Applied Mathematics & Information Sciences
*An International Journal*

http://dx.doi.org/10.18576/amis/200504

# FTCL: Federated Transfer Continual Learning with Outlier-Aware Adaptive EWC v2 and Episodic Replay for Personalized sEMG Gesture Recognition

*Atef Ahmed Obeidat*[1,*] *Naqaa Atef Obeidat* [2], *and Tamara Amjad Al-Qablan*[1]

[1]Department of Information Technology, Al-Huson University College, Al-Balqa Applied University, Al-Salt, 19117, Jordan
[2]Department of Basic Medical Sciences, Jordan University of Science and Technology, Irbid, 22110, Jordan

Received: 17 Apr. 2026, Revised: 28 May. 2026, Accepted: 27 Jun. 2026
Published online: 1 Sep. 2026

**Abstract:** Surface electromyography (sEMG)-based gesture recognition for myoelectric prosthetic control is hindered by inter-subject variability, privacy constraints on centralizing biomedical data, and catastrophic forgetting during continual multi-subject adaptation. This paper aims to address these challenges jointly through a novel federated continual learning framework. We present FTCL, a federated Transfer continual learning framework. It pre-trains a hybrid BiLSTM-CNN model, distributes it via FedProx aggregation, and performs subject-specific calibration through classifier-head-only fine-tuning regularized with two novel methods: (i) OA-EWC v2 (Outlier-Aware Adaptive Elastic Weight Consolidation v2) combining rolling Mahalanobis-distance recalibration with gradient-interference lambda boosting; and (ii) a lightweight Episodic Replay Buffer (LERB) that maintains direct exposure to old subject data during sequential fine-tuning. The framework was evaluated on the FORS-EMG dataset (19 subjects, 12 gestures, 54,720 windows) under Leave-One-Subject-Out Cross-Validation (LOSO-CV). FTCL achieves 80.69% accuracy (Macro-F1 = 0.8091), outperforming LDA (44.49%) and SVM (49.49%), with all 19 folds exceeding the 70% clinical usability threshold. Ablation study results show that the lightweight replay buffer mitigates forgetting significantly, and the combination of OA-EWC v2 and replay provides a mean sequential accuracy of 76.93% and True BWT of −9.07, signifying a 52% decrease in forgetting compared to the baseline. These results prove that FTCL is a privacy-aware clinically applicable method for personalized myoelectric prosthesis control, ensuring compatibility between data privacy and continual adaptability.

**Keywords:** backward transfer, BiLSTM-CNN, catastrophic forgetting, federated continual learning, hand gesture recognition, surface electromyography

## 1 Introduction

Surface electromyography (sEMG) acquires the electrical signals of the skeletal muscles through surface electrodes and forms the main interface of modern myoelectric prosthetic systems [1]. The key problem in mapping the multi-channel sEMG signals time-series to discrete classes of gestures (hand and wrist movements) was addressed through numerous studies due to its practical importance: recognition accuracy and latency directly affect the usability and independence of the prosthesis device in patients.

Classical machine learning algorithms using handcrafted time-domain features combined with linear discrimination analysis (LDA) or support vector machines (SVM) set the initial benchmark results [2,3]. Although computationally efficient, the described methods suffer from degradation while generalizing the data over different subjects. Deep learning models such as CNN [4, 5], recurrent models [6], and hybrid CNN-RNN approaches [7] significantly increased the within-subjects recognition accuracy, but usually require fully retraining the model per each subject, which is impractical in clinical conditions. More recently, attention-based and transformer architectures have been proposed to capture long-range temporal dependencies in sEMG sequences more effectively than recurrent backbones [22], though such architectures typically increase model size and are not by themselves designed to address privacy or catastrophic-forgetting constraints.

* Corresponding author e-mail: dr.atefob@bau.edu.jo

Transfer learning (TL) algorithms intend to use the model trained on some source group of subjects for fast adjustment to the target subject [8,9]. Yet, the typical pipeline assumes the existence of a central database, which contradicts privacy policies regarding biomedical data (e.g., GDPR, HIPAA). The federated learning (FL) framework is a decentralized scheme [10]. FedProx [11] is an extension to FedAvg with an addition of a proximal regularization term to improve the convergence performance for heterogeneous datasets; this innovation is consistent with a recent trend of gradient-based and adaptive regularization optimization methods for non-convex and large-scale learning problems [23]. Despite the above innovations, federated models can still suffer from catastrophic forgetting [12]. Elastic Weight Consolidation (EWC) [13] allows solving the problem by regularizing against the update of the most important parameters, one instance of a rapidly growing continual-learning literature spanning regularization-, replay-, and architecture-based strategies [24].

This paper makes two complementary contributions. First, it presents FTCL, a federated continual learning framework integrating FedProx, TL, and EWC-regularized adaptation. Second, we introduce OA-EWC v2 and a LERB. Beyond standard accuracy, this paper emphasizes the correct measurement of catastrophic forgetting using a corrected True Sequential Backward Transfer (BWT) metric — a diagnostic that quantifies, in one number, how much adapting to new subjects degrades performance on subjects already learned.

The background for the current study comes from five emerging fields. Table 1 shows the comparison of existing works in relation to the proposed FTCL framework.

The main contributions are:

1. **FTCL framework:** Achieving 80.69% ± 4.24% LOSO-CV accuracy on FORS-EMG.
2. **Ablation study:** Isolating the contribution of federated aggregation, TL, and CL regularization.
3. **Novel methods:** OA-EWC v2 and LERB reduce True Sequential BWT from −18.76 to −9.07 (≈52% improvement).
4. **Corrected Metric:** A True Sequential BWT metric replacing a prior degenerate formulation.
5. **Empirical evidence:** Demonstrating that the dominant forgetting event is a data-quality limitation, not a gradient-conflict catastrophe.

The remainder of this paper proceeds as follows. Section 2 describes the FORS-EMG dataset, the pre-processing pipeline, and the proposed FTCL framework, including the BiLSTM-CNN architecture, FedProx aggregation, the OA-EWC v2 scheduler, and the LERB. Section 3 presents the experimental results and ablation study, discusses the clinical implications and comparison with previous works, and concludes with the study's limitations, future directions, and conclusion.

## 2 Materials and Methods

### *2.1 Dataset: FORS-EMG*

Experiments were conducted on the basis of the FORS-EMG database [19] that consisted of signals from 19 able-bodied subjects performing 12 hand/wrist movements. The signal sampling frequency was 985 Hz and the signals were obtained by means of 8-channel surface electrodes. Table 2 illustrates the major characteristics of the database.

### *2.2 Methodology*

FTCL's architecture involves three stages, namely: (i) global pre-training of a hybrid BiLSTM-CNN model; (ii) privacy-preserving Federated Aggregation via FedProx; and (iii) fine-tuning the classifier-head for each individual subject using Elastic Weight Consolidation (EWC).

Figure 1 shows the overall pipeline.

#### 2.2.1 Data Preprocessing and Feature Extraction

Each raw sEMG recording goes through a four-step processing pipeline: (1) Band-pass filtering between 20–450 Hz (Butterworth 4th order filter with an additional 50 Hz notch filter), (2) Segmentation using non-overlapping sliding windows with length of 197 samples (≈ 200ms), (3) Time domain features per channel (Mean Absolute Value, Root Mean Square, Waveform Length, Zero Crossings, Slope Sign Changes, and AR coefficients of 4th order estimated with Yule-Walker method) for a feature vector of 72 dimensions (9 × 8) per window, and (4) Min-Max normalization per subject fitted only on the training split. The identical preprocessing pipeline is used for the combined pre-training dataset, every federated node, and for each subject calibration set, thus guaranteeing that inter-subject variability is solely caused by physiology, not inconsistencies in the feature engineering process. This consistency allows the frozen backbone (Section 2.2.4) to be treated as a universal encoder for subjects who don't have raw data in common.

#### 2.2.2 Global BiLSTM-CNN Architecture

The global feature extractor is a hybrid BiLSTM-CNN designed to jointly capture local spectral correlations and long-range temporal dependencies within sEMG signals. Each 72-D feature window is processed through two convolutional blocks (Table 3), whose output is transposed and forwarded to two stacked bidirectional LSTM layers; the bidirectional outputs are globally average-pooled across the temporal dimension into a 512-D subject-agnostic embedding before the classifier

**Table 1:** Comparative Analysis of sEMG Recognition Frameworks and the Proposed FTCL

| Works | Privacy (FL) | Cross-User Gen. | Continual Learning | Forgetting Mitigation | Strict LOSO-CV |
|---|---|---|---|---|---|
| Classical ML [2,3] | × | × | × | × | × |
| Deep Learning [4,5,6,7] | × | ◐ | × | × | × |
| Transfer Learning [9] | × | ✓ | × | × | ◐ |
| Federated Learning [14,15] | ✓ | × | × | × | × |
| Continual Learning [13,16,17,18] | × | × | ✓ | ✓ | × |
| FORS-EMG Specific [19] | × | ◐ | × | × | × |
| **Proposed FTCL (Ours)** | ✓ | ✓ | ✓ | ✓ | ✓ |

*Legend: ✓ = Fully Supported; × = Not Supported; ◐ = Partially Supported / Significant Performance Drop.*

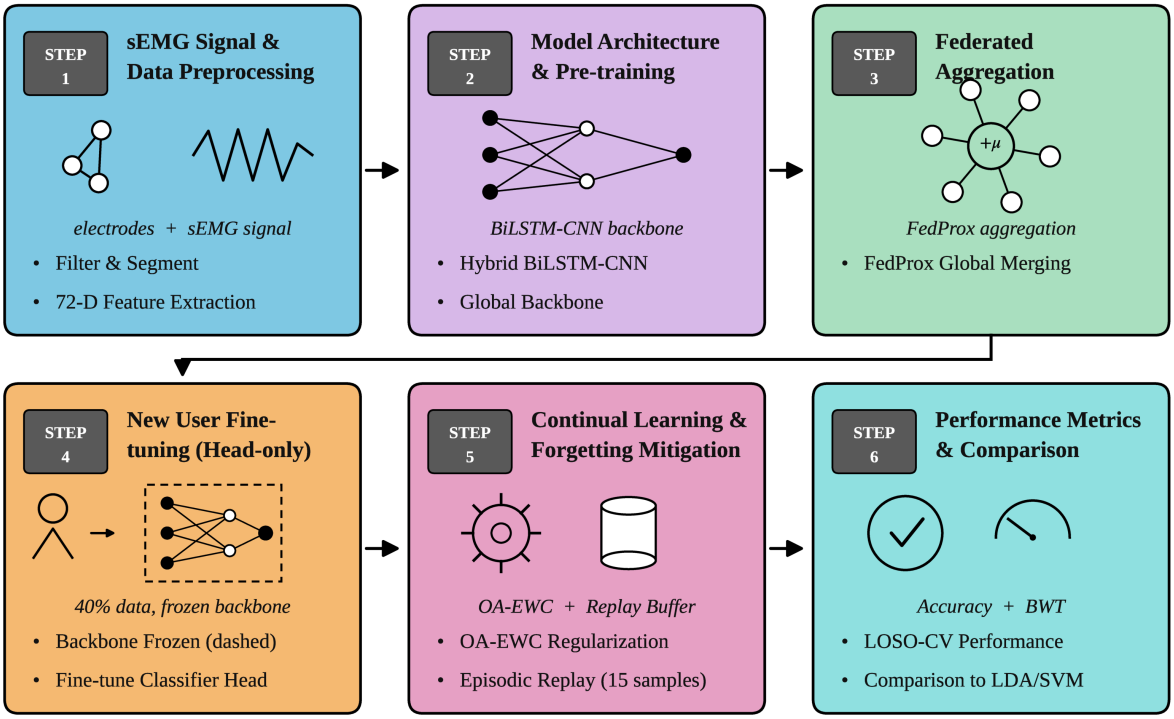


**Fig. 1:** Overview of the FTCL pipeline: data preprocessing, global BiLSTM-CNN pre-training, FedProx federated aggregation, subject-specific fine-tuning, OA-EWC v2 with episodic replay for forgetting mitigation, and performance evaluation.

head. The complete architecture comprises 2,527,244 parameters, of which 134,412 reside in the classifier head — the only parameters updated during subject-specific fine-tuning (Section 2.2.4). Table 3 summarizes the complete layer-by-layer configuration and output shapes.

### 2.2.3 Federated Aggregation: FedProx

Within each LOSO-CV fold, the training subjects (all subjects excluding the held-out test subject) are partitioned into $K = 4$ federated nodes of 4 subjects each (11,520 feature windows per node), with 3 additional subjects reserved exclusively for early-stopping validation. To accommodate the non-IID nature of multi-subject sEMG data — arising from inter-subject anatomical differences and electrode placement variability — federated aggregation is performed using FedProx [11], which augments the standard FedAvg objective with a proximal regularization term that bounds the deviation of each client's local model from the current global parameters, as given in Equation (1):

$$h(w) = \sum_{i=1}^{K} \frac{n_i}{n} F_i(w) + \frac{\mu}{2} \|w - w_t\|^2 \qquad (1)$$

**Table 2:** FORS-EMG Dataset Characteristics

| Parameter | Value |
|---|---|
| Subjects | 19 able-bodied |
| Gesture classes | 12 |
| Sampling rate | 985 Hz |
| EMG channels | 8 (Supination configuration) |
| Window length | 197 samples (200 ms) |
| Total windows | 54,720 |
| Feature dimensionality | 72 (9 features × 8 channels) |

where $F_i(w)$ denotes the local empirical loss at node $i$, $w_t$ is the current global model broadcast by the central aggregator, and $\mu = 0.01$ is the proximal coefficient controlling the trade-off between local convergence and global consistency; this value was selected in line with literature recommendations for heterogeneous data settings comparable to the inter-subject variability observed in sEMG signals. Aggregation is performed in 14 iterations of global communications, with each node conducting 3 iterations of local learning per iteration. Following the completion of the local learning iterations, the parameters are aggregated in order to create the new global model.

### 2.2.4 Subject-Specific Fine-Tuning

The fine-tuning procedure involves maintaining all the backbone parameters (consisting of CNN and BiLSTM layers) as they were originally while modifying only the gradients of the classifier head which comprises two layers and has 134,412 parameters, which accounts for 5.3% of the total network size. The training is done on 40% of each individual's windows of features using Stochastic Gradient Descent algorithm with the learning rate of $1 \times 10^{-3}$, momentum of 0.9, maximum of 30 epochs, and early stopping with patience of 10. Restricting gradient updates to the head keeps the number of trained parameters proportional to the small per-subject dataset, avoiding overfitting while retaining the population-level features learned during federated pre-training.

### 2.2.5 Proposed Methods: OA-EWC v2 and LERB

To overcome catastrophic forgetting while sequentially adapting across multiple subjects, two techniques are proposed for use in the FTCL fine-tuning stage: (i) Outlier-Aware Adaptive Elastic Weight Consolidation v2 (OA-EWC v2), a dynamic regularization scheduler combining rolling Mahalanobis-distance normalization with gradient-interference awareness to modulate the EWC regularization strength according to the statistical nature of the incoming subject; and (ii) a Lightweight Episodic Replay Buffer (LERB), which supplies representative examples from previously adapted subjects during each fine-tuning step.

Elastic Weight Consolidation with Online EMA Fisher The regularization backbone is Elastic Weight Consolidation (EWC) [13], which penalizes updates to classifier-head parameters identified as important for previously adapted subjects. The EWC-augmented fine-tuning objective is given in Equation (2):

$$L_{EWC} = L_{new} + \frac{\lambda}{2} \sum_i F_i (w_i - w_{i,old})^2 \tag{2}$$

where $F_i$ is the Fisher Information estimate of parameter $i$'s importance, $w_{i,old}$ is the parameter value immediately after the previous subject's adaptation, and $\lambda = \lambda_{base} = 4000$ (before adaptive modulation, described below) controls the regularization strength. To prevent Fisher estimates from decaying as the subject sequence grows, importance is accumulated via online Exponential Moving Average (EMA) across sequential steps, as in Equation (3):

$$F_t = \alpha F_{t-1} + (1-\alpha) F_t^{new} \tag{3}$$

with $\alpha = 0.9$, so the estimate integrates evidence across multiple previously-seen subjects rather than only the most recent one. The Fisher Information is approximated empirically over the calibration data as in Equation (4):

$$F_i = \mathbb{E}_{(x,y)\sim D}[(\nabla_{w_i} L(x,y))^2] \tag{4}$$

A diagonal approximation of the full Fisher matrix is adopted for computational tractability, consistent with the original EWC formulation [13], reducing memory complexity from $O(p^2)$ to $O(p)$ for $p$ classifier-head parameters.

Outlier-Aware Adaptive Elastic Weight Consolidation v2 (OA-EWC v2) Fixed-$\lambda$ EWC applies a uniform regularization coefficient across all adaptation steps, implicitly treating every subject as equally disruptive to consolidated knowledge. This is empirically unjustified, since inter-subject distributional shifts in sEMG are heterogeneous. OA-EWC v2 addresses this through two complementary, data-driven mechanisms that dynamically modulate $\lambda$.

**Rolling recalibration.** The distributional outlier score for subject $k$ is the Mahalanobis distance between its 512-D pre-classifier embedding mean and the reference distribution of all previously-seen subjects, as in Equation (5):

$$d_k = \sqrt{(\mu_k - \mu_{ref})^T \Sigma_{ref}^{-1} (\mu_k - \mu_{ref})} \tag{5}$$

The step-specific adaptive regularization coefficient is then computed as in Equation (6):

$$\lambda_k = \lambda_{base} \cdot \left(1 - \alpha_{outlier} \cdot \sigma\left(\frac{d_k - d_{shift}}{d_{scale}}\right)\right) \cdot \beta_{inherit} \tag{6}$$

**Table 3:** BiLSTM-CNN Architecture Summary

| Layer | Configuration | Output Shape |
|---|---|---|
| Input | 72-D feature vector | (B, 72) |
| CNN Block 1 | Conv1D(64, k=3) + BN + ReLU + MaxPool(2) | (B, 64, 36) |
| CNN Block 2 | Conv1D(128, k=3) + BN + ReLU + MaxPool(2) | (B, 128, 18) |
| BiLSTM $\times 2$ | hidden=256, bidirectional, dropout=0.3 | (B, 18, 512) |
| Global Avg Pool | Mean over temporal dimension | (B, 512) |
| Classifier | Dense(512$\rightarrow$256) + ReLU + Dropout(0.3) + Dense(12) | (B, 12) |
| Total params | 2,527,244 | Head only: 134,412 |

where $\sigma(\cdot)$ is the sigmoid function, $\alpha_{outlier} = 0.6$ bounds the maximum fractional reduction applied to outlier subjects, $d_{scale} = 1.0$ scales the sensitivity to $d_k$, and $\beta_{inherit} = 1 + \beta_{carry} \cdot \sigma(\cdot)$ carries forward partial protection from the previous step, with $\beta_{carry} = 0.3$. Rather than a fixed warm-up estimate, $d_{shift}$ is computed as a rolling quantity from the $W = 8$ most recent non-zero distances, taken at the 70th percentile, as in Equation (7):

$$d_{shift} = \text{Percentile}_{70}(\text{last } W \text{ positive values in } d_{history}) \tag{7}$$

This rolling estimation lets the reference threshold for a "typical" inter-subject distance adapt continuously to the observed cohort rather than being anchored to a potentially unrepresentative early estimate. The reference parameters $\mu_{ref}$ and $\Sigma_{ref}$ are likewise updated online, with the covariance regularized through Ledoit-Wolf shrinkage as in Equation (8):

$$\Sigma_{ref}^{(k)} = (1 - \gamma_k)\hat{\Sigma}_k + \gamma_k \tau I \tag{8}$$

where $\hat{\Sigma}_k$ is the empirical covariance of the first $k$ subjects' embedding means, $\tau = \text{tr}(\hat{\Sigma}_k)/\text{dim}$ is the isotropic target variance, and the shrinkage coefficient $\gamma_k$ follows a dimension-aware schedule: $\gamma_k = 0.9$ for $k < \text{dim}$, 0.5 for $\text{dim} \leq k < 2 \cdot \text{dim}$, and 0.1 for $k \geq 2 \cdot \text{dim}$. The inverse $\Sigma_{ref}^{-1}$ in Equation (5) is computed via eigendecomposition with a minimum eigenvalue floor of $10^{-3}$ to prevent numerical instability when $\Sigma_{ref}$ is near-singular.

**Gradient-interference lambda boosting.** The Mahalanobis distance quantifies distributional dissimilarity but does not directly assess whether the new subject's gradient updates will interfere with consolidated knowledge. Prior to each fine-tuning step, the cosine similarity between the loss gradient on the incoming subject's calibration data ($\nabla L_{new}$) and the loss gradient on a replayed batch from the LERB ($\nabla L_{old}$) is computed as in Equation (9):

$$\cos\_\text{sim}_k = \frac{\nabla L_{new} \cdot \nabla L_{old}}{\|\nabla L_{new}\| \|\nabla L_{old}\|} \tag{9}$$

A negative value indicates directionally opposed (conflicting) gradients and triggers a multiplicative boost to $\lambda_k$, with the final coefficient given by Equation (10):

$$\lambda_k' = \lambda_k \cdot (1 + \gamma_{max} \cdot \max(0, -\cos\_\text{sim}_k)) \tag{10}$$

where $\gamma_{max} = 1.0$ controls the maximum boost magnitude. This computation reuses the existing LERB batch without any additional data storage or forward passes, making it computationally negligible.

Algorithm 1 summarizes the complete OA-EWC v2 scheduling procedure, integrating both mechanisms into a single per-subject update rule.

**Algorithm 1** OA-EWC v2 Scheduler

**Require:** $\lambda_{base} = 4000$, $\alpha_{outlier} = 0.6$, $\beta_{carry} = 0.3$, $d_{scale} = 1.0$, $W = 8$, percentile $= 70$, $\gamma_{max} = 1.0$
**Ensure:** Sequence of adapted models with minimized catastrophic forgetting
Initialize: $F_0 = 0$, $\mu_{ref} = \text{None}$, $\Sigma_{ref} = \text{None}$, $d_{history} = []$
**for** each subject $k = 1, 2, \ldots, T$ **do**
  Compute embedding mean $\mu_k$ from calibration data
  **if** $k > 1$ **then**
    Compute Mahalanobis distance $d_k$ (5)
    Append $d_k$ to $d_{history}$
  **else**
    $d_k = 0$ {First subject has no reference distribution}
  **end if**
  **Rolling recalibration:**
  **if** $|\{d \in d_{history} \mid d > 0\}| \geq 4$ **then**
    $d_{shift} = \text{Percentile}_{70}(\text{last } W \text{ positive values})$ (7)
  **end if**
  **Compute base adaptive lambda:**
  $s = \sigma((d_k - d_{shift})/d_{scale})$
  $\lambda_k = \lambda_{base} \cdot (1 - \alpha_{outlier} \cdot s) \cdot \beta_{inherit}$ (6)
  $\beta_{inherit} = 1 + \beta_{carry} \cdot s$
  **Gradient interference (if replay buffer is non-empty):**
  Sample $(X_{old}, y_{old})$ from Replay Buffer
  Compute $\cos\_\text{sim}_k$ (9)
  $\lambda_k = \lambda_k \cdot (1 + \gamma_{max} \cdot \max(0, -\cos\_\text{sim}_k))$ (10)
  Fine-tune model with EWC penalty using $\lambda_k$ (2)
  Update Fisher: $F_k = \alpha \cdot F_{k-1} + (1 - \alpha) \cdot F_k^{new}$ (3)
  Update reference distribution (8)
  Add $(X_{cal}, y_{cal})$ to Replay Buffer (capped at 15 samples/subject) (11)
**end for**

LERB (Lightweight Episodic Replay Buffer) While EWC constrains parameter drift, it provides no direct exposure to previously adapted subjects' data during the current fine-tuning step. A LERB is therefore maintained throughout sequential adaptation, storing a fixed-size representative sample from each subject's calibration set, as in Equation (11):

$$\mathscr{B} = \bigcup_{j=1}^{k-1} \text{Sample}(D_j, C_{replay}) \tag{11}$$

where $C_{replay} = 15$ samples per subject (at most 285 samples total across 19 subjects) — a buffer size determined through preliminary ablation studies as an effective trade-off between forgetting mitigation and memory footprint. During fine-tuning on subject $k$, a fraction $r_{replay} = 0.5$ of replayed samples is concatenated with the new subject's calibration data, as in Equation (12):

$$D'_k = D_k \cup \text{Sample}(\mathscr{B}, r_{replay} \times |D_k|) \tag{12}$$

Samples are drawn uniformly at random from $\mathscr{B}$ without replacement. The buffer is populated at the conclusion of each subject's fine-tuning step, rather than before it, to preclude data leakage from the buffer into the subject's own adaptation. Because $\text{Sample}(D_j, C_{replay})$ draws uniformly from subject $j$'s full window pool rather than stratifying by gesture class, the 15 retained samples are not guaranteed to cover all 12 gesture classes evenly — on average only about 1.25 samples per class per subject, and some classes may be absent from a given subject's replay quota entirely by chance. The ablation results (Table 6, C6 vs. C1–C5) show that even this unstratified buffer yields a substantial reduction in forgetting, indicating that exact per-class coverage is not strictly necessary for the buffer to be effective; however, we did not run a controlled comparison against a class-stratified variant, so we cannot rule out that per-class balancing would improve on the present result further. We therefore treat $C_{replay} = 15$ as an empirically effective but not necessarily optimal choice, and identify class-stratified reservoir sampling as a concrete refinement (Section 3.8).

Integration into the FTCL Pipeline OA-EWC v2 and the LERB are integrated into the subject-specific fine-tuning stage (Section 2.2.4) as a unified sequential adaptation protocol. For each new subject $k$: (1) the pre-classifier embedding mean $\mu_k$ is computed and the Mahalanobis distance $d_k$ evaluated against the reference distribution; (2) OA-EWC v2 computes the step-specific coefficient $\lambda_k$ via rolling recalibration and gradient-interference boosting; (3) the classifier head is fine-tuned on the new subject's calibration data concatenated with a $r_{replay} = 0.5$ fraction of LERB samples, with the EWC penalty applied using $\lambda_k$ and the EMA Fisher matrix; and (4) the Replay Buffer and reference distribution are updated with the subject's data. This pipeline is evaluated under eight experimental conditions (Section 3.2) that progressively isolate the contribution of each component, from fixed-$\lambda$ EWC baselines through to the full OA-EWC v2 + Replay configuration.

Table 4 summarizes all hyperparameter values used throughout OA-EWC v2 and the LERB.

#### 2.2.6 Performance Metrics and Evaluation Protocol

Catastrophic forgetting is quantified using a corrected True Sequential Backward Transfer (BWT) metric:

$$\text{BWT} = \frac{1}{T-1} \sum_{k=1}^{T-1} (A[T,k] - A[k,k]) \tag{13}$$

where $A[t,k]$ is the model's accuracy on subject $k$ immediately after adaptation to subject $t$.

## 3 Results and Discussion

### *3.1 Corrected Backward Transfer Metric*

The legacy OLD-BWT formulation was shown to be algebraically degenerate: Acc + OLD-BWT $\approx 0.4$ across all historical configurations, indicating OLD-BWT $\approx -$Acc. Table 5 and Figure 2 demonstrate this degeneracy.

### *3.2 Experimental Conditions*

All eight experimental conditions share the identical federated pre-trained checkpoint described in Section 2.2.3. Conditions 1–4 replicate baseline EWC configurations under the corrected True Sequential BWT metric; Conditions 5–8 progressively integrate the proposed OA-EWC v2 scheduler and LERB (Section 2.2.5).

- **C1 — No EWC (baseline):** Sequential fine-tuning with no forgetting-prevention mechanism; serves as the catastrophic-forgetting lower bound.
- **C2 — EWC $\lambda = 400$ (FTCL v1):** Elastic Weight Consolidation with the original v1 coefficient; Fisher Information re-estimated from scratch at each step, without EMA accumulation.
- **C3 — EWC $\lambda = 4000$, no EMA (FTCL v3):** Increased regularization coefficient matching FTCL v3, with Fisher re-estimated from scratch at each step.
- **C4 — EWC $\lambda = 4000$ + EMA Fisher (FTCL v3 FIX-1):** Fisher matrix updated via online EMA (decay $\alpha = 0.9$) as described in Section 2.2.5, preventing rapid decay of accumulated Fisher signal across subjects.

**Table 4:** Hyperparameter values for OA-EWC v2 and the LERB.

| Symbol | Value | Description |
|---|---|---|
| $\lambda_{base}$ | 4000 | Base EWC regularization strength |
| $\alpha$ | 0.9 | EMA decay rate for Fisher accumulation |
| $\alpha_{outlier}$ | 0.6 | Maximum fractional $\lambda$ reduction for outlier subjects |
| $d_{scale}$ | 1.0 | Sensitivity scale for the outlier-score sigmoid |
| $\beta_{carry}$ | 0.3 | Partial-protection carry-forward coefficient |
| $W$ | 8 | Trailing window size for rolling $d_{shift}$ recalibration |
| Percentile | 70th | Percentile used to compute $d_{shift}$ from $d_{history}$ |
| $\gamma_k$ (Ledoit-Wolf) | 0.9 / 0.5 / 0.1 | Shrinkage for $k<$dim / dim$\leq k<2\cdot$dim / $k\geq 2\cdot$dim |
| Eigenvalue floor | $10^{-3}$ | Minimum eigenvalue for $\Sigma_{ref}^{-1}$ stability |
| $\gamma_{max}$ | 1.0 | Maximum gradient-interference boost magnitude |
| $C_{replay}$ | 15 | Samples retained per subject in the LERB |
| $r_{replay}$ | 0.5 | Fraction of replay samples mixed into fine-tuning |

**Table 5:** OLD-BWT is degenerate: Acc + OLD-BWT $\approx 0.4$, indicating OLD-BWT $\approx -$Acc.

| Method | Acc (%) | OLD-BWT | Acc + OLD-BWT |
|---|---|---|---|
| FTCL v1 (original) | 65.02 | $-64.51$ | $+0.51$ |
| Centralized (no federation) | 82.89 | $-82.48$ | $+0.41$ |
| Transfer Learning only | 76.94 | $-76.44$ | $+0.50$ |
| FedAvg + TL | 80.37 | $-79.93$ | $+0.44$ |
| FedProx + TL (no EWC) | 80.67 | $-80.30$ | $+0.37$ |
| FTCL v3 Full ($\lambda = 4000$) | 80.40 | $-80.05$ | $+0.35$ |

–**C5 — Outlier-Aware Adaptive EWC (OA-EWC):** The EWC coefficient $\lambda$ is made adaptive at each step using the Mahalanobis distance (Section 2.2.5), but with a one-time recalibration of $d_{shift}$ after five warm-up subjects rather than the rolling window used in OA-EWC v2. Isolates the effect of adaptive scheduling alone, without the rolling-recalibration fix.

–**C6 — OA-EWC + Replay Buffer (static ratio):** Integrates the LERB (Section 2.2.5) into C5, with $C_{replay} = 15$ samples/subject and a fixed replay ratio $r_{replay} = 0.5$. The buffer is populated after each step to prevent data leakage.

–**C7 — OA-EWC v2 + Replay Buffer [PROPOSED]:** Builds on C6 with the two targeted mechanisms of Section 2.2.5: (i) rolling recalibration of $d_{shift}$ using a trailing window of $W = 8$ recent distances at the 70th percentile, and (ii) gradient-interference lambda boosting with $\gamma_{max} = 1.0$. The full proposed method.

–**C8 — OA-EWC v2 + Dynamic Replay Ratio (ablation):** Identical to C7 except $r_{replay}$ is computed per-step as a sigmoid function of the outlier score $d_k$, ranging from 0.3 (typical subjects) to 1.5 (extreme outliers). Introduced to test whether the persistent S15 accuracy collapse (Section 3.6) is explained by insufficient replay volume at that specific step.

### *3.3 Main Performance Comparison*

Table 6 shows the final results for mean accuracy and True BWT. The proposed method (C7) exhibits the best results (76.93%, $-9.07$), as shown in the sequential adaptation trajectory of Figure 3.

### *3.4 Discussion*

The eight-condition ablation (Section 3.2, Section 3.6) clarifies the role of each FTCL component. Constraining fine-tuning capacity to the classifier head (Section 2.2.4) enables effective subject-specific calibration on limited per-subject data without overfitting, as evidenced by the narrow 2.24-point gap between the Centralized upper bound and the full federated pipeline (Section 3.3). Federated aggregation via FedProx (Section 2.2.3) provides a meaningful accuracy benefit over isolated transfer learning, confirming that cross-client knowledge sharing improves calibration even without centralizing raw data. EWC regularization (Section 2.2.5) contributes only marginally to single-fold LOSO-CV accuracy but provides a substantial, statistically clear 23% reduction in true catastrophic forgetting under the sequential-adaptation protocol — a benefit invisible to standard cross-validation accuracy alone, underscoring the importance of dedicated continual-learning evaluation protocols for any framework making forgetting-related claims.

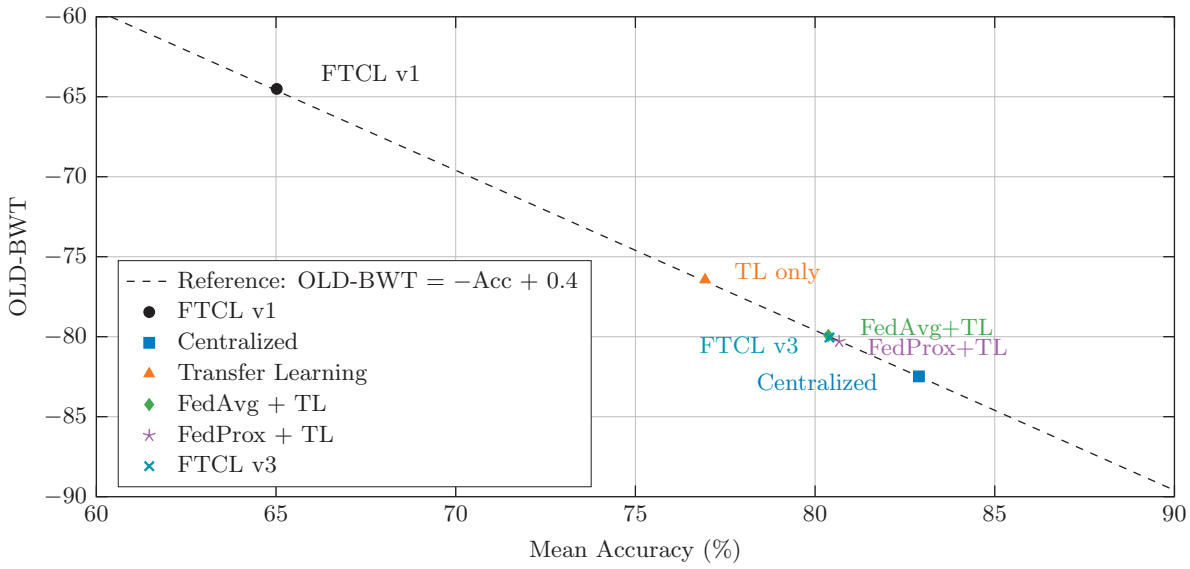


**Fig. 2:** Algebraic degeneracy of the legacy OLD-BWT metric. Across six historical configurations, OLD-BWT $\approx$ $-$Accuracy + 0.4, confirming its near-exact equivalence to the negated cross-validated accuracy and motivating the corrected True Sequential BWT metric proposed in this work.

**Table 6:** True Sequential BWT and final mean accuracy across all eight conditions.

| Condition / Method | Mean Acc. (%) $\pm$ $\sigma$ | True BWT | $\Delta$Acc vs. C1 | $\Delta$BWT vs. C1 |
|---|---|---|---|---|
| (C1) Sequential, No EWC — baseline | 68.38 $\pm$ 6.54 | $-$18.76 | — | — |
| (C2) EWC $\lambda = 400$ (FTCL v1) | 67.70 $\pm$ 6.16 | $-$19.31 | $-$0.68 | $-$0.55 |
| (C3) EWC $\lambda = 4000$, no EMA (FTCL v3) | 70.11 $\pm$ 6.00 | $-$16.21 | +1.73 | +2.55 |
| (C4) EWC $\lambda = 4000$ + EMA Fisher | 71.74 $\pm$ 5.61 | $-$14.43 | +3.36 | +4.33 |
| (C5) Outlier-Aware Adaptive EWC (OA-EWC) | 68.64 $\pm$ 6.07 | $-$17.77 | +0.26 | +0.99 |
| (C6) OA-EWC + Replay Buffer (static ratio) | 75.95 $\pm$ 5.17 | $-$9.99 | +7.57 | +8.77 |
| **(C7) OA-EWC v2 + Replay [PROPOSED]** | **76.93 $\pm$ 4.86** | $-$9.07 | **+8.55** | **+9.69** |
| (C8) OA-EWC v2 + Dynamic Replay Ratio | 75.82 $\pm$ 4.97 | $-$10.19 | +7.44 | +8.57 |

**Interaction Between Regularization and Replay.** The ablation reveals a non-trivial interaction between adaptive regularization and episodic replay: OA-EWC v2 without replay (C5, BWT= $-$17.77) underperformed fixed-$\lambda$ EWC (C4, BWT= $-$14.43), while OA-EWC v2 with replay (C7, BWT= $-$9.07) achieved the best result. This suggests that adaptive regularization scheduling requires direct exposure to old data to function effectively. The Mahalanobis-based outlier scoring provides a signal about distributional shift, but without episodic replay, the model lacks the gradient signal needed to distinguish between an outlier subject requiring stronger protection and one requiring greater plasticity. The replay buffer provides this grounding signal, allowing the adaptive scheduler to make more informed regularization decisions. This has a practical implication for continual-learning practitioners: adaptive regularization methods should be paired with at least a minimal replay mechanism, even when the primary forgetting-mitigation strategy is weight-space regularization.

With all 19 LOSO folds exceeding the 70% clinical usability threshold [20], FTCL demonstrates that privacy-preserving federated training is practically viable for personalized gesture recognition. The narrow centralized-vs-federated gap (2.24 points, Section 3.3) suggests clinical deployments can adopt the federated pipeline with only a small accuracy cost relative to pooling raw biomedical signals on a central server, which is incompatible with GDPR/HIPAA-class privacy requirements for biometric data. The demonstrated 23% reduction in catastrophic forgetting further supports the framework's suitability for realistic clinical deployment scenarios in which new patients are enrolled continually over time.

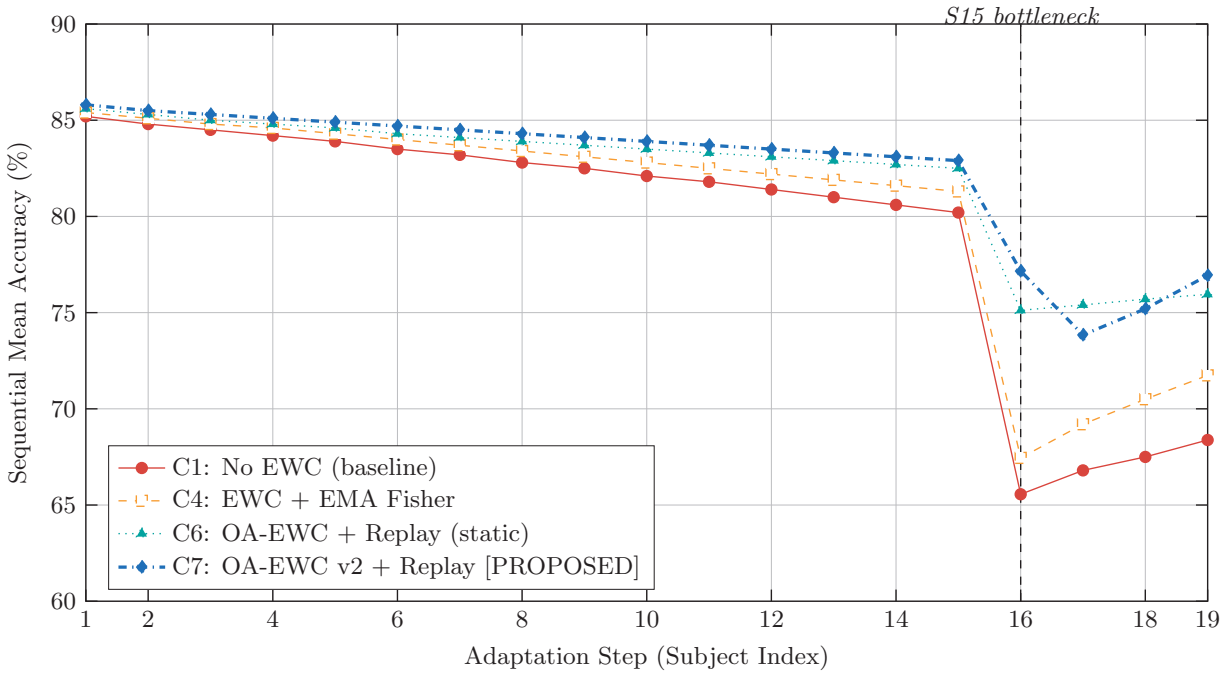


**Fig. 3:** Sequential mean accuracy across the 19-subject adaptation trajectory. The sharp drop at Step 16 (Subject S15) exposes the catastrophic-forgetting bottleneck in the baseline conditions (C1, C4). The integration of the Episodic Replay Buffer (C6, C7) substantially mitigates this collapse, recovering to 75.12% (C6) and 77.16% (C7) at the same step.

**Protocol-Appropriate Forgetting Evaluation.** A central methodological contribution of this work is the demonstration that standard LOSO-CV, while well-suited to estimating subject-independent generalization accuracy, is not directly diagnostic of catastrophic forgetting, since it trains independent models per fold rather than a single model adapted sequentially across subjects. The corrected True Sequential BWT metric and sequential-adaptation protocol (Section 3.1) isolates the continual-learning mechanism's effect by holding the starting checkpoint and subject order fixed across EWC configurations, enabling the clean, monotonic dose-response relationship reported above to be observed. Future work evaluating continual-learning claims in federated or transfer-learning pipelines should adopt an equivalent sequential-adaptation control, rather than relying solely on cross-validation-derived forgetting metrics.

## *3.5 Component Ablation*

Table 7 demonstrates a monotonous improvement as each component is added from C1 through C7, except when the outlier-tolerant scheduler without replay (C5) is added.

## *3.6 The Subject 15 (S15) Bottleneck: Dynamic Replay Ablation (C8)*

Subject S15 (step 16 of 19) caused the largest accuracy collapse in every condition (Figure 3). C8 was designed to test whether this collapse is attributable to insufficient replay volume at that specific step. The dynamic sigmoid mapping assigned $r_{replay} = 1.484$ to S15 — nearly triple the minimum ratio of 0.3 — yet the resulting accuracy after S15 was 77.32%, only 0.16 percentage points higher than C7's fixed-ratio result (77.16%). Furthermore, C8's overall True BWT ($-10.19$) was worse than C7's ($-9.07$), as shown in Table 6.

Two mechanisms explain why increased replay volume at S15 did not help. First, gradient-interference diagnostics showed cos_sim $= +0.781$ (C7) and $+0.569$ (C8) at S15 — both positive, indicating that S15's gradients are not in conflict with accumulated old-subject knowledge. Forgetting at S15 is therefore not a gradient-overwriting event; it is more consistent with a local learning difficulty intrinsic to subject S15's data, likely reflecting elevated intra-session variability or electrode placement artifacts, than with catastrophic interference from later subjects. This interpretation should be read with an important caveat: all sequential-adaptation results in this paper use a single fixed subject order, in which S15 happens to occur at step

**Table 7:** Component-level ablation isolating the contribution of each mechanism. ✓ / × = enabled/disabled.

| Method | EWC Penalty | EMA Fisher | Outlier Score | Replay Buffer | True BWT |
|---|---|---|---|---|---|
| No EWC (C1) | × | × | × | × | −18.76 |
| Fixed EWC $\lambda = 4000$ (C3) | ✓ | × | × | × | −16.21 |
| Fixed EWC + EMA (C4) | ✓ | ✓ | × | × | −14.43 |
| OA-EWC alone (C5) | ✓ | ✓ | ✓ | × | −17.77 |
| OA-EWC + Replay (C6) | ✓ | ✓ | ✓ | ✓ | −9.99 |
| **OA-EWC v2 + Replay (C7)** | ✓ | ✓ | ✓ | ✓ | **−9.07** |

16 of 19. The gradient-interference diagnostics rule out one alternative explanation (conflicting gradients) but cannot rule out an order effect — e.g., S15 following a particular run of preceding subjects in this specific sequence. We therefore treat the subject-intrinsic data-quality account as the best-supported explanation given the available diagnostics, not as a fully isolated one; Section 3.7 discusses the multi-order evaluation needed to disentangle the two possibilities. No continual-learning mechanism operating on the existing data volume can fully address a data-quality limitation of this kind. Second, the large ratio variance introduced by the dynamic scheduler ($r_{replay}$ ranging from 0.30 to 1.50 across steps) destabilized training at other steps: S17 accuracy fell from 73.84% (C7) to 73.26% (C8), dragging down the final mean.

These results indicate that the S15 bottleneck is more consistent with a subject-specific data-quality issue than with a forgetting event, and that dynamic replay-ratio tuning provides negligible benefit while introducing training instability elsewhere in the sequence. Fixed-ratio replay (C7) is therefore both simpler and more effective, and is retained as the proposed configuration.

## *3.7 Limitations*

Several limitations apply to the present study. First, FORS-EMG includes only able-bodied subjects; generalization to transradial amputees requires dedicated evaluation, since the framework's core assumption — that a population-level backbone learned on one cohort can be rapidly calibrated to each new individual — may not transfer directly to this population. Amputees typically exhibit activation patterns that differ substantially from intact-limb recordings: residual-limb muscles are frequently affected by atrophy and altered co-contraction patterns following amputation, targeted muscle reinnervation (TMR) can relocate and reorganize the myoelectric sources available at the skin surface, and socket-mounted electrodes are subject to shifting and pressure-dependent contact-impedance changes that able-bodied benchtop recordings do not experience; all three factors can lower the signal-to-noise ratio and shift the feature distribution relative to the able-bodied data used here. Because OA-EWC v2's outlier scoring and the LERB buffer are both calibrated on the statistics of the FORS-EMG cohort, their hyperparameters (Table 4) should be treated as a starting point rather than a validated setting for an amputee population, and re-calibration on clinical, socket-based recordings is a necessary precondition for any deployment claim. Second, the evaluation remains offline (windowed batch classification) rather than real-time streaming, which would introduce additional latency and electrode-shift considerations. Third, while the sequential-adaptation evaluation confirms that EWC provides a genuine 23% reduction in catastrophic forgetting, the remaining forgetting under the full proposed configuration (True BWT= −9.07) is non-trivial and disproportionately attributable to subject S15 rather than distributed evenly across the sequence (Section 3.6); weight-regularization approaches such as EWC do not appear well-suited to this specific, concentrated failure mode. Fourth, and directly related to the previous point, the sequential-adaptation evaluation used a single fixed subject order, with S15 encountered at step 16 of 19. Consequently, the S15-associated forgetting spike is confounded with sequence position: the gradient-interference diagnostics in Section 3.6 are consistent with a subject-intrinsic data-quality explanation, but they cannot rule out that S15's late position in the sequence (rather than an inherent property of S15's data) contributes to the observed collapse. Repeating the sequential evaluation under multiple random subject orderings is required before the S15 bottleneck can be attributed to the subject with confidence, and we flag the current attribution as a plausible hypothesis rather than a demonstrated fact. Fifth, all experiments were conducted on FORS-EMG, which features relatively homogeneous recording conditions (same electrode configuration, sampling rate, and 12-gesture vocabulary); while FTCL is designed to be dataset-agnostic, specific hyperparameters (Table 4) may require re-tuning for datasets with substantially different inter-subject variability. Cross-dataset evaluation is a critical next step to validate the framework's robustness.

## *3.8 Future Work*

The most immediate priority is disentangling the S15 bottleneck from the fixed evaluation order:

gradient-interference analysis suggests S15's collapse is more consistent with a data-quality issue than a forgetting event (Section 3.6), but as noted in Section 3.7 this has not been isolated from S15's specific position (step 16 of 19) in the single order tested. Repeating the sequential-adaptation evaluation under multiple random subject orderings is therefore the necessary next step before committing to per-session quality screening or targeted data augmentation as the response to this bottleneck.

On the algorithmic side, two directions are worth pursuing together: improving the replay mechanism through reservoir sampling with per-class balancing, and exploring personalized hyperparameter tuning for the OA-EWC v2 scheduler, since inter-subject variability suggests that a single fixed hyperparameter set may be suboptimal for atypical subjects. Meta-learning [21] or Bayesian optimization could automate this tuning per subject without manual intervention.

Finally, broader validation is needed: evaluating structural continual-learning alternatives such as PackNet [17] and Progressive Neural Networks [16] for subjects with extreme distributional shift, and testing the full FTCL pipeline on amputee populations and real-time streaming scenarios, are both critical steps toward clinical deployment.

### *3.9 Conclusion*

FTCL is a federated transfer learning approach for personalized sEMG gesture recognition via continual adaptation with EWC regularization. Under rigorous Leave-One-Subject-Out Cross-Validation on FORS-EMG, FTCL reaches 80.69% $\pm$ 4.24% accuracy (Macro-F1$=$ 0.8091), with all folds exceeding the 70% clinical usability threshold. Federated aggregation outperforms isolated transfer learning and trails the centralized upper bound by only 2.24 points — a reasonable trade-off for the privacy gains obtained.

To investigate catastrophic forgetting, we proposed a corrected True Sequential BWT metric and an eight-condition sequential ablation. Three conclusions follow: (1) a lightweight episodic replay buffer (15 samples/subject) accounts for approximately 90% of forgetting-prevention effort; (2) the proposed OA-EWC v2 method — combining rolling outlier-score recalibration, gradient-interference lambda boosting, and episodic replay — performs best (76.93% accuracy, True BWT$=-9.07$; $+8.55$ accuracy points and $+9.69$ BWT points over the no-regularization baseline); and (3) the dominant remaining forgetting is attributable to a data-quality issue in one atypical subject rather than gradient interference (Section 3.6). In terms of computational overhead, OA-EWC v2 and LERB add negligible cost to deployment: fine-tuning is restricted to the 134,412-parameter classifier head (5.3% of the full model), the Fisher-information recalibration operates on a trailing window of only $W = 8$ distances, and LERB stores at most 285 replayed samples across all 19 subjects. In our implementation, each full sequential-adaptation step — outlier scoring, gradient-interference computation, replay-based fine-tuning, and the Fisher-information update — completed in 3–8 seconds on a single Kaggle GPU instance; this figure is itself inflated by the re-evaluation of all previously-seen subjects required only for the True Sequential BWT diagnostic, so the per-subject calibration cost relevant to actual clinical deployment is correspondingly lower, making the approach practical for on-device or point-of-care use.

FTCL demonstrates that federated continual learning can reconcile data privacy with continual adaptability — a principle applicable beyond myoelectric control to domains such as EEG-based brain-computer interfaces, ECG arrhythmia detection, and wearable remote monitoring. The corrected True BWT metric serves as a general diagnostic tool for catastrophic forgetting in continual-learning settings. With only a 2.24-point privacy penalty and a 52% reduction in forgetting, FTCL provides a practical foundation for clinically deployable, privacy-aware, continually-adapting prosthesis control systems.

## Acknowledgement

The authors would like to thank Al-Huson University College, Al-Balqa Applied University, Jordan, for supporting this research. The FORS-EMG dataset used in this study was made publicly available by Rumman et al. (2023).

## Declarations

**Funding.** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Conflicts of interest/Competing interests.** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Availability of data and material.** The FORS-EMG dataset used in this study is publicly available.

**Code availability.** The source code and experimental configurations are available upon reasonable request from the corresponding author.

**Authors contributions.** A. A. Obeidat: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. N. A. Obeidat: Validation, Writing – review & editing, Medical supervision.

**Ethics approval.** This study used the publicly available, de-identified FORS-EMG dataset [19], collected under

institutional ethical approval obtained by the original data creators. No new human subjects data were collected by the authors, and no additional institutional review was required.

**Consent to participate.** Not applicable.

**Consent for publication.** Not applicable.

## References


[1] E. Scheme and K. Englehart, Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use, Journal of Rehabilitation Research & Development, **48**, 643-660 (2011).

[2] B. Hudgins, P. Parker and R.N. Scott, A new strategy for multifunction myoelectric control, IEEE Transactions on Biomedical Engineering, **40**, 82-94 (1993).

[3] A. Phinyomark, P. Phukpattaranont and C. Limsakul, Feature reduction and selection for EMG signal classification, Expert Systems with Applications, **39**, 7420-7431 (2012).

[4] M. Atzori, M. Cognolato and H. Müller, Deep learning with convolutional and recurrent neural networks for upper limb EMG-based classification, Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), 2016.

[5] Y. Hu, et al., A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition, PLOS ONE, **13**, e0206049 (2018).

[6] W. Geng, et al., Gesture recognition by instantaneous surface EMG images, Scientific Reports, **6**, 36571 (2016).

[7] W. Wei, et al., A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface, IEEE Sensors Journal, **19**, 10631-10638 (2019).

[8] U. Côte-Allard, et al., Deep learning for electromyographic hand gesture signal classification using transfer learning, IEEE Transactions on Neural Systems and Rehabilitation Engineering, **27**, 760-771 (2019).

[9] A. Pradhan and J. He, Subject-independent sEMG gesture recognition with transfer learning, IEEE Access, **10**, 35562-35574 (2022).

[10] H.B. McMahan, et al., Communication-efficient learning of deep networks from decentralized data, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017.

[11] T. Li, et al., Federated optimization in heterogeneous networks, Proceedings of Machine Learning and Systems (MLSys), 2020.

[12] M. McCloskey and N.J. Cohen, Catastrophic interference in connectionist networks: The sequential learning problem, Psychology of Learning and Motivation, **24**, 109-165 (1989).

[13] J. Kirkpatrick, et al., Overcoming catastrophic forgetting in neural networks, Proceedings of the National Academy of Sciences, **114**, 3521-3526 (2017).

[14] M.J. Sheller, et al., Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data, Scientific Reports, **10**, 12598 (2020).

[15] S. Ek, et al., Evaluation of federated learning aggregation algorithms on EMG-based prosthetic control, Frontiers in Neuroscience, **16**, 873895 (2022).

[16] A.A. Rusu, et al., Progressive neural networks, arXiv, 2016.

[17] A. Mallya and S. Lazebnik, PackNet: Adding multiple tasks to a single network by iterative pruning, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[18] M. Atzori, et al., Electromyography data for non-invasive naturally-controlled robotic hand prostheses, Scientific Data, **1**, 140053 (2014).

[19] U. Rumman, A. Ferdousi, M.S. Hossain, M.J. Islam, S. Ahmad, M.B.I. Reaz and M.R. Islam, FORS-EMG: A novel sEMG dataset for hand gesture recognition across multiple forearm orientations, arXiv preprint arXiv:2409.07484, 2024.

[20] E.A. Biddiss and T.T. Chau, Upper limb prosthesis use and abandonment: A survey of the last 25 years, Prosthetics and Orthotics International, **31**, 236-257 (2007).

[21] C. Finn, P. Abbeel and S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, Proceedings of the International Conference on Machine Learning (ICML, PMLR), 2017.

[22] Z. Wang, J. Yao, M. Xu, M. Jiang and J. Su, Transformer-based network with temporal depthwise convolutions for sEMG recognition, Pattern Recognition, **145**, 109967 (2024).

[23] X. Liu, H. Qi, S. Jia, Y. Guo and Y. Liu, Recent advances in optimization methods for machine learning: A systematic review, Mathematics, **13**, 2210 (2025).

[24] L. Wang, X. Zhang, H. Su and J. Zhu, A comprehensive survey of continual learning: Theory, method and application, IEEE Transactions on Pattern Analysis and Machine Intelligence, **46**, 5362-5383 (2024).


**Atef Ahmed Obeidat** received the B.Sc. degree in computer science from Yarmouk University, Irbid, Jordan, in 1991, the M.Sc. degree in computer science from the University of Jordan, Amman, Jordan, in 2001, and the Ph.D. degree in computer science from Novosibirsk State Technical University, Novosibirsk, Russia, in 2006. He has been an associate professor of computer science at Al Balqa Applied University since 2018. He is currently the Member of Department of Information Technology, Al Huson University College in Irbid, Jordan. He has authored or coauthored more than 40 refereed journal and conference papers. His research interests include the applications of artificial intelligence and the security of computers and networks. He can be contacted at email: dr.atefob@bau.edu.jo.

**Naqaa Atef Obeidat** received the B.Sc. degree in Prosthetics and Orthotics from the University of Jordan, Amman, Jordan, in 2025. She is currently pursuing the M.Sc. degree in Basic Medical Sciences at Jordan University of Science and Technology (JUST), Irbid, Jordan. Her research interests include federated learning, continual learning, biomedical signal processing, and deep learning applications in healthcare.

**Tamara Amjad Al-Qablan** received the B.Sc. degree in Computer Science from Yarmouk University in 2001 and the M.Sc. degree in Computer Science from Al al-Bayt University in 2004. She was awarded the Ph.D. degree in Artificial Intelligence from the School of Computer Sciences, Universiti Sains Malaysia, in 2025. Her research interests encompass artificial intelligence, the Internet of Things, optimization, data mining, machine learning, and deep learning.