causal gait-context receptive-field-matched temporal convolutional network (RF-TCN) achieved strict-test MAE 0.1326 Nm/kg, RMSE 0.2069, R^2 0.4427, and r 0.6529 over 1564 windows; five-seed MAE was 0.1338 ± 0.0011 Nm/kg. Relative to the earlier causal-context reference, MAE decreased by 2.47%, while other metrics showed no supported improvement. A controlled K7-versus-K9 ablation showed only a small average MAE advantage for K9 and no statistically resolved overall effect. Sparse sEMG added no benefit beyond gait context. In stateful streaming replay, GPU complete-update p95/p99 latency was 21.78/25.03 ms with 0/6069 steady-state 50-ms deadline misses. These results support leakage-controlled, training-stable surrogate forecasting and GPU computational feasibility, not validated biological kinetics or hard-real-time exoskeleton control.

Keywords: marker-based motion capture; lower-limb biomechanics; joint torque proxy; causal forecasting; temporal convolutional network; gait context; sparse surface electromyography; streaming computation; subject-specific modeling

1. Introduction

Lower-limb exoskeletons increasingly rely on individualized control policies rather than fixed assistance profiles. Human-in-the-loop optimization, simulation-trained controllers and biological joint-moment estimation have each shown that assistance quality improves when control is aligned with the mechanical state and needs of the wearer [1–3]. A remaining systems challenge is timing. Camera acquisition, signal conditioning, computation, communication and actuator response introduce delay; consequently, a controller that only reconstructs the present mechanical state may react after a substantial part of the desired assistance window has elapsed. Predicting near-future joint dynamics is therefore attractive for feed-forward assistance, gain scheduling and predictive safety constraints.

Laboratory-grade joint kinetics are commonly obtained from three-dimensional optical motion capture, force plates and inverse dynamics or musculoskeletal modeling. These workflows provide high-quality biomechanical estimates but are expensive, spatially constrained and difficult to deploy outside an instrumented laboratory [4–6]. Low-cost video systems have consequently received substantial attention. OpenCap demonstrated that multi-view smartphone video can recover clinically useful movement dynamics [4], while recent two-camera videogrammetry work showed that inexpensive image-based marker tracking can achieve high kinematic accuracy

when the cameras are explicitly calibrated and validated [5]. These studies motivate accessible visual motion capture, but they also highlight an important distinction: the accuracy of a reconstructed 3D trajectory depends on camera calibration, geometry and validation, and low-cost systems should not be presented as laboratory ground truth without such evidence.

A parallel literature has replaced or complemented classical inverse dynamics with learned mappings from wearable or kinematic measurements to joint moments. Optical kinematics, inertial measurement units (IMUs), surface electromyography (sEMG) and multimodal combinations have all been used to estimate lower-limb kinetics [7-17]. Long short-term memory (LSTM) and other temporal models can predict multiple joint torques from sEMG and joint angles [11-13], IMU-only models can recover useful joint-moment information [9,10], and recent transformer-based work has reported accurate multi-joint angle and torque estimation from richer multi-channel sEMG [15]. Recent studies have directly investigated sEMG-based lower-limb torque calibration and hip-exoskeleton prediction [16,17], while EMG-mediated gait adaptation and adaptive human-exoskeleton collaboration demonstrate the relevance of biological signals to assistive control [23,24]. Collectively, the literature confirms that temporal learning is suitable for biomechanical estimation, but reported performance depends strongly on sensing richness, target construction, subject generalization and experimental protocol.

For deployment, three distinctions are especially important. First, a model can be architecturally causal while its inputs are not: centered filters, future-sample interpolation, completed future gait cycles or retrospectively repaired marker trajectories can silently introduce information unavailable at run time. Second, a model-derived target that shares upstream measurements with its predictors should not be confused with an independently measured biomechanical ground truth. Third, low numerical inference time alone does not establish hard-real-time suitability if preprocessing, deadline misses and worst-case timing are omitted. These distinctions are particularly important in anticipatory prediction, where future leakage can artificially simplify the task and source-coupled labels can be over-interpreted as biological validation.

This study evaluates a complete sensing-to-surrogate-forecasting pipeline rather than treating model architecture in isolation. A dual-view marker-based video workflow reconstructs approximate 3D lower-limb trajectories; an engineering Lagrangian model generates physics-derived torque proxies offline; and a separate stateful Online-Causal pathway constructs predictor inputs available only at or before the current time. The scientific objective is not to validate biological joint moments. Instead, it is to characterize whether a low-cost causal state representation can anticipate the future output of a fixed physics-based surrogate under a leakage-safe protocol. The primary endpoint is 100-ms forecasting with a context-only receptive-field-matched temporal convolutional network (RF-TCN) selected on validation data. Receptive-field matching is used here as a controlled temporal-design refinement, not as a claim of a new convolutional mechanism. Supporting analyses characterize 0-200 ms horizons, preprocessing causality, the available sparse sEMG montage, broad model classes and chronological computational latency.

The study addresses five research questions: (RQ1) does strict input causality materially change proxy-forecasting performance relative to retrospective preprocessing; (RQ2) how does predictive fit change across 0-200 ms anticipation horizons; (RQ3) under the available four-channel distal-muscle sEMG montage, does sEMG add information beyond marker-derived gait context; (RQ4) what error change, training-seed variability and controlled receptive-field effect are observed at the validation-selected 100-ms RF-TCN endpoint; and (RQ5) what latency distribution and deadline-miss rate are observed for the final stateful RF-TCN software path? The scope is explicitly subject-specific and within-session. All outputs are physics-derived torque proxies, and the primary claims concern causal surrogate forecasting and computational characterization rather than biological joint-moment validity or closed-loop control efficacy.

2. Related Work

2.1. Accessible motion capture and biomechanics

Marker-based optical motion capture remains a reference method for quantitative human movement analysis, but the cost and operational complexity of multi-camera laboratories have motivated lower-cost alternatives. OpenCap combines multi-view video with musculoskeletal modeling to estimate movement dynamics from smartphone recordings [4]. More recently, Peña-Trabalon et al. developed a calibrated two-camera videogrammetry system and validated marker-derived 3D coordinates against Vicon, illustrating that low-cost camera systems can support quantitative kinematic analysis when acquisition geometry is explicitly controlled [5]. Other validation studies of low-cost video systems similarly emphasize that camera placement and calibration substantially affect 3D accuracy. The present work follows the accessible-motion-analysis direction but uses a simpler orthogonal-view fusion procedure rather than calibrated stereo triangulation; accordingly, it treats the reconstructed coordinates as approximate 3D motion-capture trajectories and confines the main claim to system feasibility.

Classical dynamics software such as OpenSim formalizes the relationship between kinematics, segment properties and joint-level dynamics [6]. Several data-driven studies use laboratory inverse-dynamics outputs as supervised

targets for models operating on reduced sensor sets [7–10]. Our work adopts the same broad separation between offline biomechanical target generation and deployable predictor inputs, but the target is an engineering torque proxy because no force-plate ground reaction force was available. This distinction is maintained throughout the evaluation.

2.2. Data-driven lower-limb torque estimation

Machine-learning approaches to lower-limb kinetics span optical motion features, wearable inertial sensors and muscle activity. Mundt et al. showed that neural networks can infer gait joint angles and moments from motion-capture-derived variables [7,8]. Hossain et al. estimated lower-extremity joint moments and three-dimensional ground reaction forces from IMUs across multiple walking conditions [10], while Altai et al. compared several time-series neural architectures for IMU-based lower-limb moment prediction [9]. Multi-source signal fusion has also been used to predict limb joint angles for exoskeleton wearers [20]. These studies illustrate the practical value of replacing laboratory inputs with reduced sensing once a supervised kinetic target is available.

sEMG is attractive because muscle electrical activation can precede observable motion. Zhang et al. predicted four lower-limb joint torques across daily activities using LSTM networks and transfer learning [12], and later compared EMG-driven neuromusculoskeletal models with LSTM estimators [13]. Wang et al. combined sEMG characteristics and joint angles with LSTM/Gaussian-process regression [11]. Kizyte et al. demonstrated that EMG representation and input selection influence ankle-torque prediction [14]. Recent work has extended this direction to multi-channel transformer models [15] and domain-adaptive calibration [16]. These results establish the potential of rich myoelectric sensing, but they do not imply that a small distal-muscle subset will necessarily add information beyond gait kinematics. Our modality ablation directly tests that question under the available four-channel configuration.

Foroutannia et al. used EMG and additional sensor signals to predict future hip-joint position for exoskeleton control [17], and earlier exoskeleton studies linked gait-phase estimation with assistive-torque prediction [18,19]. Data-driven hip-joint power prediction has also been used to optimize variable-stiffness exoskeleton assistance [22]. Chen et al. demonstrated direct EMG-mediated gait adaptation on a lower-extremity exoskeleton with healthy participants [23], whereas He et al. combined sEMG-based motion estimation with adaptive exoskeleton control across variable walking environments [24]. These works motivate anticipatory control but differ in target definition, sensing and causal validation. The present study focuses specifically on whether a full preprocessing-and-inference chain remains causal and computationally feasible when forecasting physics-derived multi-joint torque proxies.

2.3. Temporal models and causal deployment

Recurrent networks, transformers and temporal convolutions are all plausible sequence models for biomechanical signals. Temporal convolutional networks (TCNs) use causal dilated convolutions and residual connections to obtain a finite temporal receptive field without recurrent state [21]. For a fixed input history, however, the nominal window length and the network receptive field are not equivalent: if the receptive field is shorter than the supplied window, the oldest samples cannot influence the endpoint representation. Receptive-field matching is an established sequence-model design principle and is not presented here as architectural novelty. In this study it serves a narrower engineering question: whether the fixed 500-ms streaming history is more consistently used when the causal TCN receptive field is made approximately equal to that buffer. The broader contribution is the controlled, leakage-safe system evaluation rather than a new TCN operator.

3. Materials and Methods

3.1. Study design and problem formulation

The analyzed dataset comprises one recording from an adult volunteer participant with a body mass of 60 kg and is treated as a subject-specific, within-session feasibility study. The archived repository does not retain age, sex or height metadata. The author operated the acquisition computer and supervised the recording, while the participant completed the locomotion protocol. The recording contains standing, progressive treadmill walking from 1.25 to 4.25 km/h, faster locomotion at 4.75-5.25 km/h, deceleration, cross-step trials at 2.50-3.00 km/h, small jumps, squat-to-stand and a short backward-walking segment. After video-sEMG alignment, the effective overlap used by the analysis workflow was 306.45 s. The backward segment was too short to remain evaluable after the strict temporal split and purge procedure. The 100-ms horizon was designated as the primary model-development endpoint; the 0-200 ms horizon sweep, preprocessing comparison and runtime replay are retained as supporting system-level analyses.

Let x(t) denote the predictor feature vector available at sample t and let τ(t) be the four-dimensional physics-derived torque-proxy vector. With a causal history of W=500 samples and forecast horizon h, the learning problem is

$$X(t)=[x(t-W+1),\ldots,x(t)],\hat{\tau}(t+h)=f_\theta[X(t)],h\in\{0,50,100,150,200\}ms(1)$$

The predictor never receives the future torque label or future-derived gait state. Torque proxies are computed offline and serve only as supervised targets.

3.2. Video motion capture and sEMG acquisition

Two camera views were recorded: a sagittal/side view and a frontal view. The original videos were 3840×2160 pixels. The front recording had a nominal frame rate of 60000/1001 Hz (≈59.94 Hz), the side recording approximately 59.59 Hz, and kinematic processing used 60 Hz. The archived records do not preserve camera brand/model, camera-to-subject distance or camera height, so these values are not inferred. Physical markers were tracked at four right-lower-limb locations: A, hip; B, knee; C, hindfoot/lace region; and D, toe/forefoot. Marker material and diameter were not preserved in the archived metadata. Representative synchronized raw frames from the two views are shown in Fig. 1(a), and the tracked marker configuration is illustrated in Fig. 1(b). The manuscript figure does not display the participant's face.

sEMG was stored as four pulse-code modulation (PCM) channels and processed at 1000 Hz. The recorded channels were right gastrocnemius, right peroneus longus, left gastrocnemius and left peroneus longus. Acquisition-device model, electrode model, amplifier gain and hardware filter settings were not recoverable from the archived records. Consequently, claims are restricted to the recorded signal channels and software processing that can be verified from the repository.

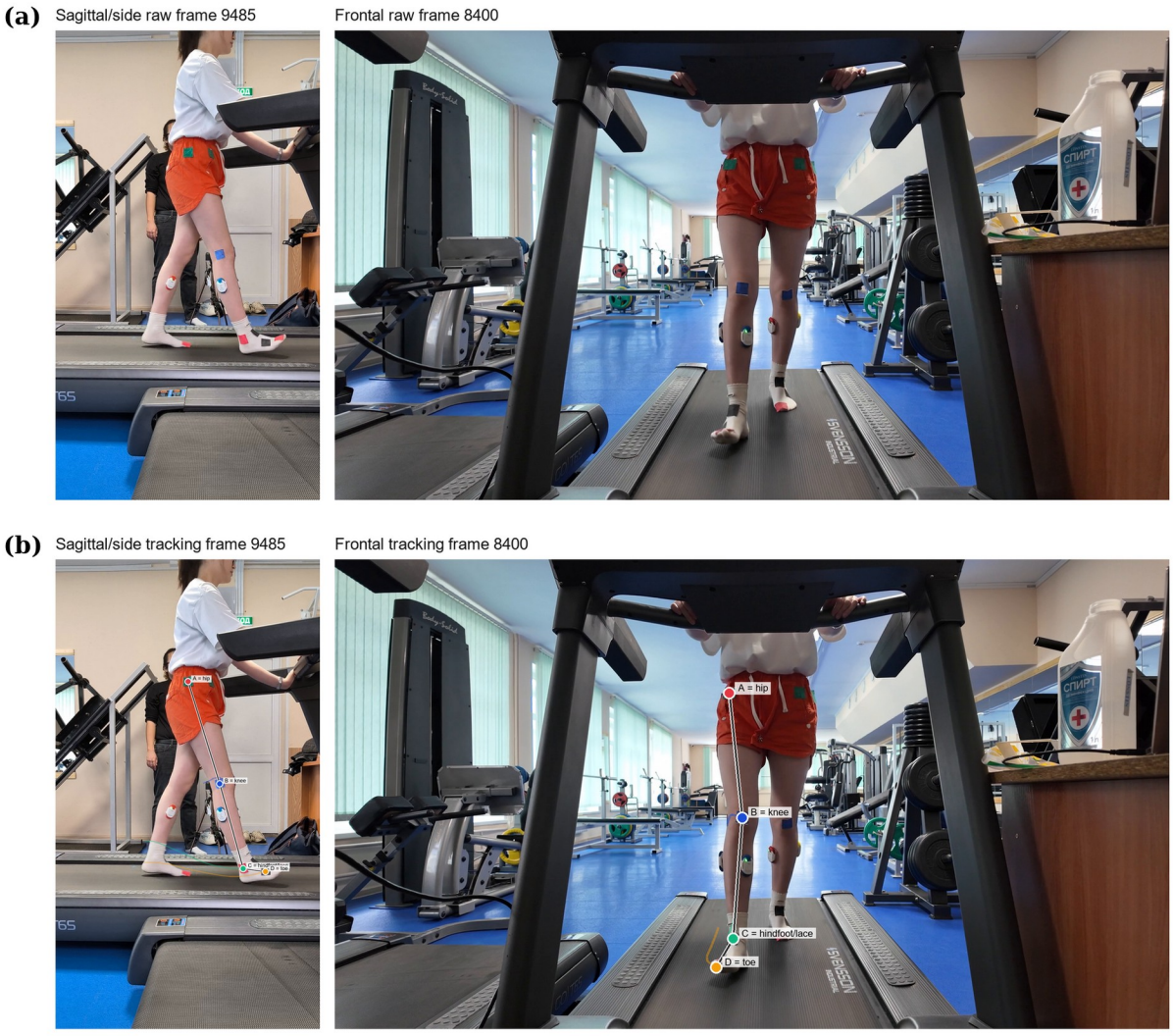


Fig. 1. Dual-view marker-based motion capture during treadmill walking. (a) Temporally aligned sagittal/side and frontal raw video frames. (b) Corresponding marker-tracking result showing the four right-leg landmarks used in this study: A, hip; B, knee; C, hindfoot/lace; and D, toe.

Item	Verified specification
Camera views	2: sagittal/side and frontal
Original resolution	3840×2160 pixels for both views
Video rate	front ≈59.94 Hz; side ≈59.59 Hz; processed at 60 Hz
Markers	A hip; B knee; C hindfoot/lace; D toe/forefoot

Item	Verified specification
Surface electromyography (sEMG)	4 pulse-code modulation (PCM) channels at 1000 Hz
Muscles	bilateral gastrocnemius and peroneus longus
Raw duration	front 349.13 s; side 372.92 s; PCM ≥351117 samples/channel
Aligned analysis duration	306.45 s; 306450 EMG samples; 18881 kinematic frames
Hardware synchronization	No hardware trigger; train-only offset 559 frames (9.317 s), frozen before val/test

Table 1. Verified acquisition parameters recoverable from the archived experiment and code.

3.3. Marker tracking, scaling and approximate 3D fusion

Frame-wise marker centers from the two views were paired after temporal alignment. The side view supplied the longitudinal and vertical components, while the frontal view supplied the mediolateral component. View-specific manual origins translated the trajectories into a common physical frame. Scaling used recorded geometric references, including a 100-mm C–D reference and 230-mm shoe length, together with marker-height calibration values retained in the reconstruction workflow. Because this procedure does not estimate full camera intrinsics/extrinsics or perform stereo triangulation, the term approximate 3D is used deliberately. An example of the reconstructed right-leg trajectory cloud and the instantaneous A–B–C–D link geometry is shown in Fig. 2.

$$r_i(t)=[X_i, side(t), Y_i, front(t), Z_i, side(t)]^T, i \in \{A, B, C, D\} \quad (2)$$

The Offline-Reference pathway used the existing gap-filled marker trajectory produced by retrospective linear interpolation with boundary filling. The Online-Causal pathway used the unfilled marker stream. In streaming mode, a missing marker could be held from its last valid observation for at most 30 video frames. If a contact-related marker remained invalid beyond the allowed hold, the corresponding confirmed contact state was retained rather than forcing a false no-contact transition.

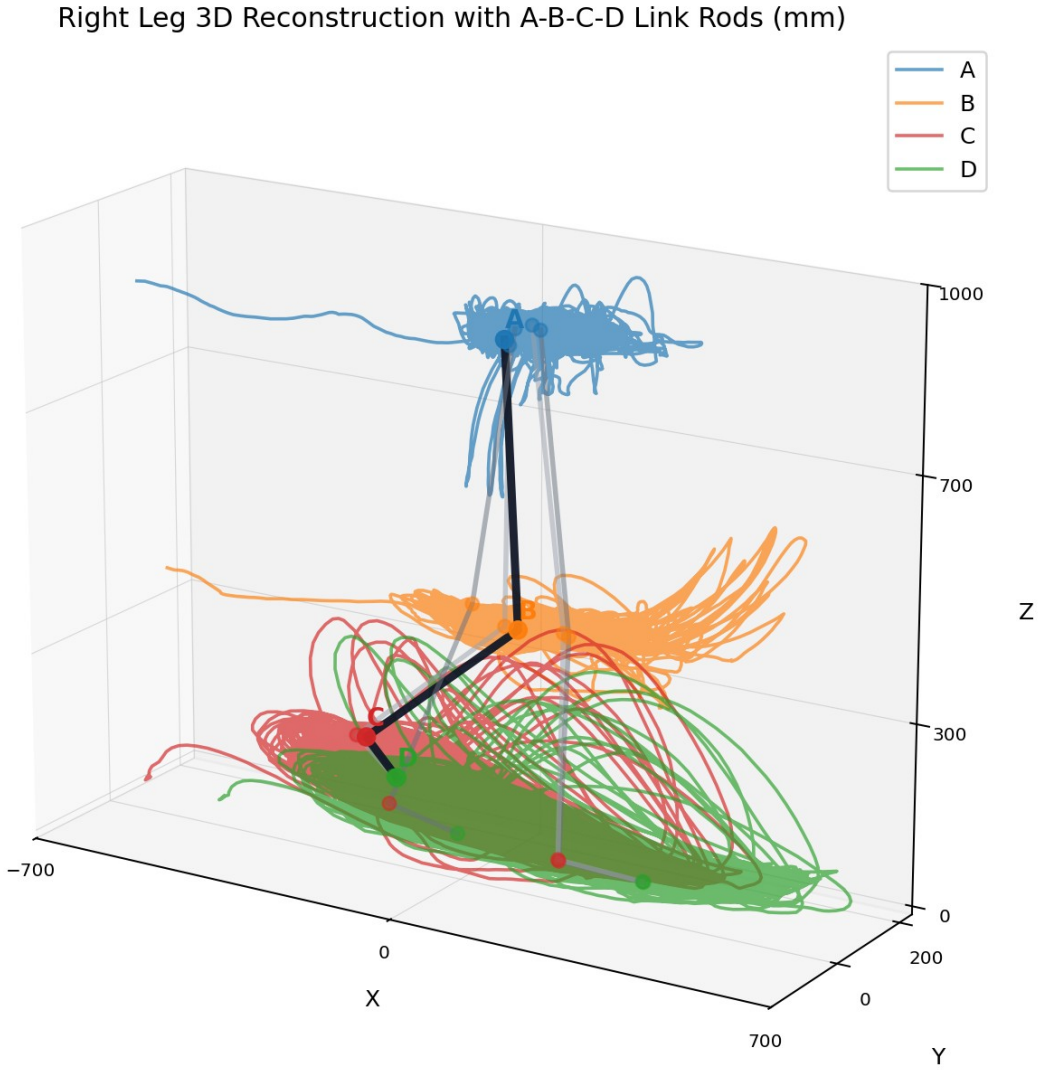


Fig. 2. Approximate three-dimensional right-leg reconstruction obtained from the fused dual-view marker trajectories. Colored traces denote the four tracked landmarks (A-D); the dark A-B-C-D links illustrate the instantaneous segment geometry used in the kinematic and dynamics pipeline. Coordinates are shown in millimetres.

3.4. Physics-derived torque-proxy model

The reconstructed right-leg marker trajectories defined segment geometry and joint kinematics. Segment lengths were obtained from the A-B, B-C and C-D geometry; the toe-tip geometry additionally used the 230-mm shoe-length reference. The participant body mass was 60.0 kg and was used as the body-mass parameter in the primary dynamics analysis. Segment mass fractions were 0.1000 (thigh), 0.0465 (shank), 0.0145 (foot) and 0.0030 (toe). Segment center-of-mass locations were placed at 0.433 of thigh and shank length and 0.50 of foot and toe length. An unmeasured-body proxy center of mass was placed 0.225 m above the hip reference. The segment fractions, center-of-mass locations and unmeasured-body proxy are modeling assumptions rather than participant-specific anthropometric measurements.

The internal rigid-body contribution follows a Lagrangian formulation. With generalized coordinates q, kinetic energy T and potential energy V,

$$L(q,\dot{q}) = T(q,\dot{q}) - V(q), d/dt(\partial L/\partial \dot{q}) - \partial L/\partial q = Q \quad (3)$$

External loading was represented by a stance-gated residual-force approximation. The generalized contribution of an external force proxy Fext acting through Jacobian J(q) is

$$Q_{ext} = J(q)^T F_{ext} \quad (4)$$

The implementation combines internal Lagrangian terms and the J^TF contribution under its joint-sign convention, and then converts absolute joint torques to the relative-joint representation used by the prediction task. The final four targets are hip yaw, hip pitch, knee pitch and ankle pitch. They are normalized by the participant body mass Mbody = 60 kg:

$$\bar{\tau}(t) = \tau proxy(t)/Mbody, Mbody = 60\,kg \quad (5)$$

Because force-plate ground reaction forces (GRFs) and a validated laboratory inverse-dynamics reference were not available, these quantities are consistently termed physics-derived joint torque proxies. The model is used to create a repeatable engineering supervision signal, not to claim ground-truth biological joint moments.

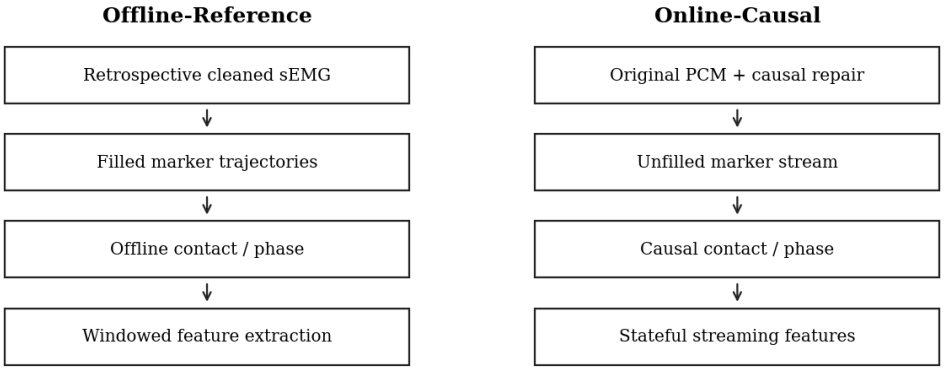
The supervision target is therefore source-coupled to the visual measurements: the torque proxy is generated from reconstructed marker kinematics and modeled contact/loading, while several predictor variables (contact and phase) are also derived from the marker stream. The primary 100-ms task remains temporally predictive rather than a same-sample algebraic remapping because the predictor at time t uses only information available at or before t, whereas the label is the offline proxy at t + 100 ms and depends on future reconstructed motion. This temporal separation does not create an independent biomechanical reference. Accordingly, prediction metrics quantify how well the causal model anticipates the future output of the fixed surrogate model; they do not quantify biological joint-moment accuracy.

Model quantity	Value / assumption
Participant body mass	60.0 kg; used for dynamics parameterization and normalization
Segment mass fractions	thigh 0.1000; shank 0.0465; foot 0.0145; toe 0.0030
Segment center-of-mass (COM) fractions	thigh 0.433L; shank 0.433L; foot 0.50L; toe 0.50L
Unmeasured-body COM proxy	hip + [0, 0, 0.225] m
External load	stance-gated residual-force proxy for ground reaction force (GRF); no force plate
Prediction targets	hip yaw, hip pitch, knee pitch, ankle pitch
Normalization	Nm/kg using participant body mass (60 kg)

Table 2. Biomechanical assumptions used to construct the physics-derived supervision targets.

3.5. Offline-Reference and Online-Causal preprocessing

The core experimental control is shown in Fig. 3. The Offline-Reference uses the pre-existing retrospectively cleaned sEMG export and filled marker trajectories. The Online-Causal route starts from original PCM samples and the unfilled marker stream, and every state update uses only information available at or before the current time. To prevent synchronization from becoming an unrecognized leakage path, the video-sEMG offset was estimated exclusively on the training partition. For each candidate frame offset k0, training-split EMG-time frames were paired with kinematic frames k0+t, and a composite alignment score combined signal-level correlation, first-difference correlation and rolling-energy correlation. The training-only optimum was 559 video frames (9.317 s; composite score 0.359). This value was frozen before validation and test processing and used for all final results; validation and test samples did not enter the alignment search. Because no hardware trigger was available and locomotor signals are periodic, the selected offset is treated as a training-derived registration parameter rather than an independently verified physical time-zero measurement, and residual synchronization error remains possible.



Matched split, model configuration and 100-ms prediction target

After the offset had been selected and frozen, a training-only local sensitivity check evaluated the same composite score at 529, 544, 559, 574 and 589 frames (±30 frames around the selected value). Scores were 0.3327, 0.3473, 0.3594, 0.3444 and 0.3336, respectively. Thus, 559 frames was the highest-scoring candidate within the examined local neighborhood. This analysis was descriptive only: it did not reselect the offset, did not use validation or test data, and does not establish a hardware-verified physical synchronization origin.

Fig. 3. Controlled comparison between retrospective Offline-Reference preprocessing and the strictly Online-Causal predictor-input pipeline.

3.6. Causal sEMG processing and feature construction

The online sEMG cleaner updates each channel sample-by-sample. A trailing-history median and median absolute deviation (MAD) are used to identify invalid samples, including non-finite values, near-zero dropouts and large robust outliers. Invalid samples are replaced by the last valid observation; before any valid sample is available, the current trailing median is used. No centered filter, future neighbor or backward fill is permitted in the online route.

For a generic cleaned sEMG channel e(t), causal 50-ms features are computed over N=50 samples:

$$r(t)=|[e(t)]|, env(t)=(1/N)\Sigma(k=0\ldots N-1)|[e(t-k)]|, rms(t)=\sqrt{[(1/N)\Sigma(k=0\ldots N-1)e(t-k)^2]} \quad (6)$$

The engineered set further includes backward slopes and bilateral difference/ratio features. Across four channels this yields 32 EMG-derived variables. The full multimodal input adds six causal gait-context variables for a total of 38 channels. Feature normalization is fitted only on the training partition.

3.7. Causal contact and gait-phase estimation

Rear-foot and forefoot contact are inferred from the C-marker and D-marker heights using thresholds of 0.08 m and 0.04 m, respectively. A two-frame on/off debounce is applied at the 60-Hz video rate, implying approximately 33.3 ms of contact-decision delay. Contact states are propagated to the 1000-Hz predictor timeline by causal zero-order hold.

A gait cycle begins at detected stance onset. Let t_s be the most recent stance-onset time and T_med the median duration of the five most recently completed stance cycles; before sufficient history is available, T_med=0.700 s. The causal phase is

$$\varphi(t)=clip[(t-t_s)/T_{med},0,1], p_{\sin}=\sin(2\pi\varphi), p_{\cos}=\cos(2\pi\varphi) \quad (7)$$

The six gait-context inputs are stance state, rear contact, fore contact, phase, sin(2πφ) and cos(2πφ). Future heel strikes or future completed cycles are never used to revise the current phase.

3.8. Temporal models and receptive-field-consistent RF-TCN selection

Two model roles are distinguished so that the selected 100-ms predictor is not conflated with supporting system analyses. The primary 100-ms model is the context-only RF-TCN described below. The horizon sweep, modality ablation and Offline-Reference versus Online-Causal comparison retain the earlier activity-conditioned causal residual TCN as a supporting system-reference model. That reference model has hidden width 128, kernel size 7, dropout 0.15 and five residual blocks with dilations 1, 2, 4, 8 and 16; each block contains two causal Conv1d layers. An auxiliary eight-class activity/locomotion-state head predicts stand, speed-up walking, fast jog, slow-down walking, cross-step, small jump, squat or backward walking, and its 16-dimensional soft embedding conditions the four-output torque head. This auxiliary target is distinct from the continuous causal stance-phase variables supplied as predictors.

For a causal one-dimensional convolution with dilation d and kernel coefficients w_k,

$$y(t)=\Sigma(k=0\ldots K-1)w_k x(t-dk) \quad (8)$$

which guarantees that y(t) depends only on current or earlier samples. For the five-block, two-convolution-per-block stack, the nominal endpoint receptive field is $R = 1 + 2(k - 1)\Sigma d_i$ samples. With k = 7 and dilations {1, 2, 4, 8, 16}, R = 373 samples; with k = 9, R = 497 samples. The 500-sample (500-ms) history was a fixed bounded streaming buffer used before RF-TCN candidate screening; at the 50-ms update stride it spans ten update intervals. It was not selected because it equals a universal gait-cycle duration, and no physiological optimality of 500 ms is claimed. Under the kernel-7 reference, approximately the oldest 127 ms of the supplied buffer cannot influence the endpoint representation, whereas kernel 9 makes almost the entire fixed buffer accessible. This receptive-field consistency, rather than a claim of a new TCN mechanism, motivated the RF-TCN refinement.

The final RF-TCN uses only the six causal gait-context channels and removes the auxiliary activity-conditioning branch. Its encoder retains five residual blocks and hidden width 128 but uses kernel size 9; the final causal state is passed through Linear(128,128), Gaussian error linear unit (GELU), dropout 0.15 and Linear(128,4), giving 1,353,220 trainable parameters. Five kernel-9 context-only candidates were screened using validation MAE only: hidden widths 128, 96 and 64 at dropout 0.15, plus widths 96 and 64 at dropout 0.10. The width-128/dropout-0.15 configuration (C1) was selected with validation mean MAE 0.11017 Nm/kg. No strict-test metric was used to choose among these five candidates.

Training settings were shared wherever applicable: AdamW, learning rate 2×10^{-4}, weight decay 1×10^{-4}, batch size 128, maximum 80 epochs, validation patience 12 and gradient clipping at 1.0. The supporting system-reference model uses SmoothL1 torque loss plus 0.25-weighted activity-state cross-entropy, whereas RF-TCN uses SmoothL1 torque loss only. After the RF-TCN architecture and 100-ms endpoint were frozen, training stability was evaluated across seeds 2024-2028 with identical data, hyperparameters and strict-test anchors. The existing seed-2026 K9 checkpoint was reused and only the four missing K9 seeds were trained; no seed result was used to modify the architecture.

$$\mathscr{L} = \mathscr{L}_{tau} + 0.25\,\mathscr{L}_{a}(9)$$

where $\mathscr{L}\tau$ is SmoothL1 regression loss over the four normalized torque proxies and $\mathscr{L}a$ is the auxiliary eight-class cross-entropy activity-state loss. RF-TCN minimizes $\mathscr{L}\tau$ alone. Architectural baselines (Ridge, multilayer perceptron (MLP), LSTM, gated recurrent unit (GRU), standard TCN and a causal transformer) are retained as matched-input sensitivity analyses rather than as candidates selected against the final RF-TCN.

To isolate the kernel/receptive-field factor from the simultaneous removal of the auxiliary activity-conditioning head, a separate controlled ablation compared two otherwise identical context-only simple-head TCNs. K7 used kernel size 7 (nominal receptive field 373 ms) and K9 used kernel size 9 (497 ms); both used hidden width 128, dropout 0.15, dilations {1, 2, 4, 8, 16}, the same six causal-phase inputs, the same SmoothL1 objective, optimizer settings, chronological split, training-only alignment and 1564 strict-test anchors. The comparison was repeated for seeds 2024-2028. Existing K9 results were reused and only the missing K7 runs were trained. This post-selection ablation was used to characterize the temporal-design factor, not to reopen architecture selection.

Parameter	**Supporting system-reference**	**Selected RF-TCN (100 ms)**
Input sampling	1000 Hz	1000 Hz
History window	500 ms (500 samples)	500 ms (500 samples)
Update stride	50 ms (20 Hz)	50 ms (20 Hz)
Forecast horizons	0, 50, 100, 150, 200 ms	100 ms primary endpoint
Predictor inputs	sEMG + causal phase (38 channels)	causal phase only (6 channels)
Hidden width	128	128
Kernel / dilations	7 / 1, 2, 4, 8, 16	9 / 1, 2, 4, 8, 16
Nominal receptive field	373 samples (≈373 ms)	497 samples (≈497 ms)
Residual blocks	5; two causal one-dimensional convolution layers/block	5; two causal one-dimensional convolution layers/block
Conditioning / head	8-class activity head; 16-d embedding	Linear 128 → 128 → 4; no auxiliary head
Loss	SmoothL1 + 0.25 cross-entropy	SmoothL1
Optimizer	AdamW; learning rate 2×10^{-4}; weight decay 1×10^{-4}	same
Training	batch 128; max 80; patience 12; clip 1.0	batch 128; max 80; patience 12; clip 1.0; final stability seeds 2024-2028
Trainable parameters	1,092,748 (full 38-channel model)	1,353,220

Table 3. Model roles and verified temporal configuration. The selected receptive-field-matched temporal convolutional network (RF-TCN) has a 497-ms nominal receptive field, closely matching the fixed 500-ms causal history at the 1000-Hz predictor timeline; the supporting system-reference model is retained for horizon, modality and preprocessing analyses.

3.9. Leakage-safe split, statistics and streaming benchmark

Windows are divided chronologically rather than randomly. The final split contains 158.937 s training, 44.720 s validation and 84.386 s testing, with 18.407 s removed as purge intervals to prevent overlap leakage across split boundaries. At 100 ms, the strict test set contains 1564 prediction windows. Input and target scalers are fitted on training data only, early stopping and RF-TCN candidate selection use validation data, and the strict test set is not used to choose among C1-C5 RF-TCN candidates.

Performance is reported as mean absolute error (MAE), root mean square error (RMSE), range-normalized RMSE (NRMSE), coefficient of determination R^2 and Pearson correlation r. For joint j with n samples, $MAE = n^{-1}\Sigma_i|\tau_i - \hat{\tau}_i|$ and $RMSE = \sqrt{[n^{-1}\Sigma_i(\tau_i - \hat{\tau}_i)^2]}$; NRMSE divides RMSE by the observed target range within the evaluated set. Because overlapping windows are temporally dependent, uncertainty is estimated with temporal block resampling rather than independent-window resampling. Joint-specific uncertainty for the original 100-ms system-reference condition uses a 5-s block bootstrap with 1000 repetitions. Paired 5-s moving-block bootstrap comparisons use 5000 repetitions on identical timestamps for the principal modality contrast, Offline-Reference versus Online-Causal preprocessing, RF-TCN versus the original causal-context reference, and the controlled seed-2026 K9-versus-K7 contrast. RF-TCN stochastic training stability and the controlled K7/K9 comparison are summarized across five seeds by mean, standard deviation, range and paired per-seed MAE differences on the same strict-test anchors. Bootstrap intervals quantify within-recording temporal uncertainty only; seed variation quantifies sensitivity to stochastic training only. Neither is a population confidence interval across participants.

The 50-ms update period is used as a scheduling target for computational profiling, not as a hard-real-time certification criterion. The final RF-TCN benchmark reports cold-start latency separately from steady-state timing. After 50 warm-up updates, 6069 chronological updates are timed on the central processing unit (CPU) and an NVIDIA CUDA-enabled graphics processing unit (GPU), and median, p95, p99, maximum complete-update latency and the fraction of missed 50-ms deadlines are reported. The timed path starts from original raw PCM samples and the unfilled marker-coordinate trajectory, performs stateful causal context/contact/phase updates, assembles the 500-ms six-channel window, applies training-fitted normalization, transfers tensors and executes RF-TCN inference. Precomputed cached feature arrays are not used as the timed feature source.

The stateful streaming implementation reproduces causal features without precomputing full-recording feature, contact or phase arrays. Prefix-invariance tests verify that outputs up to time t are unchanged if the recording is truncated at t, and stateful predictions are compared with batch-causal execution at strict-test anchors. The final timing benchmark should be interpreted as marker-coordinate/PCM-to-prediction software profiling. It does not include camera image acquisition, marker detection/tracking, dual-view reconstruction, operating-system scheduling guarantees, controller execution, communication or actuator dynamics; therefore it is not a sensor-to-actuator latency measurement.

4. Results

4.1. Causality and streaming equivalence

All prefix-invariance checks passed for cleaned sEMG, marker/contact state, gait phase and the complete online feature stream. Batch-causal and stateful execution agreed to numerical precision: maximum absolute differences were 0 for cleaned sEMG, 1.60×10^{-7} for engineered sEMG, 0 for contact, 5.66×10^{-7} for phase and 5.66×10^{-7} for the full feature vector. Using the final trained checkpoint, predictions at all 1564 strict test anchors differed by at most 2.11×10^{-5} (mean absolute difference 1.69×10^{-6}), satisfying the 10^{-4} equivalence criterion.

4.2. Anticipatory prediction across 0–200 ms

Using the supporting full-input system-reference model, predictive performance generally degraded as anticipation increased beyond 50 ms (Table 4). Mean MAE was 0.145 Nm/kg at 0 ms and 0.144 Nm/kg at 50 ms, then increased to 0.152, 0.169 and 0.186 Nm/kg at 100, 150 and 200 ms, respectively. Mean R^2 decreased consistently from 0.442 at 0 ms to 0.215 at 200 ms, and mean Pearson r decreased from 0.640 to 0.492. Thus, the shortest two horizons were similar in absolute error, whereas predictive fit declined progressively as the target was moved farther into the future. These supporting horizon results characterize the broader causal pipeline and are separate from the primary 100-ms RF-TCN endpoint.

Horizon (ms)	MAE	RMSE	NRMSE	R^2	Pearson r
0	0.145	0.202	0.126	0.442	0.640
50	0.144	0.206	0.129	0.435	0.628
100	0.152	0.219	0.137	0.383	0.614
150	0.169	0.242	0.152	0.278	0.528
200	0.186	0.254	0.159	0.215	0.492

Table 4. Mean four-joint Online-Causal performance across prediction horizons for the supporting full-input system-reference model (seed 2026).

4.3. Primary 100-ms torque-proxy forecasting performance

At the primary 100-ms endpoint, the validation-selected RF-TCN achieved mean MAE 0.1326 Nm/kg, RMSE 0.2069 Nm/kg, NRMSE 0.1297, R^2 0.4427 and Pearson r 0.6529 against the physics-derived proxy targets on 1564 strict-test windows (Table 5). Knee pitch had the lowest MAE (0.0838 Nm/kg), while ankle pitch had the strongest fit (R^2 0.6327; r 0.8000). Hip pitch achieved MAE 0.1719 Nm/kg and r 0.6491. Hip yaw remained the weakest degree of freedom by fit (R^2 0.1625; r 0.4117), despite a lower absolute MAE than hip pitch or ankle pitch. These values quantify surrogate-forecasting agreement and are not measures of agreement with independently measured biological joint moments.

Joint	MAE	RMSE	NRMSE	R^2	Pearson r
Hip yaw	0.125	0.175	0.113	0.163	0.412
Hip pitch	0.172	0.252	0.127	0.418	0.649
Knee pitch	0.084	0.130	0.106	0.558	0.751
Ankle pitch	0.150	0.271	0.173	0.633	0.800
Mean	0.133	0.207	0.130	0.443	0.653

Table 5. Joint-specific strict-test performance at 100 ms for the validation-selected RF-TCN using causal gait-context inputs only (n = 1564 windows).

4.4. Available sparse sEMG modality ablation

Within the available four-channel distal-muscle montage, input representation had a larger effect than modest architectural changes in the supporting screening experiments (Table 6; Fig. 4). Causal-phase-only input achieved the lowest mean MAE (0.136 Nm/kg), whereas contact-only input achieved the highest mean R^2 (0.451) and a closely similar MAE of 0.137 Nm/kg. The full sEMG + causal-phase input achieved MAE 0.152 Nm/kg, R^2 0.383 and r 0.614, and engineered sEMG alone achieved MAE 0.238 Nm/kg with mean R^2 -0.036. Paired 5-s block bootstrap analysis supported this within-configuration pattern. Relative to contact-only, sEMG + causal phase increased mean MAE by 0.0145 Nm/kg (95% confidence interval (CI), 0.0069 to 0.0260), with ΔR^2 = -0.0676 (95% CI, -0.1088 to -0.0240) and Δr = -0.0326 (95% CI, -0.0606 to -0.0040). Relative to causal-phase-only, the full multimodal condition increased mean MAE by 0.0160 Nm/kg (95% CI, 0.0080 to 0.0235). The contact-only versus causal-phase-only difference was small and uncertain (phase-only minus contact-only ΔMAE = -0.00145 Nm/kg; 95% CI, -0.0119 to 0.0151). These results justify a context-only final model for this recording; they do not test whether a richer multi-muscle sEMG montage would improve true joint-moment estimation.

Input set	Mean MAE	Mean R^2	Mean r
Causal phase only	0.136	0.443	0.657
Contact only	0.137	0.451	0.647
sEMG + causal phase	0.152	0.383	0.614
sEMG + contact	0.153	0.387	0.603
Engineered sEMG only	0.238	-0.036	0.076

A secondary residual-information audit was performed on training and validation data only. After subtracting predictions from the existing 100-ms context model, Ridge probes were fitted to predict the remaining torque-proxy residuals from the 32 engineered sEMG variables at causal lead times from 0 to 250 ms. The best mean validation incremental R^2 was -0.0607 (0-ms lead), providing no evidence that these particular four-channel features explained stable residual information beyond gait context in this recording. This diagnostic result is configuration-specific and should not be interpreted as evidence against sEMG as a modality in general.

Table 6. Input-modality ablation at 100 ms using the supporting activity-conditioned TCN and the available four-channel distal-muscle sEMG montage.

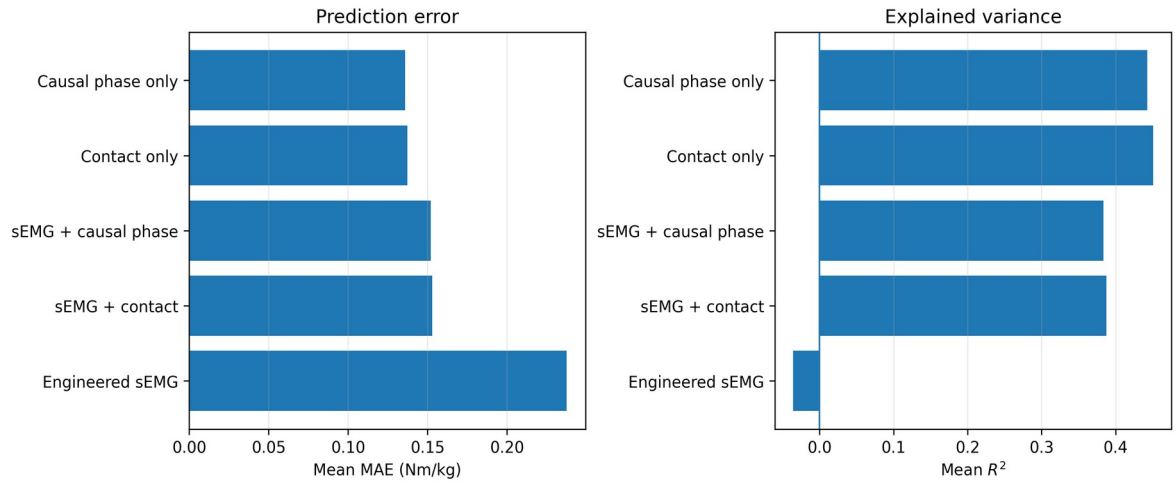


Fig. 4. Input-modality ablation at 100 ms for the supporting activity-conditioned TCN. The comparison is restricted to the available four-channel distal-muscle sEMG montage; lower MAE and higher R^2 indicate better proxy-forecasting performance.

4.5. Model comparison

The matched-input architecture benchmark did not identify the activity-conditioned TCN as uniformly superior (Table 7; Fig. 5). Standard TCN achieved the lowest mean MAE (0.147 Nm/kg). GRU and LSTM followed at 0.150 and 0.151 Nm/kg, the activity-conditioned TCN achieved 0.152 Nm/kg, the causal transformer 0.153 Nm/kg and MLP 0.155 Nm/kg. MLP and standard TCN achieved nearly identical mean R^2 (0.411 and 0.410), while MLP had the highest mean Pearson correlation (r = 0.633). Ridge regression was markedly worse than all nonlinear models. This comparison uses identical sEMG + causal-phase inputs and therefore addresses broad model-class sensitivity; it is supporting evidence rather than a selection contest against the primary context-only RF-TCN.

Model	Mean MAE	Mean R^2	Mean r
Standard TCN	0.147	0.410	0.629
GRU	0.150	0.388	0.596
LSTM	0.151	0.355	0.562
Activity-conditioned TCN	0.152	0.383	0.614
Transformer	0.153	0.341	0.594
MLP	0.155	0.411	0.633
Ridge	0.338	-1.482	0.316

Table 7. Architecture comparison at 100 ms using identical sEMG + causal-phase inputs.

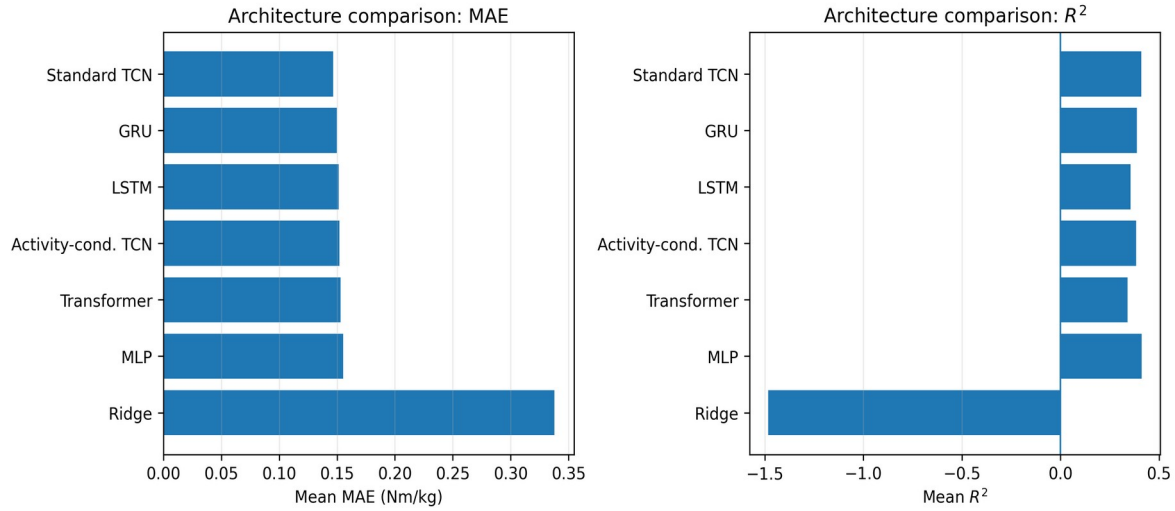


Fig. 5. Architecture comparison under matched 100-ms sEMG + causal-phase inputs. Standard TCN achieved the lowest mean MAE, whereas MLP and standard TCN had nearly identical mean R^2. These screening results motivate a focus on representation and temporal design rather than a claim of universal architecture superiority.

4.6. RF-TCN refinement analysis and uncertainty

Validation-only screening selected the width-128, kernel-9 RF-TCN (C1) at mean validation MAE 0.11017 Nm/kg, compared with 0.11525 Nm/kg for the reused original causal-context reference. On the strict 1564-window 100-ms test set, RF-TCN achieved MAE 0.13262 Nm/kg versus 0.13599 Nm/kg for the original context reference, an absolute ΔMAE (RF-TCN - original) of -0.00336 Nm/kg and a relative reduction of 2.47%. The paired 5-s moving-block bootstrap 95% CI was -0.00503 to -0.00072 Nm/kg. The corresponding RMSE decreased from 0.20881 to 0.20693 Nm/kg, but its bootstrap interval crossed zero; mean R^2 was essentially unchanged (0.44275 versus 0.44267), and Pearson r changed from 0.65676 to 0.65291. Thus, the supported result is a small within-recording reduction in absolute proxy error rather than a broad performance improvement. No control-relevance threshold was defined for proxy MAE, so the absolute change of 0.00336 Nm/kg should not be interpreted as an established actuator-level or clinical benefit.

The package-level change was not uniform across joints: relative MAE decreased by 3.47% for hip pitch, 3.00% for knee pitch and 5.63% for ankle pitch, but increased by 3.53% for hip yaw. The subsequent controlled kernel-only ablation provided a narrower estimate of the receptive-field contribution. Across seeds 2024-2028, K7 (373-ms receptive field) achieved mean MAE 0.134970 ± 0.000449 Nm/kg, whereas K9 (497 ms) achieved 0.133802 ± 0.001061 Nm/kg, an average difference of -0.001168 Nm/kg (approximately 0.87% lower for K9). K9 had lower MAE in 3 of 5 seeds. For seed 2026, the paired 5000-repetition 5-s moving-block bootstrap gave mean ΔMAE (K9 - K7) = -0.00310 Nm/kg with 95% CI -0.00616 to 0.00015, which crossed zero. Joint-specific seed-2026 intervals excluded zero for knee pitch and narrowly for ankle pitch, but not for hip yaw, hip pitch or the four-joint mean. Receptive-field matching is therefore treated as a plausible but modest temporal-design factor rather than a standalone, uniformly supported source of improvement.

Controlled model	Kernel	Nominal RF	Five-seed MAE	Five-seed R^2	Five-seed r
K7 context-only TCN	7	373 ms	0.134970 ± 0.000449	0.44207 ± 0.00234	0.65254 ± 0.00279
K9 RF-TCN (final)	9	497 ms	0.133802 ± 0.001061	0.44448 ± 0.00527	0.65382 ± 0.00423

Table 8. Controlled kernel-only ablation at the primary 100-ms endpoint. K7 and K9 are identical context-only simple-head TCNs except for kernel size and the resulting nominal receptive field. K9 achieved lower MAE in 3/5 seeds; the seed-2026 paired 5-s moving-block-bootstrap 95% CI for mean K9 - K7 ΔMAE was -0.00616 to 0.00015 Nm/kg.

4.7. Offline-Reference versus Online-Causal preprocessing

Causalizing the predictor input did not reduce performance in the matched system-reference comparison. Mean MAE was 0.1525 Nm/kg for Offline-Reference and 0.1520 Nm/kg for Online-Causal (-0.35% relative difference). Mean RMSE changed from 0.2204 to 0.2192 Nm/kg, R^2 from 0.380 to 0.383 and r from 0.6148 to 0.6145. Paired 5-s block-bootstrap analysis showed no clear difference: for Offline-Reference minus Online-Causal, ΔMAE was 0.00054 Nm/kg (95% CI, -0.00049 to 0.00189), ΔR^2 was -0.00356 (95% CI, -0.01421 to 0.00425), and Δr was 0.00033 (95% CI, -0.00537 to 0.00524). The small numerical advantage of the online condition is not interpreted as evidence that causal preprocessing is intrinsically superior; the supported conclusion is that the deployable causal route preserved retrospective-reference accuracy.

Metric	Offline-Reference	Online-Causal	Difference
Mean MAE (Nm/kg)	0.1525	0.1520	-0.35%
Mean RMSE (Nm/kg)	0.2204	0.2192	-0.53%
Mean NRMSE	0.1376	0.1369	-0.50%
Mean R^2	0.380	0.383	+0.0036
Mean Pearson r	0.6148	0.6145	-0.0003

Table 9. Matched Offline-Reference versus Online-Causal performance at 100 ms for the supporting full-input system-reference model.

4.8. Training stability and block-bootstrap uncertainty

After the RF-TCN configuration was frozen, five-seed replication showed low sensitivity to stochastic initialization and optimization. Strict-test MAE values were 0.13291, 0.13372, 0.13262, 0.13489 and 0.13486 Nm/kg for seeds 2024-2028, respectively. Across seeds, mean ± sample standard deviation (SD) was 0.13380 ± 0.00106 Nm/kg for MAE, 0.20623 ± 0.00109 Nm/kg for RMSE, 0.12929 ± 0.00076 for NRMSE, 0.44448 ± 0.00527 for R^2 and 0.65382 ± 0.00423 for Pearson r. The MAE range was 0.13262-0.13489 Nm/kg and the across-seed MAE coefficient of variation was approximately 0.79%, indicating that the selected RF-TCN endpoint was not strongly dependent on a single random seed. Exact matched original causal-context baseline results were not available for all five seeds, so this analysis supports RF-TCN training stability rather than five-seed superiority over the baseline.

Seed	MAE	RMSE	NRMSE	R^2	Pearson r	n
2024	0.13291	0.20653	0.12950	0.44318	0.65325	1564
2025	0.13372	0.20443	0.12806	0.45355	0.66102	1564

Seed	MAE	RMSE	NRMSE	R^2	Pearson r	n
2026	0.13262	0.20693	0.12972	0.44267	0.65291	1564
2027	0.13489	0.20607	0.12915	0.44324	0.65204	1564
2028	0.13486	0.20718	0.13002	0.43976	0.64990	1564
Mean ± SD	0.13380 ± 0.00106	0.20623 ± 0.00109	0.12929 ± 0.00076	0.44448 ± 0.00527	0.65382 ± 0.00423	5 seeds

Table 10. Five-seed strict-test stability of the final RF-TCN at the 100-ms endpoint. Seed 2026 reuses the validation-selected checkpoint; seeds 2024, 2025, 2027 and 2028 were trained only after the architecture was frozen. All seeds use the same 1564 strict-test anchors.

For the principal package-level RF-TCN comparison, the paired 5000-repetition 5-s moving-block bootstrap yielded ΔMAE = -0.00336 Nm/kg (RF-TCN - original context; 95% CI, -0.00503 to -0.00072). For the controlled seed-2026 kernel-only comparison, K9 - K7 mean ΔMAE was -0.00310 Nm/kg with 95% CI -0.00616 to 0.00015. Across the five training seeds, the K9 - K7 MAE differences were -0.00206, -0.00117, -0.00310, +0.00034 and +0.00014 Nm/kg, respectively. Figure 6 summarizes the paired seed behavior and seed-2026 block-bootstrap intervals. These intervals quantify within-recording temporal uncertainty and should not be interpreted as between-participant uncertainty.

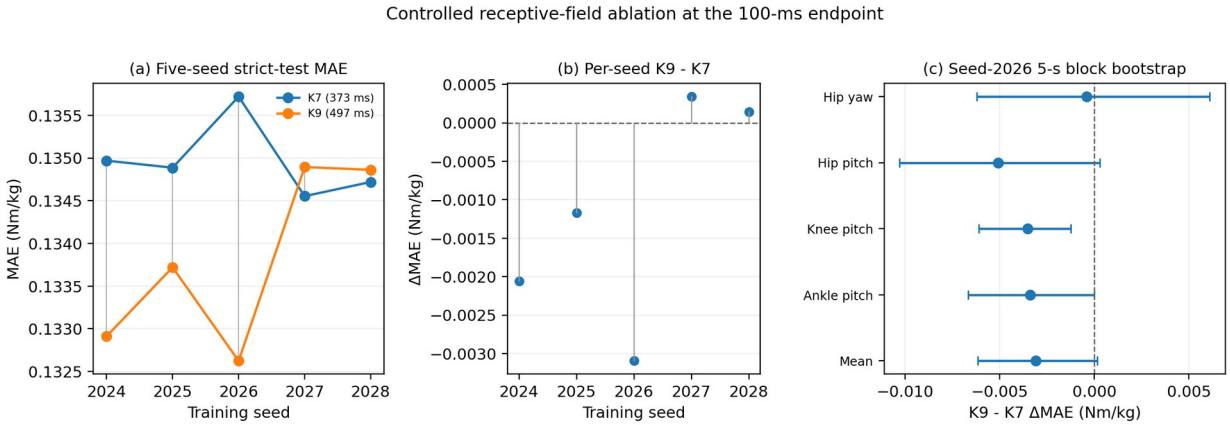


Fig. 6. Controlled receptive-field ablation at 100 ms. (a) Five-seed strict-test MAE for otherwise identical K7 (373-ms receptive field) and K9 (497-ms receptive field) context-only TCNs. (b) Per-seed K9 - K7 MAE differences; negative values favor K9. (c) Seed-2026 paired 5-s moving-block-bootstrap ΔMAE intervals (5000 repetitions) by joint and for the four-joint mean. The mean interval crosses zero, so the kernel-only effect is interpreted as a small numerical tendency rather than a uniformly supported improvement.

4.9. Activity-specific behavior

Performance of the supporting system-reference model varied across activity segments. Accelerating and decelerating walking produced comparatively strong knee and ankle results, whereas fast jogging was more difficult, particularly for hip pitch. Small-jump estimates were based on only 14 strict test windows and are therefore descriptive. Several squat target traces had extremely small within-segment variance, causing R^2 to become numerically unstable despite finite MAE/RMSE. For near-constant activity subsets, absolute errors are more interpretable than R^2. Standing contained only three test windows and was classified as insufficient support. Activity-specific RF-TCN estimates were not separately promoted as confirmatory endpoints.

4.10. Stateful computational-latency benchmark

The final RF-TCN online streaming computational benchmark timed the stateful marker-coordinate/PCM-to-prediction software path after 50 warm-up updates, with cold start reported separately. Across 6069 steady-state updates, CPU complete-update median, p95 and p99 latencies were 33.84, 67.15 and 71.37 ms, respectively; 1658 updates (27.32%) exceeded the 50-ms scheduling period. On the RTX 5060 Laptop GPU, median, p95 and p99 complete-update latencies were 19.32, 21.78 and 25.03 ms, the maximum was 34.77 ms, and no timed update exceeded 50 ms (0/6069). GPU model inference itself was a small part of the total cost (median 1.47 ms; p95 2.51 ms), whereas the stateful feature/context update dominated the median computational cost (17.31 ms). Cold-start complete-update latencies were 139.19 ms on CPU and 274.69 ms on GPU and were excluded from the steady-state percentiles. These results demonstrate GPU computational feasibility for the implemented 20-Hz software loop, not hard-real-time certification or camera-to-actuator latency.

Table 11. Final RF-TCN steady-state online streaming computational timing after 50 warm-up updates. Cold-start latency is reported separately and excluded from percentile summaries. The timed path begins from raw PCM and the unfilled marker-coordinate trajectory; it does not include camera tracking, dual-view reconstruction, controller/communication or actuator dynamics.

Device	Median	p95	p99	Max	50-ms misses	Cold start
CPU	33.84 ms	67.15 ms	71.37 ms	105.63 ms	1658/6069 (27.32%)	139.19 ms

RTX 5060 Laptop GPU	19.32 ms	21.78 ms	25.03 ms	34.77 ms	0/6069 (0%)	274.69 ms

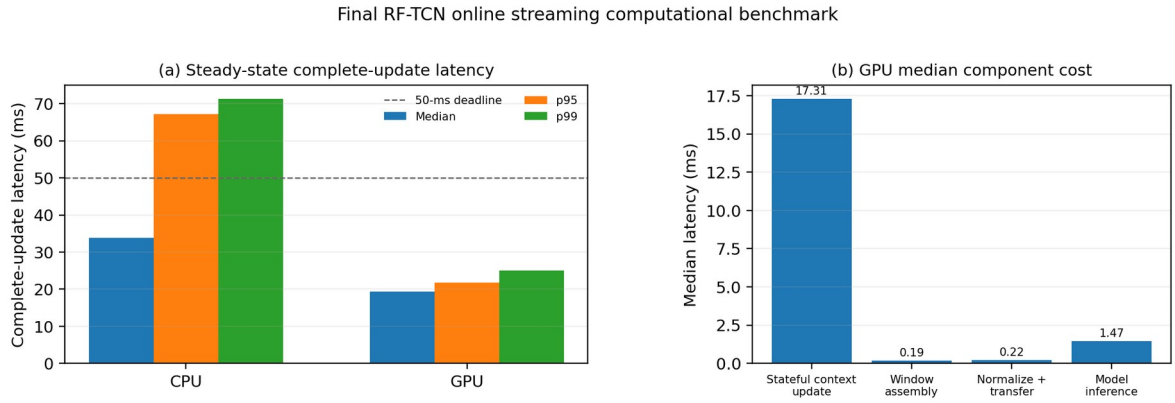


Fig. 7. Final RF-TCN online streaming computational benchmark. (a) Complete-update steady-state median, p95 and p99 latency on CPU and GPU relative to the 50-ms scheduling period; CPU missed 27.32% of deadlines, whereas GPU missed 0/6069 timed updates. (b) Median GPU component costs, showing that stateful feature/context updating dominated model inference. Cold-start latency is analyzed separately. This is software computational profiling, not hard-real-time or sensor-to-actuator certification.

5. Discussion

5.1. Main findings

This study asked whether an inexpensive motion-capture-derived pipeline could forecast a fixed physics-derived torque surrogate under genuinely online-causal conditions. The main findings are: (1) training-only synchronization and stateful preprocessing removed future-data dependence from predictor construction; (2) 100-ms proxy forecasting remained feasible with a context-only RF-TCN and was training-stable across five seeds; (3) the available four-channel distal-muscle sEMG montage did not add predictive value beyond marker-derived gait context; (4) the package-level RF-TCN refinement reduced MAE by 2.47%, but the controlled kernel-only K7-versus-K9 comparison showed only a small, non-uniform five-seed advantage for K9; and (5) the final RF-TCN stateful benchmark achieved GPU p99 25.03 ms with zero steady-state 50-ms deadline misses, while CPU timing did not satisfy the same scheduling target. These findings concern a subject-specific computational surrogate and should not be read as validation of biological joint moments or safety-critical exoskeleton control.

5.2. Scope of the supervised target and source coupling

The torque proxies are not independently measured labels. They are outputs of an offline engineering dynamics mapping driven partly by the same reconstructed marker trajectories from which contact and phase are obtained. The correct interpretation is therefore narrower than conventional joint-moment estimation: the model forecasts the future output of a fixed surrogate from causal past state. At the primary horizon, x(t) contains only information available at or before t, whereas the target τproxy(t + 100 ms) depends on reconstructed motion 100 ms later and is unavailable to the predictor. This temporal separation prevents a direct same-sample algebraic identity, but it does not remove common-source dependence. The progressive decline in R^2 and r with increasing horizon is compatible with a non-trivial forecasting problem, yet it cannot substitute for external kinetic validation. Force-plate-supported inverse dynamics or another independently validated reference is required before the proxy can be interpreted as biological joint torque.

5.3. Why causalization did not reduce accuracy

The near-equality of Offline-Reference and Online-Causal performance is methodologically useful because retrospective interpolation and signal repair can otherwise make a forecasting pipeline appear more deployable than it is. Replacing those operations with past-only state updates did not impose a measurable penalty in the supporting reference model. One interpretation is that the dominant structure of this recording is low-frequency and gait-locked: a bounded causal history, persistent marker state and debounced contact/phase provide sufficient continuity for the proxy-forecasting task. This result should not be generalized to other motion-capture systems and is not itself a novel modeling contribution; rather, it verifies that the reported forecast metrics are not dependent on future-aware predictor preprocessing.

The result also provides a methodological caution for forecasting studies. A model can appear to predict 100-200 ms ahead while receiving features that themselves depend on future samples. Our train-only synchronization, prefix tests and stateful replay were designed to eliminate this ambiguity. The consistent decrease in R^2 and r with horizon is compatible with a genuine increase in prediction difficulty as the target is moved further into the future, even though MAE at 0 and 50 ms was essentially unchanged.

5.4. Interpretation of the sparse distal-muscle sEMG ablation

Within the available four-channel distal-muscle montage, gait-context-only representations produced lower proxy-forecasting error than inputs containing the recorded sEMG. This result should not be generalized to sEMG as a modality. The four channels sample only bilateral gastrocnemius and peroneus longus, whereas hip, knee and ankle moments depend on a substantially broader muscle set including proximal thigh and hip musculature. The recorded myoelectric field is therefore under-complete for a multi-joint inference task, and the experiment is best interpreted as an ablation of this specific low-cost montage rather than a test of the physiological value of muscle activation signals.

A second factor is target construction. The torque proxy is generated from kinematics, segment geometry, contact state and an approximate external-load model. Causal contact and phase are mechanically and statistically close to this target-generation process, creating common-source structure that favors marker-derived context. Therefore, the ablation cannot establish that contact or phase is superior to sEMG for true biological joint moments. It shows only that, for this source-coupled proxy and this sparse distal montage, marker-derived gait context is the more informative low-dimensional representation. Richer sEMG coverage should be evaluated against an independently validated kinetic target before physiological conclusions are drawn.

For this recording, the lack of incremental value justified omitting sEMG from the selected final predictor, but it does not justify excluding sEMG from future systems. The train/validation residual-information audit reached the same configuration-specific conclusion: Ridge probes of context-model residuals produced no positive mean validation incremental R^2 across the tested causal EMG leads. A stronger experiment would add quadriceps, hamstring and gluteal coverage, document electrode and amplifier characteristics, and evaluate incremental value against force-plate-supported kinetics across participants and sessions.

5.5. RF-TCN temporal design and architectural complexity

The architecture results argue for controlled temporal design rather than indiscriminate complexity. Under identical 38-channel inputs, standard TCN, GRU, LSTM, transformer, MLP and the activity-conditioned TCN produced broadly comparable nonlinear-model performance, with no architecture dominating every metric. This is consistent with the structure of the present task: the input history is short (500 ms), the context is low-dimensional and locomotion is strongly periodic, conditions under which a causal dilated convolutional encoder is a natural baseline. Receptive-field matching itself is not a new sequence-modeling mechanism. Its role here is practical: the fixed 500-ms buffer contains 500 samples, whereas the kernel-7 stack reaches 373 ms and kernel 9 reaches 497 ms.

The post-selection five-seed replication supports the narrower claim that the RF-TCN endpoint is training-stable within this dataset. Mean strict-test MAE was 0.13380 ± 0.00106 Nm/kg, with all five K9 seeds between 0.13262 and 0.13489 Nm/kg. The controlled K7/K9 experiment further prevents over-attributing the package-level 2.47% improvement to receptive-field matching. K9 reduced mean MAE from 0.134970 to 0.133802 Nm/kg across seeds, but it was better in only 3/5 seeds, and the seed-2026 mean block-bootstrap interval crossed zero. Accordingly, the evidence supports a modest numerical tendency associated with making almost the entire fixed history accessible, not a strong or universal kernel-size effect.

The controlled result also sharpens the interpretation of model complexity. The useful contribution of the final predictor is not a claim that changing a TCN kernel constitutes a novel architecture; rather, it shows that a simple causal context model can be competitive when its temporal support is matched to the available history and when leakage controls are enforced. The fact that knee and ankle showed clearer seed-2026 kernel effects than hip yaw or hip pitch suggests that temporal-context requirements may be degree-of-freedom dependent, but the present single-recording design is insufficient to assign a physiological mechanism to those differences. Future model development should therefore prioritize externally validated kinetics, cross-subject/session adaptation, and biomechanics-aware multi-task or physics-informed objectives before further architecture proliferation.

5.6. Computational deployment and timing

The final timing results are substantially clearer when model inference is separated from the surrounding stateful computation. On GPU, RF-TCN inference had median latency 1.47 ms, while the full timed software update had median/p95/p99 latencies of 19.32/21.78/25.03 ms with no steady-state 50-ms deadline misses. The dominant median cost was the stateful feature/context update (17.31 ms), not the neural network itself. CPU execution was not sufficient for the same scheduling target: p95/p99 were 67.15/71.37 ms and 27.32% of updates missed 50 ms. These observations support GPU computational feasibility for the implemented 20-Hz research loop and identify preprocessing/state maintenance as the more important optimization target than network compression.

The benchmark remains narrower than an end-to-end exoskeleton timing claim. It begins from raw PCM and an unfilled marker-coordinate trajectory, not raw camera frames, and it excludes image tracking, dual-view reconstruction, controller execution, communication and actuator dynamics. Cold-start latency was also much larger than steady-state timing. Therefore, zero steady-state GPU misses in this software benchmark should not be interpreted as deterministic hard-real-time certification or evidence of closed-loop safety.

5.7. Relation to validated kinetic-estimation studies

Direct numerical comparison with validated torque-estimation studies should be made cautiously. Studies trained on force-plate-supported inverse-dynamics targets and richer EMG/kinematic inputs estimate a different quantity and often report higher R^2 or lower normalized errors [11–15]. The present target is a source-coupled engineering proxy, so its numerical error cannot be interpreted on the same biological scale until external kinetic validation is available. The appropriate contribution is therefore a subject-specific feasibility demonstration of causal surrogate forecasting, synchronization discipline, streaming equivalence and computational profiling, not superiority over laboratory-ground-truth joint-moment models.

The system framing remains relevant to future assistive-device research because it connects accessible visual motion capture, an offline dynamics surrogate and a measured online forecast loop. If the proxy is subsequently validated against independent kinetics, such a pipeline could be investigated for latency compensation or supervisory prediction. The present data do not establish that the forecast is suitable as a direct exoskeleton torque command.

5.8. Limitations

- The dataset contains a single participant. All performance estimates are subject-specific and within-session; the bootstrap quantifies temporal sampling uncertainty but cannot substitute for between-participant validation.
- The torque outputs are physics-derived proxies rather than independently measured joint moments. External loading is approximated and no force plate or laboratory inverse-dynamics reference was available; biological joint-moment accuracy and absolute physical calibration are therefore unknown. In addition, proxy targets and gait-context predictors partly share the same marker source. The reported task is future surrogate forecasting, not independent sensor-to-ground-truth torque validation.
- The two-view 3D reconstruction uses view scaling and manual origin alignment rather than calibrated stereo triangulation. Its absolute 3D accuracy has not been validated against Vicon, Qualisys or another reference system.
- Archived age, sex and height metadata, camera model, EMG acquisition hardware, electrode model, gain and hardware filter settings are incomplete. The participant body mass used in the reported analysis was 60 kg. The paper reports only parameters traceable to the archived experiment and repository and does not infer missing equipment specifications.
- The four-channel sEMG montage samples only gastrocnemius and peroneus longus bilaterally and is insufficient to draw general conclusions about richer myoelectric sensing.
- Video-sEMG synchronization was not hardware-triggered. A composite signal-alignment search performed only on the training partition selected a 559-frame (9.317-s) offset, which was frozen before validation/test processing and used for all final results. A local training-only sensitivity check confirmed 559 as the highest score among 529, 544, 559, 574 and 589 frames, but periodic locomotor signals can contain multiple plausible correlation peaks; the numerical optimum is therefore a registration parameter, not an independently verified physical time zero.
- No physical exoskeleton closed-loop experiment was performed. The final RF-TCN benchmark measured marker-coordinate/PCM-to-prediction software timing only. GPU steady-state p99 was below 50 ms with no timed misses, but camera tracking, 3D reconstruction, communication, control computation and actuator dynamics were not included, cold-start latency was larger, and CPU timing missed 27.32% of 50-ms deadlines. No hard-real-time, safety, stability or actuator-level benefit is inferred.

- The RF-TCN is a within-recording model-development analysis rather than an externally validated population model. C1 was selected from five kernel-9 context-only candidates using validation data before strict-test inference, while the broader decision to focus on gait-context inputs was informed by earlier modality screening. The paired bootstrap therefore supports within-recording temporal error reduction, not independent confirmatory generalization.
- The original causal-context reference and RF-TCN differ in both temporal receptive-field design and auxiliary activity conditioning, so the package-level 2.47% MAE reduction cannot be attributed exclusively to kernel size. The separate controlled K7/K9 ablation reduced five-seed mean MAE by only 0.001168 Nm/kg (about 0.87%), was favorable in 3/5 seeds and had a seed-2026 mean bootstrap interval that crossed zero. Receptive-field matching is therefore reported as a modest design tendency rather than a statistically uniform architectural improvement.

6. Conclusions

This subject-specific feasibility study integrated low-cost dual-view marker motion capture, an approximate offline physics model and a strictly online-causal predictor to forecast physics-derived lower-limb torque proxies. The central result is methodological rather than a claim of biological torque recovery: training-only synchronization, causal predictor construction and stateful replay enabled leakage-controlled 100-ms forecasting

of a fixed model-derived surrogate. The selected six-channel RF-TCN achieved strict-test MAE 0.1326 Nm/kg and five-seed MAE 0.1338 ± 0.0011 Nm/kg. Relative to the earlier causal-context reference, the complete RF-TCN refinement package reduced MAE by 2.47% with a paired temporal-bootstrap interval below zero, but RMSE, R^2 and r did not show supported improvement and hip-yaw MAE worsened. A controlled five-seed K7-versus-K9 experiment isolated the receptive-field factor and showed a smaller average MAE advantage for the 497-ms design (0.133802 vs 0.134970 Nm/kg), favorable in 3/5 seeds; the seed-2026 four-joint bootstrap interval crossed zero. Thus, the evidence favors careful causal temporal design over architectural complexity, without establishing a novel or universally superior TCN mechanism. In the final stateful software benchmark, GPU complete-update p95/p99 latencies were 21.78/25.03 ms with zero steady-state 50-ms deadline misses, whereas CPU missed 27.32% of deadlines. Because the study contains one participant, uses a source-coupled torque proxy without force-plate/inverse-dynamics validation, and does not include the full camera-to-actuator chain, external kinetic validation and multi-participant, multi-session evaluation remain prerequisites before claims about biological joint moments or exoskeleton control can be made.

Author Contributions

Ange Tong: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing - original draft, Writing - review & editing.

Funding

This research received no external funding.

Institutional Review Board Statement

No formal ethics-committee review was sought at the time of data collection. The study involved one adult volunteer completing non-invasive treadmill and transitional locomotor tasks while marker-based video and surface electromyography were recorded. No clinical intervention, patient data, invasive procedure or vulnerable population was involved. This statement reports the actual study history; no retrospective ethics approval or exemption is asserted.

Informed Consent Statement

The participant voluntarily agreed to take part in the recording protocol and to research use of the resulting motion and sEMG data. The manuscript figures do not display identifiable facial information.

Data Availability Statement

Source code and selected derived CSV/JSON results supporting the final reported workflow are available in the public reproducibility repository: https://github.com/Alfred-Duncan/RF-TCN, curated at commit 1d5360a4ed3dbb423b076b563bc2883ccbdd39c8. The repository includes the video/marker-processing utilities, train-only synchronization and local sensitivity analysis, online-causal feature pipeline, torque-proxy generation code, final RF-TCN training/evaluation code, five-seed stability results, controlled K7-versus-K9 receptive-field ablation, final streaming computational benchmark and paper-result export files. Raw participant videos, raw sEMG recordings and trained model checkpoints are not publicly distributed because of participant privacy and large binary-file constraints. Full reruns therefore require authorized local access to the original experimental files; requests for restricted raw data may be directed to the corresponding author and are subject to participant consent, privacy constraints and applicable institutional requirements.

Acknowledgments

The author thanks the volunteer participant for taking part in the locomotion data-collection session.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Slade P, Kochenderfer MJ, Delp SL, Collins SH. Personalizing exoskeleton assistance while walking in the real world. Nature. 2022;610:277–282. doi:10.1038/s41586-022-05191-1.
2. Luo S, Jiang M, Zhang S, Zhu J, Yu S, Dominguez Silva I, et al. Experiment-free exoskeleton assistance via learning in simulation. Nature. 2024;630:353–359. doi:10.1038/s41586-024-07382-4.
3. Molinaro DD, Scherpereel KL, Schonhaut EB, Evangelopoulos G, Shepherd MK, Young AJ. Task-agnostic exoskeleton control via biological joint moment estimation. Nature. 2024;635:337–344. doi:10.1038/s41586-024-08157-7.
4. Uhlrich SD, Falisse A, Kidziński Ł, Muccini J, Ko M, Chaudhari AS, Hicks JL, Delp SL. OpenCap: Human movement dynamics from smartphone videos. PLoS Comput Biol. 2023;19:e1011462. doi:10.1371/journal.pcbi.1011462.
5. Peña-Trabalon A, Moreno-Vegas S, Estebanez-Campos MB, Nadal-Martinez F, Garcia-Vacas F, Prado-Novoa M. A Low-Cost Validated Two-Camera 3D Videogrammetry System Applicable to Kinematic Analysis of Human Motion. Sensors. 2025;25:4900. doi:10.3390/s25164900.

6. Delp SL, Anderson FC, Arnold AS, Loan P, Habib A, John CT, Guendelman E, Thelen DG. OpenSim: open-source software to create and analyze dynamic simulations of movement. IEEE Trans Biomed Eng. 2007;54:1940–1950. doi:10.1109/TBME.2007.901024.
7. Mundt M, Thomsen W, Witter T, Koeppe A, David S, Bamer F, Potthast W, Markert B. Prediction of lower limb joint angles and moments during gait using artificial neural networks. Med Biol Eng Comput. 2020;58:211–225. doi:10.1007/s11517-019-02061-3.
8. Mundt M, Koeppe A, David S, Bamer F, Potthast W, Markert B. Prediction of ground reaction force and joint moments based on optical motion capture data during gait. Med Eng Phys. 2020;86:29–34. doi:10.1016/j.medengphy.2020.10.001.
9. Altai Z, Boukhennoufa I, Zhai X, Phillips A, Moran J, Liew BXW. Performance of multiple neural networks in predicting lower limb joint moments using wearable sensors. Front Bioeng Biotechnol. 2023;11:1215770. doi:10.3389/fbioe.2023.1215770.
10. Hossain MSB, Guo Z, Choi H. Estimation of Lower Extremity Joint Moments and 3D Ground Reaction Forces Using IMU Sensors in Multiple Walking Conditions: A Deep Learning Approach. IEEE J Biomed Health Inform. 2023;27:2829–2840. doi:10.1109/JBHI.2023.3262164.
11. Wang M, Chen Z, Zhan H, Zhang J, Wu X, Jiang D, Guo Q. Lower Limb Joint Torque Prediction Using Long Short-Term Memory Network and Gaussian Process Regression. Sensors. 2023;23:9576. doi:10.3390/s23239576.
12. Zhang L, Soselia D, Wang R, Gutierrez-Farewik EM. Lower-Limb Joint Torque Prediction Using LSTM Neural Networks and Transfer Learning. IEEE Trans Neural Syst Rehabil Eng. 2022;30:600–609. doi:10.1109/TNSRE.2022.3156786.
13. Zhang L, Soselia D, Wang R, Gutierrez-Farewik EM. Estimation of Joint Torque by EMG-Driven Neuromusculoskeletal Models and LSTM Networks. IEEE Trans Neural Syst Rehabil Eng. 2023;31:3722–3731. doi:10.1109/TNSRE.2023.3315373.
14. Kizyte A, Lei Y, Wang R. Influence of Input Features and EMG Type on Ankle Joint Torque Prediction With Support Vector Regression. IEEE Trans Neural Syst Rehabil Eng. 2023;31:4286–4294. doi:10.1109/TNSRE.2023.3323364.
15. Wang Z, Chen C, Chen H, Zhou Y, Wang X, Wu X. Dual Transformer Network for Predicting Joint Angles and Torques From Multi-Channel EMG Signals in the Lower Limbs. IEEE J Biomed Health Inform. 2025. doi:10.1109/JBHI.2025.3555255.
16. Zhang Y, Ling Z, Cao GZ, Li LL, Diao D, Cui F. A Fast Calibration Method for an sEMG-Based Lower Limb Joint Torque Estimation Model. Biomed Signal Process Control. 2024;93:106188. doi:10.1016/j.bspc.2024.106188.
17. Foroutannia A, Akbarzadeh-T MR, Akbarzadeh A. A deep learning strategy for EMG-based joint position prediction in hip exoskeleton assistive robots. Biomed Signal Process Control. 2022;75:103557. doi:10.1016/j.bspc.2022.103557.
18. Li M, Deng J, Zha F, Qiu S, Wang X, Chen F. Towards Online Estimation of Human Joint Muscular Torque with a Lower Limb Exoskeleton Robot. Appl Sci. 2018;8:1610. doi:10.3390/app8091610.
19. Ma Y, Wu X, Wang C, Yi Z, Liang G. Gait Phase Classification and Assist Torque Prediction for a Lower Limb Exoskeleton System Using Kernel Recursive Least-Squares Method. Sensors. 2019;19:5449. doi:10.3390/s19245449.
20. Xie H, Li G, Zhao X, Li F. Prediction of Limb Joint Angles Based on Multi-Source Signals by GS-GRNN for Exoskeleton Wearer. Sensors. 2020;20:1104. doi:10.3390/s20041104.
21. Bai S, Kolter JZ, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271. 2018.
22. Pang Z, Zhang L, Liu F, Meng A, Yu H, Hu B. Optimizing the stiffness of variable-stiffness exoskeleton based on a data-driven human hip joint power prediction model. Biomed Signal Process Control. 2026;112:108612. doi:10.1016/j.bspc.2025.108612.
23. Chen W, Lyu M, Ding X, Wang J, Zhang J. Electromyography-controlled lower extremity exoskeleton to provide wearers flexibility in walking. Biomed Signal Process Control. 2023;79:104096. doi:10.1016/j.bspc.2022.104096.
24. He Y, Li F, Li J, Liu J, Wu X. An sEMG based adaptive method for human-exoskeleton collaboration in variable walking environments. Biomed Signal Process Control. 2022;74:103477. doi:10.1016/j.bspc.2021.103477.