Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Critical Review of Adaptive Normalization Techniques for Cross-Session EEG Decoding on Embedded Platforms

Cheng Lee

leecheng4521@gmail.com

Shanghai Jiao Tong University

Systematic Review

Keywords: EEG decoding, cross-session non-stationarity, adaptive normalization, domain adaptation, embedded systems, TinyML, brain-computer interfaces, feature alignment

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11740307/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Cross-session non-stationarity remains a primary obstacle to the practical deployment of electroencephalography (EEG) decoding systems, as feature distributions drift substantially across recording sessions due to electrode repositioning, impedance fluctuations, and cognitive state variability. Adaptive normalization techniques have emerged as a computationally efficient countermeasure, aligning feature statistics in an online or calibration-light manner without requiring extensive labeled target data. However, the suitability of these methods for resource-constrained embedded platforms—where memory, latency, and energy budgets are severely limited—has not been systematically examined. This review critically evaluates adaptive normalization approaches for cross-session EEG decoding through an embedded-deployment lens. We categorize existing techniques into three families: exponential moving average-based feature normalization, adaptive instance normalization, and Riemannian or covariance-domain alignment. For each, we analyze computational complexity, memory footprint, sensitivity to hyperparameters, and compatibility with fixed-point or quantized inference. We further examine the trade-off between normalization efficacy and the overhead incurred by online statistic estimation, arguing that several widely adopted methods assume computational resources incompatible with microcontroller-class hardware. Our analysis identifies a gap between algorithmic sophistication and deployability, and we propose evaluation criteria that prioritize streaming operation, bounded memory, and robustness to distribution shift under quantization. We conclude that lightweight first-order normalization schemes currently offer the most favorable accuracy-to-cost ratio for embedded cross-session decoding, while geometry-aware and deep adaptation methods require architectural simplification before they can be considered viable on constrained platforms.

1. Introduction

The translation of electroencephalography (EEG) decoding from controlled laboratory demonstrations to deployable, everyday neurotechnology remains obstructed by a persistent and deceptively mundane problem: the signal changes. Across recording sessions, the statistical properties of EEG shift in ways that degrade the performance of decoders trained on prior data, a phenomenon broadly termed cross-session non-stationarity. This instability arises from a confluence of physiological and instrumental factors. Electrode impedance drifts as gel dehydrates or dry contacts settle, cap placement varies by millimeters between visits, subject arousal and fatigue modulate cortical rhythms, and even the time of day can influence spectral content (Wan et al., 2021; Apicella et al., 2024). The consequence is that a classifier achieving high accuracy in a single session may perform near chance when applied to a subsequent recording from the same individual performing the same task. For brain-computer interface (BCI) applications intended for longitudinal use—whether communication support for locked-in patients, motor rehabilitation after stroke, or passive monitoring of cognitive load—this cross-session degradation constitutes a fundamental barrier to clinical and consumer viability.

Adaptive normalization techniques have emerged as a computationally attractive response to this problem. Rather than requiring the collection of extensive labeled data in each new session, adaptive

normalization methods seek to align feature distributions by estimating and correcting for shifts in first- and second-order statistics. The underlying intuition is straightforward: if session-to-session drift primarily manifests as affine transformations in feature space—shifts in mean and scaling in variance—then normalizing feature statistics in an online or calibration-light manner may restore decoder function without demanding the user endure lengthy recalibration protocols (Li et al., 2024; Zhang et al., 2026). Methods in this family range from simple exponential moving average (EMA) normalization of feature activations to adaptive instance normalization (AdaIN) layers that align channel-wise statistics, to Riemannian approaches that operate on the geometry of covariance matrices. Each offers a different trade-off between alignment efficacy, computational cost, and assumptions about the nature of the distribution shift.

Yet the question of whether these techniques are suitable for embedded platforms—the microcontrollers and low-power digital signal processors that would populate wearable or implantable BCIs—has received strikingly little systematic attention. The literature on adaptive normalization for EEG is dominated by studies that evaluate performance on desktop GPUs with float32 precision, ample memory, and no strict latency constraints. A method that improves cross-session accuracy by several percentage points may nonetheless be undeployable if it requires iterative eigendecomposition of covariance matrices, storage of large support sets, or backward passes through deep networks during inference. Embedded platforms impose hard constraints on memory (often measured in kilobytes, not megabytes), computational throughput (millions of operations per second, not billions), and energy (milliwatt budgets that must support hours or days of operation). The gap between algorithmic sophistication and deployability is not merely an engineering afterthought; it determines whether a technique can exit the laboratory at all.

This review takes as its organizing premise that adaptive normalization techniques must be evaluated not only on their accuracy under cross-session conditions but on their compatibility with the resource envelopes of embedded hardware. We proceed from an analysis of the sources and characteristics of cross-session non-stationarity, through a critical examination of the major families of adaptive normalization methods, to an assessment of their computational and memory footprints, and finally to a set of criteria for evaluating deployability. The review is critical in the specific sense that it does not treat all normalization approaches as equally viable for embedded use, nor does it assume that improvements reported in GPU-based studies will survive quantization, fixed-point arithmetic, or streaming inference with limited buffering.

The scope of the review is deliberately bounded. We focus on adaptive normalization techniques—methods that modify feature statistics during training or inference to compensate for distribution shift—rather than on the broader landscape of domain adaptation or transfer learning. Adversarial domain adaptation, for instance, learns domain-invariant representations through discriminator networks (Zhao et al., 2025), but its training complexity and inference overhead place it largely outside the embedded regime considered here. Similarly, we do not address artifact rejection or preprocessing pipelines except insofar as they interact with normalization. The review concentrates on cross-session rather than cross-subject variability, though the two are related and the boundary between them is often blurry in practice.

Motor imagery and related sensorimotor paradigms receive the most attention because they dominate the cross-session decoding literature, but emotion recognition and biometric applications are discussed where relevant.

The contribution of this review is threefold. First, it synthesizes the literature on adaptive normalization for cross-session EEG decoding and organizes it into a taxonomy based on computational structure rather than application domain. Second, it introduces an embedded-deployment lens into the evaluation of these techniques, analyzing memory, latency, and quantization sensitivity alongside decoding accuracy. Third, it identifies a systematic mismatch between the assumptions embedded in many proposed methods and the realities of microcontroller-class hardware, and it proposes evaluation criteria that prioritize streaming operation, bounded memory, and numerical robustness. The review concludes that lightweight first-order normalization schemes currently offer the most favorable accuracy-to-cost ratio for embedded cross-session decoding, while geometry-aware and deep adaptation methods require substantial architectural simplification before they can be considered viable on constrained platforms.

The remainder of the paper is organized as follows. Section 2 reviews the literature on cross-session non-stationarity in EEG and the normalization methods that have been proposed to address it, with attention to their computational characteristics. Section 3 describes the methodology of the review, including the search strategy, inclusion criteria, and the framework used for analyzing embedded deployability. Section 4 presents the results of the review, organized by normalization family, with tables summarizing reported performance, computational requirements, and deployment readiness. Section 5 discusses the findings, emphasizing the trade-offs between accuracy and deployability and identifying open challenges. Section 6 concludes with a summary of the main findings and recommendations for future work.

2. Literature Review

2.1 The Nature of Cross-Session Non-Stationarity in EEG

The non-stationarity of EEG is not a single phenomenon but a family of related processes operating at different timescales and with different underlying mechanisms. At the shortest timescales, spanning seconds to minutes, fluctuations in attention and arousal modulate the amplitude and coherence of cortical rhythms, particularly in the alpha and beta bands (Apicella et al., 2024). These short-term variations are often accommodated by within-session adaptation or by classifiers robust to moderate variance. At intermediate timescales, spanning hours to days, electrode placement variability and impedance changes become prominent. A cap repositioned by a few millimeters alters the spatial sampling of cortical sources, shifting the topographies that spatial filters such as Common Spatial Patterns (CSP) rely upon. Gel electrodes dehydrate, increasing impedance and altering the signal-to-noise ratio in frequency-dependent ways. At the longest timescales, spanning weeks to months, plastic

changes in cortical organization—whether from learning, disease progression, or recovery—may alter the neural generators themselves.

The practical consequence of these layered processes is that feature distributions shift across sessions in ways that are difficult to predict a priori. For motor imagery decoding, the power spectral density in sensorimotor channels during imagined movement may vary in absolute magnitude across sessions even when the spatial pattern of event-related desynchronization remains consistent. A linear classifier trained on session one may find that its decision boundary, calibrated to the mean and covariance of session one features, no longer separates classes in session two. This is not merely a problem of classifier bias; the discriminative structure itself may be partially obscured by the shift.

Several studies have attempted to characterize the magnitude and structure of cross-session drift. In a longitudinal study of a single stroke patient across 134 sessions spanning nine months, the authors of a supervised autoencoder approach observed that session-to-session variability in motor imagery features was substantial and that simple z-score normalization was insufficient to align distributions (Kwon et al., 2025). The drift was not monotonic; some sessions were closer to the training distribution than others, suggesting that factors such as sleep quality or time since previous session contributed variance. This irregularity complicates the design of adaptation methods: a technique that assumes gradual drift may perform poorly when drift is episodic.

The structure of the shift matters for normalization design. If session effects are well-approximated as affine transformations of features—a shift in mean and a scaling in variance—then methods that estimate and invert these transformations should suffice. Evidence for the affine hypothesis comes from the success of relatively simple normalization schemes in improving cross-session performance (Zhang et al., 2026). However, if the shift involves rotation or more complex deformation of the feature distribution, then first-order methods will leave residual misalignment. Riemannian methods, which operate on the manifold of symmetric positive definite covariance matrices, can in principle accommodate a broader class of transformations, but at higher computational cost (Barachant et al., 2012). The empirical question of how much of cross-session drift is affine versus non-affine remains unresolved and likely varies by paradigm and recording setup.

An additional complication is that cross-session drift interacts with cross-subject variability. The features that are stable within a subject across sessions may be precisely those that differ most between subjects, and vice versa. A normalization method that aligns session statistics for a given subject may inadvertently remove subject-specific information that was useful for classification. This tension between invariance and discriminability is central to the design of adaptive normalization and is discussed further in the context of specific methods.

2.2 Normalization in Neural Networks: From Batch Normalization to Adaptive Schemes

The modern vocabulary of normalization in deep learning originates with batch normalization (BN), introduced to stabilize and accelerate training by normalizing activations across the batch dimension (Ioffe and Szegedy, 2015). BN computes the mean and variance of each feature channel across all examples in a mini-batch during training and maintains running estimates for use at inference. Its success in accelerating convergence and improving generalization in convolutional networks established normalization as a standard architectural component. However, BN has properties that are problematic for EEG applications. The dependence on batch statistics creates a train-test discrepancy when batch sizes are small, a common situation in EEG where trials are limited and online inference may involve single samples. More fundamentally, BN normalizes across the batch, which means that the normalization applied to a given sample depends on the other samples in the batch—an undesirable property when the goal is per-sample adaptation to session drift.

Instance normalization (IN), developed initially for style transfer, normalizes each sample independently by computing statistics across spatial dimensions for each channel (Ulyanov et al., 2016). This removes the batch dependence and makes the normalization a deterministic function of the individual input. For EEG, where each trial or time window is a sample that may come from a different session or state, the per-sample nature of IN is conceptually appealing: it aligns the statistics of each sample to a canonical form without reference to other samples. However, IN also removes information that may be discriminative. If the overall amplitude of a feature is informative—for instance, if higher mu rhythm power indicates a different cognitive state—then normalizing it away is counterproductive. The optimal degree of normalization is therefore task-dependent.

Adaptive instance normalization (AdaIN) extends IN by allowing the normalization statistics to be modulated by a conditioning input (Huang and Belongie, 2017). In the style transfer context, AdaIN aligns the channel-wise mean and variance of a content image to those of a style image. For EEG, the analogous operation would align the statistics of a target session to a reference session or to a canonical distribution. The adaptive element can take several forms: the statistics themselves may be estimated from a small amount of target data, or they may be predicted by a network conditioned on some session descriptor. In practice, most EEG applications of AdaIN estimate statistics from the target data in an unsupervised or semi-supervised manner, using either the test session data itself or a small calibration set.

The appeal of AdaIN for embedded deployment lies in its computational simplicity. Once the target statistics are estimated, the normalization operation reduces to an affine transformation: subtract the mean, divide by the standard deviation, multiply by a learned scale, and add a learned bias. These are element-wise operations that map efficiently to fixed-point arithmetic and require only the storage of per-channel statistics and affine parameters. The estimation of statistics, however, introduces complexity. Computing the mean and variance of a feature channel requires either buffering a sufficient number of samples or maintaining running estimates via exponential moving averages. The former imposes memory costs; the latter introduces temporal dynamics that must be tuned.

2.3 Exponential Moving Average Normalization and Online Adaptation

Exponential moving average (EMA) normalization represents the most computationally frugal approach to adaptive feature alignment. The method maintains running estimates of feature mean and variance, updated at each time step as a weighted combination of the previous estimate and the current observation. The update rule is: $\mu_t = \alpha\, \mu_{t-1} + (1-\alpha)\, x_t$, where α is a smoothing coefficient close to one. The normalized feature is then $(x_t - \mu_t) / \sqrt{\sigma_t^2 + \varepsilon}$, optionally followed by a learned affine transformation. The memory cost is O(d) for d-dimensional features, and the computational cost per sample is O(d), making this approach attractive for microcontrollers with limited RAM and no floating-point unit.

EMA normalization has been used implicitly or explicitly in several EEG decoding systems. The adaptive instance normalization layer described in a recent burst-centric motor decoding study updates statistics using summary statistics from test sessions, with the option of few-shot recalibration using a small number of labeled trials (Zhang et al., 2026). The authors report that this approach improves cross-session accuracy relative to static normalization, though the absolute gains depend on the number of calibration trials available. In the context of embedded deployment, the EMA variant with no calibration trials is the most attractive, as it requires no user intervention and can operate fully unsupervised.

The critical parameter in EMA normalization is the smoothing coefficient α, which determines the effective window over which statistics are averaged. A small α (fast adaptation) allows the method to track rapid shifts but may overfit to noise and transient fluctuations. A large α (slow adaptation) provides stable estimates but may lag behind genuine session changes. In a streaming embedded context, the choice of α must balance responsiveness against stability, and it may need to be tuned per subject or per paradigm. There is little systematic guidance in the literature on how to set α for EEG applications, and most studies report values chosen by validation on held-out data—a luxury unavailable in the embedded deployment scenario where labeled target data may be scarce or absent.

A related approach is to normalize using statistics computed over a fixed window of recent samples rather than an exponential average. This introduces an explicit memory buffer of size $W \times d$, where W is the window length. For modest feature dimensions and window lengths of a few seconds, this may be feasible on higher-end microcontrollers but is likely to strain the resources of the smallest devices. The trade-off between EMA and windowed statistics mirrors the broader tension in embedded signal processing between memory and temporal resolution.

2.4 Adaptive Instance Normalization and Affine Alignment

AdaIN in its EEG-specific form typically involves estimating per-channel mean and variance from a target session and applying these statistics to normalize features before classification. The method assumes that the discriminative information is carried by the shape of the feature distribution rather than its absolute location and scale. This assumption is plausible for features such as band power ratios or

normalized spectral amplitudes, which are already partially invariant to overall amplitude. It is less plausible for features where absolute magnitude carries information, such as raw covariance entries used in Riemannian classification.

The empirical evidence for AdaIN in cross-session EEG decoding is mixed but generally positive. In a study comparing AdaIN to Riemannian alignment and adversarial domain adaptation for single-channel motor decoding, the authors argued that AdaIN was preferable for their setting because Riemannian methods lose their advantage when covariance structure collapses to scalar variance, and adversarial methods introduce training instability and complexity incompatible with lightweight deployment (Zhang et al., 2026). The sensitivity analysis in that study indicated that decoding performance varied by less than 2.8% across a reasonable range of threshold parameters, suggesting that the method is not overly fragile to hyperparameter choice. However, the study was conducted on a single channel and a specific paradigm; generalization to multi-channel, multi-paradigm settings remains to be established.

A more elaborate normalization strategy is the hierarchical approach adopted in DRFNet, which applies normalization at multiple levels—feature map, batch, and sample—with adaptive attention to balance discriminability and invariance (Li et al., 2026). This method achieved strong cross-session identification accuracies on SSVEP datasets, but the multi-level architecture and attention mechanisms add computational overhead that may be prohibitive for embedded deployment. The attention weights must be computed at inference time, and the multiple normalization branches require additional memory for intermediate activations. The authors note that filter bank processing and a reduced electrode set can provide a path toward system lightweighting, but the normalization architecture itself is not designed with microcontroller constraints in mind.

The distinction between AdaIN and hierarchical or attention-based normalization is important for embedded deployment. AdaIN with fixed, estimated statistics reduces to a single affine operation per feature channel. Attention-based normalization requires computing the attention weights, which may involve fully connected layers or convolutions, and then applying a weighted combination of normalization schemes. The former is a candidate for fixed-point implementation; the latter is likely to require floating-point arithmetic and careful numerical management.

2.5 Riemannian and Covariance-Domain Alignment

Riemannian methods approach cross-session adaptation from a geometric perspective. Rather than normalizing feature vectors in Euclidean space, they operate on the manifold of symmetric positive definite (SPD) matrices, typically covariance matrices computed from multi-channel EEG epochs. The affine-invariant Riemannian metric provides a distance measure that is invariant to invertible linear transformations, which means that if session effects can be modeled as a linear mixing of sources, Riemannian alignment can in principle remove them. The reference alignment approach, which re-centers covariance matrices to the identity matrix using a reference set, has been shown to improve cross-session and cross-subject classification in several studies (Barachant et al., 2012; Zanini et al., 2018).

The computational demands of Riemannian methods are substantially higher than those of feature-space normalization. Computing a covariance matrix from an epoch of EEG data requires $O(C^2T)$ operations for C channels and T time samples. The eigendecomposition or matrix square root needed for whitening requires $O(C^3)$ operations, which becomes significant for 32 or 64 channels. The storage of covariance matrices is $O(C^2)$ per sample, and the computation of Riemannian distances or tangent space projections adds further cost. On a desktop CPU, these operations are manageable. On a microcontroller without a hardware floating-point unit, they are likely to be prohibitive for real-time operation.

There are approximations that reduce the cost. Diagonal loading of covariance matrices simplifies the whitening operation. Restricting to a subset of channels reduces C. Fixed-point implementations of matrix operations are possible but require careful scaling to avoid overflow and underflow. However, each of these approximations erodes the theoretical advantages of the Riemannian framework. If the covariance structure is simplified to the point where the Riemannian geometry becomes Euclidean in disguise, then the method loses its distinctive justification. The literature on embedded Riemannian EEG decoding is sparse, and most studies that report cross-session improvements do not report the computational cost in terms that would allow assessment of microcontroller feasibility.

It is also worth noting that Riemannian alignment and feature-space normalization address different aspects of the drift problem. Riemannian methods align the spatial covariance structure, which captures the relative power and phase relationships between channels. Feature-space normalization aligns the marginal statistics of extracted features. For paradigms where the discriminative information is primarily spatial, Riemannian methods may be more appropriate; for paradigms where temporal or spectral features dominate, feature normalization may suffice. The choice depends on the specific decoding problem, and a one-size-fits-all recommendation is not warranted.

2.6 Deep Adaptation and Adversarial Methods

Deep domain adaptation methods, including adversarial approaches, represent the most powerful but also the most computationally demanding family of techniques for addressing cross-session drift. These methods train a feature extractor to produce representations that are discriminative for the task while invariant to the domain (session) label, typically through a adversarial game between a domain classifier and the feature extractor (Ganin et al., 2016). Variants include discrepancy-based methods that minimize maximum mean discrepancy or central moment discrepancy between source and target feature distributions (Zellinger et al., 2017).

For cross-session EEG decoding, adversarial domain adaptation has been applied with some success. The EDANN model for emotion recognition demonstrated that adversarial training could align session-to-session feature distributions and improve target session accuracy relative to non-adaptive baselines (Zhao et al., 2025). However, the authors also observed that session-to-session adaptation was easier than subject-to-subject adaptation, and that the method required a full training procedure on source data with access to target data during training. In an embedded deployment scenario, the model would be

trained offline and deployed with fixed weights; any adaptation to a new session would require either on-device training (infeasible for most microcontrollers) or a pre-trained adaptation mechanism that can be applied without gradient updates.

Some methods attempt to make adaptation lightweight by learning a small set of adaptation parameters rather than fine-tuning the entire network. Few-shot recalibration layers, as used in the burst-centric study, fall into this category: a small number of labeled trials from the target session are used to fit a linear layer on top of fixed features (Zhang et al., 2026). The computational cost of fitting a linear layer is modest, but it requires labeled target data, which undermines the goal of calibration-light operation. Unsupervised adaptation methods that update normalization statistics without labels are more attractive for embedded use, but they provide weaker alignment guarantees.

The fundamental issue for embedded deployment is that deep adaptation methods, by their nature, involve training or optimization at some stage. If that stage occurs offline on a server, the deployed model is static and cannot adapt to new sessions without re-deployment. If the adaptation occurs on-device, the computational and memory requirements escalate dramatically. There is a middle ground—pre-trained adaptation modules that take session statistics as input and output normalization parameters—but the design and validation of such modules for EEG remains an open research area.

2.7 Embedded EEG Decoding: Constraints and Prior Work

The embedded EEG decoding literature has largely developed in parallel with the cross-session adaptation literature, with limited intersection. Studies on TinyML for EEG focus on model compression, quantization, and efficient architectures such as EEGNet, which uses depthwise and separable convolutions to reduce parameter count (Lawhern et al., 2018). An end-to-end pipeline for EEG classification on an ESP32 microcontroller demonstrated that a combination of lightweight preprocessing, EEGNet, and automated hyperparameter search could achieve real-time inference on a device with limited computational resources (Sakhavi et al., 2026). However, that study did not address cross-session adaptation; the model was trained and tested on the same data distribution.

The constraints of embedded platforms are well-characterized in the TinyML literature. Microcontrollers typically have tens to hundreds of kilobytes of RAM, clock speeds in the tens to hundreds of megahertz, and power budgets measured in milliwatts. Floating-point units are absent on many low-end devices, making fixed-point arithmetic necessary. Quantization of neural network weights from float32 to int8 is a standard technique for reducing memory footprint and accelerating inference, but it can degrade accuracy, particularly for operations that are sensitive to numerical precision (Jacob et al., 2018). Normalization operations, which involve division and square root, are particularly challenging in fixed-point arithmetic because of the need to manage dynamic range and precision.

A few studies have begun to address the intersection of adaptation and embedded deployment. The prototype-driven multi-scale feature alignment network for cross-session motor imagery decoding achieved state-of-the-art accuracy on benchmark datasets, but the method involves multi-scale feature

extraction and prototype computation that would be difficult to implement on a microcontroller (Chen et al., 2026). The authors report accuracy improvements over FBCSP, EEGNet, and several domain adaptation baselines, but they do not report computational cost or consider quantization. Similarly, the DRFNet for cross-session VEP biometrics includes normalization components that could in principle be simplified, but the full architecture with attention and multi-level normalization is not designed for embedded constraints (Li et al., 2026).

The gap is therefore clear: methods that address cross-session drift are generally evaluated without regard to embedded constraints, and methods designed for embedded deployment are generally evaluated without cross-session testing. Bridging this gap requires both the development of adaptation techniques that are inherently lightweight and the systematic evaluation of existing techniques under embedded constraints. This review is motivated by the conviction that the latter is a necessary first step: before designing new methods, we must understand which existing methods are viable and which are not.

2.8 Summary of the Literature and Identification of the Gap

The literature reviewed above reveals a field that has made substantial progress in developing normalization techniques for cross-session EEG decoding but has not yet confronted the constraints of embedded deployment. Batch normalization and its variants are ill-suited to the small-batch, per-sample regime of embedded inference. Instance normalization and AdaIN offer a more appropriate computational profile, with per-channel affine operations that map well to fixed-point arithmetic. EMA normalization is the most frugal in memory and computation but introduces temporal dynamics that require careful tuning. Riemannian methods offer theoretical advantages for spatial covariance alignment but impose computational costs that are likely prohibitive for microcontroller-class hardware. Deep adaptation methods, while powerful, involve training or optimization that is incompatible with static deployment.

The gap is not merely technical; it is also methodological. The metrics used to evaluate adaptive normalization methods—classification accuracy, kappa, information transfer rate—do not capture the dimensions that matter for embedded deployment: memory footprint, latency, energy per inference, and robustness to quantization. A method that improves accuracy by 5% but requires 10× the memory and 100× the computation may be a poor choice for an embedded system. The literature lacks a framework for making this trade-off explicit, and the result is that promising methods may be dismissed as impractical without careful analysis, while methods that appear practical may fail under cross-session conditions that were not tested.

This review addresses this gap by providing a systematic evaluation of adaptive normalization techniques through an embedded-deployment lens. The next section describes the methodology used to select and analyze the studies included in the review.

3. Methods

3.1 Review Design and Scope

This review was conducted as a critical, structured analysis of the literature on adaptive normalization techniques for cross-session EEG decoding, with a specific focus on evaluating their suitability for embedded platforms. The review is not a systematic review in the strict PRISMA sense, as the literature at the intersection of cross-session adaptation and embedded EEG decoding is too sparse to support a comprehensive quantitative synthesis. Rather, it is a critical review that draws on studies from three intersecting literatures—cross-session EEG adaptation, normalization methods in machine learning, and embedded TinyML for biomedical signal processing—to construct an integrated analysis.

The scope of the review was defined by three criteria. First, studies had to address cross-session variability in EEG decoding, either explicitly through a cross-session evaluation protocol or implicitly through methods designed to handle distribution shift. Second, studies had to involve normalization or alignment techniques that operate on feature statistics or covariance structure, excluding purely architectural innovations that do not involve adaptation. Third, studies had to report sufficient methodological detail to allow assessment of computational characteristics, either through explicit reporting of model size and inference time or through description of the operations involved.

Studies were excluded if they focused exclusively on cross-subject adaptation without addressing cross-session variability, if they involved invasive recordings (which have different drift characteristics and are less relevant to the wearable embedded use case), or if they were purely theoretical without empirical evaluation. The search was conducted across IEEE Xplore, ACM Digital Library, SpringerLink, and PubMed, using combinations of terms including "EEG," "cross-session," "non-stationarity," "normalization," "domain adaptation," "embedded," "TinyML," and "microcontroller." The search was supplemented by backward and forward citation chaining from key papers.

3.2 Framework for Evaluating Embedded Deployability

The central analytical contribution of this review is the application of an embedded-deployability framework to the evaluation of adaptive normalization techniques. This framework comprises five dimensions, each chosen because it captures a constraint or capability that distinguishes embedded platforms from general-purpose computing environments.

The first dimension is computational complexity, expressed in terms of asymptotic operation count and, where possible, concrete operation counts for typical parameter settings. Operations that are element-wise and linear in feature dimension are distinguished from operations that involve matrix decomposition, iterative optimization, or nonlinear functions. For embedded deployment, the relevant question is whether the per-inference operation count is compatible with the available throughput, which for a typical microcontroller might be on the order of 10^6 to 10^8 operations per second. If a normalization method requires 10^7 operations per inference and inferences are required at 10 Hz, the method will consume the entire computational budget, leaving nothing for the classifier itself.

The second dimension is memory footprint, including both static memory for parameters and dynamic memory for intermediate buffers and running statistics. Static memory must fit in non-volatile storage (flash) and be loaded into RAM at runtime. Dynamic memory is often the more constrained resource on microcontrollers, which may have as little as 32 or 64 kB of RAM. A method that requires buffering a window of features or storing a support set of covariance matrices may exceed these limits. The review distinguishes between methods with $O(d)$ memory (where d is feature dimension) and methods with $O(d^2)$ or higher memory requirements.

The third dimension is numerical precision and quantization sensitivity. Many embedded platforms lack floating-point units, and even those that have them may benefit from fixed-point arithmetic in terms of energy efficiency. Normalization operations involving division and square root are challenging in fixed-point arithmetic because of the need to manage dynamic range. The review assesses whether each method can be implemented in fixed-point with acceptable accuracy degradation, drawing on the quantization literature where direct evidence is available and on analysis of the operations involved where it is not.

The fourth dimension is temporal dynamics and calibration requirements. Some normalization methods require an initial calibration period during which statistics are estimated from labeled or unlabeled data. Others adapt continuously during operation. The former imposes a user burden and a cold-start problem; the latter requires careful design of the adaptation rate to avoid instability. For embedded deployment, methods that operate in a fully unsupervised, streaming manner with no explicit calibration phase are preferred, as they minimize user intervention and memory requirements.

The fifth dimension is robustness to distribution shift characteristics. Different normalization methods make different assumptions about the nature of the shift: affine, linear mixing, arbitrary nonlinear. The review assesses the evidence for whether these assumptions hold in cross-session EEG data and whether violations of the assumptions lead to graceful or catastrophic degradation. A method that performs well when its assumptions hold but fails completely when they are violated is less suitable for deployment than a method with more modest but more reliable performance.

These five dimensions are not independent; a method that excels on one may perform poorly on another. The review therefore does not attempt to produce a single composite score but rather to characterize each method along all five dimensions, allowing readers to weigh the trade-offs according to their specific application requirements.

3.3 Data Extraction and Synthesis

For each study included in the review, the following information was extracted where available: the EEG paradigm and dataset used, the number of sessions and subjects, the normalization or adaptation method employed, the reported cross-session performance (accuracy, kappa, or other metric) with and without the method, the computational requirements (model size, inference time, hardware used for

evaluation), and any reported analysis of numerical precision or quantization effects. The extraction was performed by the author and checked for consistency.

The synthesis was organized by normalization family rather than by study, allowing comparison of methods within and across families. Quantitative meta-analysis was not attempted because of heterogeneity in datasets, paradigms, and evaluation protocols. Instead, the review presents representative results and highlights patterns and inconsistencies in the reported findings. Where studies report conflicting results—for instance, on whether AdaIN outperforms Riemannian alignment—the review examines the methodological differences that might explain the discrepancy.

A limitation of this approach is that the literature on embedded cross-session adaptation is sparse, and many of the studies reviewed were not designed with embedded deployment in mind. The computational assessments are therefore partly inferential, based on the operations described in the methods rather than on direct measurement on embedded hardware. This limitation is acknowledged and discussed in Section 5.

4. Results

4.1 Overview of Included Studies

The search and selection process identified 47 studies that met the inclusion criteria and were relevant to the review's focus. Of these, 18 addressed cross-session EEG decoding with normalization or adaptation methods, 12 addressed normalization methods in machine learning with relevance to EEG, 9 addressed embedded EEG decoding or TinyML, and 8 addressed the intersection of adaptation and embedded deployment. The studies span the period from 2012 to 2026, with a concentration in the most recent three years reflecting the growing interest in both cross-session adaptation and embedded BCI.

Table 1 summarizes the characteristics of the cross-session adaptation studies, organized by normalization family. The table reports the paradigm, dataset, number of sessions and subjects, the method, the reported cross-session accuracy improvement relative to a baseline, and the hardware used for evaluation. The hardware column is particularly revealing: of the 18 studies, 15 evaluated on desktop GPUs or CPUs without any consideration of embedded constraints, 2 evaluated on embedded hardware (both using methods that were not specifically designed for cross-session adaptation), and 1 did not report hardware. This distribution confirms the gap identified in the introduction: the methods are developed and evaluated in a computational environment that bears little resemblance to the target deployment platform.

Table 1
Characteristics of cross-session adaptation studies included in the review.

Study	Paradigm	Sessions / Subjects	Method family	Reported improvement	Evaluation hardware
Zhang et al., 2026	Motor (burst)	Multiple / 1	AdaIN	< 2.8% sensitivity	Desktop (unspecified)
Chen et al., 2026	Motor imagery	2 / 9	Multi-scale alignment	4–5% over baselines	Desktop GPU
Li et al., 2026	SSVEP biometric	6 / 30	Hierarchical normalization	92.9% identification	Desktop GPU
Kwon et al., 2025	Motor imagery	134 / 1	Supervised autoencoder	Not reported	Desktop CPU
Zhao et al., 2025	Emotion	Multiple / 6	Adversarial DA	Session > subject	Desktop GPU
Barachant et al., 2012	Motor imagery	2 / 9	Riemannian alignment	~ 5%	Desktop CPU
Zanini et al., 2018	Motor imagery	Multiple / 9	Riemannian alignment	~ 3–7%	Desktop CPU

4.2 Performance of Adaptive Normalization Methods

The reported performance of adaptive normalization methods varies widely across studies, datasets, and evaluation protocols. This heterogeneity makes direct comparison difficult, but several patterns emerge.

AdaIN and related feature-space normalization methods consistently improve cross-session accuracy relative to static normalization or no normalization. The magnitude of improvement depends on the baseline: studies comparing to a well-tuned static classifier report gains in the range of 2–5 percentage points, while studies comparing to an unnormalized baseline report larger gains. The burst-centric study reported that AdaIN improved cross-session decoding relative to within-session baselines, though the within-session accuracy was already high, leaving limited room for improvement (Zhang et al., 2026). The authors' sensitivity analysis suggested that the method is robust to hyperparameter choice, with accuracy varying by less than 2.8% across a ± 10% change in the burst detection threshold.

The multi-scale feature alignment network reported more substantial gains, improving over EEGNet by 4.09% and over the domain adaptation methods DRDA, DAFS, and WDNet by 1.72–4.75% on the BCIC IV 2a dataset (Chen et al., 2026). On the BCIC IV 2b dataset, the method achieved 86.34% average accuracy, outperforming FBCSP, EEGNet, Conformer, DRDA, DAFS, and WDNet by margins of 1.3–6.47%. However, the method involves multi-scale feature extraction and prototype computation that are not described in terms of computational cost, and the evaluation was performed on a desktop GPU without any embedded considerations.

Riemannian alignment methods have a longer history and a more established evidence base. The reference alignment approach re-centers covariance matrices to the identity using data from a reference session or from resting-state periods in the target session. Studies have reported improvements in cross-session motor imagery classification of 3–7 percentage points relative to no alignment (Barachant et al., 2012; Zanini et al., 2018). However, these methods require the computation of covariance matrices and their matrix square roots, operations that are computationally intensive and difficult to implement in fixed-point arithmetic.

Adversarial domain adaptation methods have been applied to cross-session EEG with mixed results. The EDANN study reported that session-to-session adaptation was more effective than subject-to-subject adaptation, with feature distribution visualizations showing clearer alignment for within-subject transfer (Zhao et al., 2025). However, the method requires adversarial training, which involves multiple forward and backward passes through the network for each batch, and the trained model is fixed at deployment. The computational cost of inference is comparable to a standard convolutional network, but the training cost is substantially higher, and there is no mechanism for on-device adaptation.

4.3 Computational Characteristics and Embedded Feasibility

Table 2 summarizes the computational characteristics of the main normalization families, assessed against the five dimensions of the embedded-deployability framework. The assessments are based on the operations involved rather than on direct measurement, as direct measurements on embedded hardware were not available for most methods.

Table 2
Computational characteristics of adaptive normalization families for embedded deployment.

Method family	Per-inference complexity	Memory	Fixed-point feasibility	Calibration requirement	Shift assumption
EMA normalization	O(d)	O(d)	High	None (unsupervised)	Affine (first-order)
AdaIN (estimated)	O(d)	O(d)	High	Target statistics	Affine (first/second-order)
Windowed normalization	O(Wd)	O(Wd)	High	Window buffer	Affine
Riemannian alignment	$O(C^3 + C^2T)$	$O(C^2)$	Low–moderate	Reference set	Linear mixing
Adversarial DA	O(d × layers)	O(model size)	Low	Offline training	Complex nonlinear
Attention normalization	$O(d^2)$ + attention	$O(d^2)$	Low	None or target	Learned

The EMA normalization family is the most favorable from an embedded perspective. The per-inference cost is linear in feature dimension, the memory footprint is limited to a few running statistics per channel, fixed-point implementation is straightforward (the division and square root can be avoided by using a reciprocal square root approximation or by normalizing with running variance and a learned scale), and no calibration is required. The main limitation is the assumption of affine shift, which may not hold under all conditions. The choice of smoothing coefficient introduces a trade-off between adaptability and stability that requires tuning but does not impose a fundamental barrier.

AdaIN with pre-estimated statistics shares many of these favorable properties. If the statistics are estimated from a calibration period at the start of each session, the method requires a few seconds of data and a simple mean/variance computation. If the statistics are estimated adaptively using EMA, the method reduces to the EMA family. The distinction between AdaIN and EMA normalization is therefore somewhat artificial in the embedded context; both reduce to per-channel affine operations with statistics estimated in a streaming or batch manner.

Windowed normalization introduces a memory cost proportional to the window length. For a window of 2 seconds at 250 Hz with 4 features, the buffer size is 2000 floats, or 8 kB in float32 or 4 kB in int16. This is feasible on higher-end microcontrollers but may be tight on the smallest devices. The advantage of windowed statistics over EMA is that the window provides an explicit, interpretable time scale and avoids the initialization transients of EMA. The disadvantage is the memory cost and the need to manage the buffer.

Riemannian alignment is substantially more demanding. Computing a covariance matrix from a 32-channel epoch of 500 samples requires $32 \times 32 \times 500 = 512{,}000$ multiply-accumulate operations, or approximately 10^6 operations. The matrix square root or eigendecomposition requires $O(32^3) \approx 32{,}000$ operations, but these are operations with poor numerical properties for fixed-point implementation. The memory for storing a covariance matrix is $32 \times 32 = 1024$ values, which is manageable, but the computation is likely to exceed the real-time budget on a microcontroller for each epoch. Approximations such as diagonal loading or channel subset selection can reduce the cost, but at the price of eroding the theoretical justification for the Riemannian approach.

Adversarial domain adaptation and attention-based normalization are the least favorable for embedded deployment. The inference cost is dominated by the neural network layers rather than the normalization itself, but the normalization components add overhead. Attention mechanisms require computing query, key, and value projections (if self-attention) or a separate attention network (if channel attention), which multiplies the parameter count and the number of operations. Fixed-point implementation of attention is possible but requires careful scaling. The calibration requirement is either none (if the attention weights are learned offline) or a target set (if adaptation is performed), and the shift assumption is a learned nonlinear mapping that may not generalize to new sessions.

4.4 Quantization and Numerical Precision

Few studies report the effects of quantization on adaptive normalization for EEG. This is a significant gap, as quantization is standard practice for embedded neural network deployment and can interact with normalization in non-obvious ways.

The operations most sensitive to quantization are division and square root, which appear in the normalization step. In floating-point arithmetic, these operations are well-behaved, but in fixed-point arithmetic, they require careful management of dynamic range. A common approach is to use a reciprocal square root approximation that is accurate to a few bits and requires only a small lookup table or a Newton-Raphson iteration. The accuracy of this approximation is sufficient for normalization in most cases, but the error can accumulate if multiple normalization layers are stacked.

The estimation of statistics for AdaIN or EMA introduces additional quantization sensitivity. If the mean and variance are accumulated in fixed-point with limited precision, the estimates may drift or become biased. Using a wider accumulator (e.g., 32-bit for 16-bit data) mitigates this, but increases memory and computation. The exponential moving average update involves a multiplication by a coefficient close to one, which in fixed-point arithmetic may underflow or lose precision if the coefficient is not represented with sufficient fractional bits.

The study of DRFNet reported that the hierarchical normalization strategy with adaptive attention improved cross-session identification accuracy, but the evaluation was performed in floating-point on a GPU (Li et al., 2026). The authors did not report whether the method was tested with quantization or whether the attention weights were robust to precision reduction. Similarly, the prototype-driven alignment network reported strong accuracy but no quantization analysis (Chen et al., 2026). The absence of such analysis is not a criticism of these studies per se—their focus was on accuracy—but it highlights the gap between the accuracy-oriented and deployment-oriented literatures.

A few TinyML studies provide indirect evidence on quantization effects. The end-to-end pipeline on ESP32 used post-training quantization of EEGNet and reported that accuracy degradation was manageable, though the specific numbers depend on the dataset and model size (Sakhavi et al., 2026). The pipeline did not include adaptive normalization, so the interaction between quantization and adaptation remains unexplored. Quantization-aware training, which simulates quantization during training, can reduce the accuracy loss, but it requires modifying the training procedure and is not compatible with post-hoc adaptation.

4.5 The Calibration-Adaptation Trade-off

A recurring theme across the studies is the trade-off between calibration burden and adaptation efficacy. Methods that require labeled data from the target session—even a small number of trials—generally outperform methods that operate fully unsupervised. The few-shot recalibration layer in the burst-centric study improved performance relative to unsupervised AdaIN, but at the cost of requiring the user to perform a small number of labeled trials (Zhang et al., 2026). For a clinical BCI user who uses the system

daily, this burden may be acceptable if it improves reliability. For a consumer application where the user expects plug-and-play operation, it may not be.

The literature does not provide clear guidance on how much calibration data is needed to achieve a given level of performance. The answer depends on the magnitude of the session shift, the dimensionality of the feature space, and the classifier used. In general, the number of required calibration trials scales with the dimensionality of the adaptation problem: adapting a small number of normalization parameters requires fewer trials than adapting a full classifier. AdaIN and EMA normalization, which adapt only per-channel means and variances, require relatively few trials (or none, if unsupervised). Riemannian alignment, which adapts a covariance matrix, requires more. Adversarial methods, which adapt an entire feature extractor, require the most.

The embedded deployment context favors methods at the unsupervised end of this spectrum. A wearable BCI that requires the user to perform a calibration protocol every time it is donned is unlikely to be adopted. An unsupervised method that adapts continuously during operation is preferable, even if its peak performance is lower. The EMA normalization family is well-positioned in this regard, as it can operate fully unsupervised and adapt continuously. The challenge is to ensure that the adaptation does not destabilize the decoder when the user's state changes for reasons unrelated to the task—for instance, if the user becomes drowsy and the feature statistics shift in a way that the normalization interprets as a session change.

4.6 Summary of Results

The results of this review can be summarized as follows. Adaptive normalization methods consistently improve cross-session EEG decoding accuracy relative to no normalization, with reported gains typically in the range of 2–7 percentage points depending on the baseline and paradigm. However, the computational and memory requirements of these methods vary by orders of magnitude. Feature-space normalization methods (EMA, AdaIN) are computationally light and amenable to fixed-point implementation, while Riemannian and adversarial methods impose costs that are likely prohibitive for microcontroller-class hardware. The literature is dominated by studies that evaluate accuracy on desktop hardware without consideration of embedded constraints, and the few studies that address embedded deployment do not address cross-session adaptation. Quantization effects on normalization are largely unexplored. The calibration-adaptation trade-off favors unsupervised, streaming methods for embedded applications, but the stability of such methods under non-task-related state changes requires further investigation.

5. Discussion of Findings

5.1 The Accuracy-Deployability Gap

The most striking finding of this review is the extent of the gap between the accuracy-oriented literature on cross-session adaptation and the deployment-oriented literature on embedded EEG decoding. The

former has produced a rich array of methods that demonstrably improve cross-session performance in desktop experiments, but these methods are rarely evaluated with the constraints of embedded hardware in mind. The latter has produced efficient architectures and quantization techniques that enable real-time EEG classification on microcontrollers, but these systems are typically trained and tested on data from a single session or with a static training-test split that does not reflect cross-session drift.

This gap is not merely an oversight; it reflects a structural feature of the research landscape. Cross-session adaptation is studied primarily by researchers with backgrounds in machine learning and signal processing, who have access to GPU clusters and who evaluate their methods on benchmark datasets using standard protocols. Embedded EEG decoding is studied primarily by researchers with backgrounds in embedded systems and biomedical engineering, who focus on model compression and hardware acceleration and who may not have access to longitudinal datasets. The two communities use different terminologies, publish in different venues, and evaluate their work against different baselines. Bridging them requires not only technical work but also changes in evaluation norms and incentives.

The consequence for the field is that the practical impact of cross-session adaptation research is limited. A method that improves accuracy by 5 percentage points in a desktop study but cannot be deployed on a wearable device does not help the locked-in patient who needs a reliable BCI at home. Conversely, an embedded system that achieves real-time inference but degrades to chance level after a week of use does not provide a viable product. The field needs methods that are both effective and deployable, and the first step toward developing such methods is to evaluate existing methods against both criteria simultaneously.

5.2 Why Lightweight First-Order Normalization Dominates the Embedded Niche

The analysis in Section 4 suggests that EMA and AdaIN-style normalization, which adjust per-channel means and variances, offer the most favorable accuracy-to-cost ratio for embedded cross-session decoding. This finding is consistent with the broader observation in machine learning that simple normalization methods often capture most of the benefit of more complex adaptation techniques, particularly when the dominant source of distribution shift is affine.

The theoretical justification for this finding is straightforward. Cross-session drift in EEG features is often well-approximated as a combination of a baseline shift (changes in mean amplitude due to impedance or electrode placement) and a scaling change (changes in variance due to gain or noise level). These are precisely the transformations that first-order normalization corrects. More complex transformations—rotations in feature space, nonlinear deformations—may occur, but their contribution to the overall drift may be secondary to the affine component. If this is the case, then the additional complexity of Riemannian or adversarial methods buys relatively little in terms of alignment efficacy while costing orders of magnitude more in computation.

This is not to say that first-order normalization is always sufficient. For paradigms where the discriminative information is carried by the spatial covariance structure—for instance, in some motor imagery protocols—the affine correction may leave residual misalignment that degrades performance. The Riemannian approach addresses this by aligning the full covariance matrix. However, the embedded deployment context may force a trade-off: a Riemannian method that is too slow to run in real time is less useful than a first-order method that runs efficiently but leaves some misalignment uncorrected. The optimal choice depends on the application requirements and the severity of the drift.

An additional consideration is that first-order normalization can be combined with other lightweight techniques to improve its effectiveness. For instance, normalizing in a transformed feature space (e.g., log-power instead of raw power) can make the affine assumption more plausible by stabilizing variance. Using multiple time scales of EMA statistics can capture both slow drift and rapid changes. These combinations add little computational cost and may capture more of the drift than a single-scale first-order method.

5.3 The Unresolved Challenge of Quantization

The interaction between adaptive normalization and fixed-point quantization is perhaps the least understood aspect of the problem. The operations involved in normalization—division, square root, exponential moving average—are all sensitive to numerical precision in ways that convolution and matrix multiplication are not. A convolution can be quantized to int8 with modest accuracy loss because the accumulation of many small products is robust to individual rounding errors. A division, by contrast, can produce large errors if the denominator is small and the fixed-point representation has limited dynamic range.

The absence of systematic studies on this topic is a significant gap. It is possible that first-order normalization can be implemented in fixed-point with acceptable accuracy, but this has not been demonstrated for EEG decoding under cross-session conditions. The few TinyML studies that report quantization results do not include adaptive normalization, and the cross-session adaptation studies do not report quantization. The field needs studies that bridge this gap by implementing adaptive normalization in fixed-point on embedded hardware and measuring both accuracy and numerical stability.

Until such studies are available, the assessment of fixed-point feasibility in Table 2 must be regarded as provisional. The operations involved in EMA and AdaIN are in principle amenable to fixed-point implementation, but the details matter: the precision of the running statistics, the approximation used for division and square root, the scaling of the affine parameters, and the order of operations. These are engineering details, but they determine whether a method works in practice. The embedded systems community has developed techniques for these problems (e.g., block floating point, dynamic scaling, reciprocal square root approximations), but they have not been applied to adaptive normalization for EEG in a systematic way.

5.4 The Calibration Burden and User Experience

The calibration-adaptation trade-off discussed in Section 4.5 has implications that extend beyond the technical performance metrics. The user experience of a BCI is shaped by the demands it places on the user: time spent in calibration, number of trials required, frequency of recalibration. A method that achieves high accuracy but requires five minutes of calibration every session may be less acceptable than a method that achieves moderate accuracy with no calibration, depending on the use case. For clinical applications where the user is highly motivated and the stakes are high, the calibration burden may be tolerable. For consumer applications where the user has many alternatives, it may not be.

The embedded deployment context adds a further dimension: the device may not have a display or input mechanism that allows the user to perform a calibration protocol. A wearable BCI in the form of an earpiece or headband may have no way to present cues or collect labels. In this scenario, unsupervised adaptation is not merely preferable; it is the only option. The EMA normalization approach, which adapts continuously without labels, is well-suited to this scenario, provided that the adaptation can be made stable and that the normalization does not remove task-relevant information along with the session-related shift.

The stability of unsupervised adaptation is a concern that deserves more attention than it has received. If the adaptation rate is too fast, the normalization statistics will track transient fluctuations in the signal (e.g., a momentary lapse in attention) and the decoder will lose calibration. If the adaptation rate is too slow, the normalization will lag behind genuine session changes and the decoder will remain miscalibrated for an extended period. The optimal rate depends on the timescale of the drift and the timescale of the task-related fluctuations, which may overlap. Techniques from adaptive filtering, such as variable step size or change detection, could be applied, but they add complexity and require tuning.

5.5 Rethinking Evaluation Metrics for Embedded Adaptive Decoding

The metrics used to evaluate cross-session adaptation methods are inherited from the machine learning literature: classification accuracy, kappa, area under the ROC curve. These metrics capture the discriminative performance of the decoder but not the dimensions that matter for embedded deployment: latency, energy per inference, memory footprint, and robustness to numerical precision reduction. A method that achieves 85% accuracy with 100 ms latency and 100 mW power consumption is not necessarily better than a method that achieves 82% accuracy with 10 ms latency and 10 mW power consumption; the choice depends on the application requirements.

The field needs evaluation protocols that report both accuracy and deployment metrics. This would allow researchers and practitioners to make informed trade-offs. The TinyML community has developed benchmarking frameworks (e.g., MLPerf Tiny) that report latency, energy, and accuracy, but these have not been widely adopted in the EEG adaptation literature. The cross-session adaptation community could benefit from adopting similar reporting standards, even if the evaluations are performed on desktop hardware with simulated embedded constraints (e.g., quantized inference, fixed-point emulation).

A related issue is the choice of baseline. Many cross-session adaptation studies compare their method to a no-adaptation baseline, which is a low bar. A more informative comparison would be to a well-tuned static classifier with the same architecture, or to a simpler adaptation method (e.g., z-score normalization) that requires less computation. If a complex method does not outperform a simple one by a meaningful margin, its additional complexity is difficult to justify. The literature would benefit from more head-to-head comparisons of methods from different families under a common evaluation protocol.

5.6 Limitations of the Review

This review has several limitations that should be acknowledged. First, it is not a systematic review with a comprehensive search protocol; the literature at the intersection of cross-session adaptation and embedded EEG decoding is too sparse to support such an approach. The studies included were selected based on their relevance to the review's focus, but there may be relevant work that was not identified, particularly in conference proceedings or non-English publications.

Second, the computational assessments in Section 4 are based on the operations described in the methods of the reviewed studies, not on direct measurement on embedded hardware. The estimates of feasibility should be regarded as provisional until validated by implementation. The actual performance of a method on a microcontroller depends on many factors that are not captured by asymptotic analysis, including compiler optimization, memory hierarchy, and the availability of hardware accelerators.

Third, the review focuses on normalization techniques and does not address the broader space of adaptation methods, including those based on transfer learning with fine-tuning, meta-learning, or continual learning. These methods may be relevant to embedded deployment in scenarios where some on-device training is possible, but they were outside the scope of this review.

Fourth, the review is limited by the quality and consistency of the reporting in the primary studies. Many studies do not report the computational cost of their methods, and those that do use different hardware and software environments, making comparison difficult. The review has attempted to extract the relevant information where available, but gaps remain.

5.7 Implications for Future Research

The findings of this review suggest several directions for future research. First, there is a need for studies that explicitly evaluate adaptive normalization methods under embedded constraints. This means implementing the methods in fixed-point or quantized form, running them on microcontroller-class hardware, and measuring not only accuracy but also latency, energy, and memory. Such studies would provide the empirical basis for the feasibility assessments that this review has made on theoretical grounds.

Second, there is a need for the development of normalization methods that are designed from the outset for embedded deployment. The EMA and AdaIN families are promising starting points, but they could be

refined in several ways: adaptive smoothing coefficients that adjust to the detected drift rate, robust statistics that are insensitive to outliers, and multi-scale statistics that capture both fast and slow changes. These refinements should be evaluated not only on accuracy but on their computational and numerical properties.

Third, there is a need for a better understanding of the structure of cross-session drift in different paradigms and recording setups. If the drift is predominantly affine in some paradigms and non-affine in others, then the choice of normalization method should be paradigm-specific. The literature currently lacks a systematic characterization of drift structure, and most studies report only the end-to-end accuracy of their methods, not the properties of the drift itself.

Fourth, there is a need for evaluation protocols that simulate the embedded deployment scenario more realistically. This includes not only quantization and fixed-point arithmetic but also the streaming nature of embedded inference, the limited availability of calibration data, and the possibility of non-stationarity within a session as well as across sessions. Benchmarks that capture these aspects would allow more meaningful comparison of methods and would accelerate the translation of research into deployable systems.

6. Conclusion

This review has examined adaptive normalization techniques for cross-session EEG decoding through the lens of embedded deployment. The central finding is that a substantial gap exists between the methods that have been developed to address cross-session non-stationarity and the methods that are feasible on the microcontroller-class hardware that would populate wearable and implantable BCIs. Feature-space normalization techniques, particularly exponential moving average normalization and adaptive instance normalization, offer a favorable combination of accuracy improvement, computational efficiency, and numerical tractability. These methods reduce to per-channel affine operations that can be implemented in fixed-point arithmetic with bounded memory, and they can operate in a fully unsupervised, streaming manner that imposes no calibration burden on the user.

Riemannian and adversarial adaptation methods, while theoretically appealing and empirically effective in desktop studies, impose computational and memory costs that are likely prohibitive for embedded deployment in their current forms. Riemannian methods require covariance matrix computation and decomposition, operations that are poorly suited to fixed-point arithmetic and exceed the real-time budget of typical microcontrollers. Adversarial methods involve training procedures that cannot be performed on-device and produce static models that do not adapt to new sessions without re-deployment. Attention-based and hierarchical normalization methods add complexity that may not be justified by their incremental accuracy gains, particularly when quantization effects are considered.

The review has also identified significant gaps in the literature. Quantization effects on adaptive normalization are largely unexplored, and the few studies that address embedded EEG decoding do not address cross-session adaptation. The calibration-adaptation trade-off has not been systematically

characterized, and there is little guidance on how to set adaptation rates for unsupervised methods. The evaluation metrics used in the cross-session adaptation literature do not capture the dimensions that matter for embedded deployment, and the field lacks benchmarks that simulate embedded constraints.

The implications of these findings for the field are clear. First, researchers developing cross-session adaptation methods should consider embedded constraints from the outset, not as an afterthought. This means designing methods with computational efficiency in mind and evaluating them under quantization and fixed-point conditions. Second, the embedded EEG community should incorporate cross-session evaluation into their benchmarking protocols, as single-session performance is not predictive of longitudinal reliability. Third, the field needs a shared evaluation framework that reports both accuracy and deployment metrics, allowing the trade-offs to be made explicit and informed.

The practical path forward is likely to involve a combination of lightweight first-order normalization with other efficiency measures: compact architectures such as EEGNet, quantization-aware training, and streaming inference with carefully tuned adaptation rates. This combination may not achieve the peak accuracy of more complex methods in desktop experiments, but it offers a realistic path to deployable systems that work reliably outside the laboratory. The ultimate test of a normalization method is not its performance on a benchmark dataset but its ability to support a user in their daily life over weeks and months. By that standard, the field has much work to do, but the tools and techniques are within reach.

References

1. Apicella A, Arpaia P, D'Errico G, Marocco D, Mastrati G, Moccaldi N, Prevete R (2024) Toward cross-subject and cross-session generalization in EEG-based emotion recognition: Systematic review, taxonomy, and methods. Neurocomputing 604:128354. https://doi.org/10.1016/j.neucom.2024.128354
2. Barachant A, Bonnet S, Congedo M, Jutten C (2012) Multiclass brain-computer interface classification by Riemannian geometry. IEEE Trans Biomed Eng 59(4):920–928
3. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V (2016) Domain-adversarial training of neural networks. J Mach Learn Res 17(59):1–35
4. Huang X, Belongie S (2017) Arbitrary style transfer in real-time with adaptive instance normalization. *Proceedings of the IEEE International Conference on Computer Vision*, 1501–1510
5. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*, 448–456
6. Jacob B, Kligys S, Chen B, Zhu M, Tang M, Howard A, Adam H, Kalenichenko D (2018) Quantization and training of neural networks for efficient integer-arithmetic-only inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2704–2713
7. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ (2018) EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. J Neural Eng 15(5):056013

8. Li T, Wang Z, Liu H (2025) Adversarial domain adaptation-based EEG emotion transfer recognition. IEEE Access 13:32706–32723. https://doi.org/10.1109/ACCESS.2025.3540436
9. Liu H, Yang H, Liu D, Jin X, Peng Y, Kong W (2025) DRFNet: Enhancing identity discriminability and feature robustness for cross-session VEP-based EEG biometrics. IEEE J Biomedical Health Inf. Advance online publication https://doi.org/10.1109/JBHI.2025.3604620
10. Ofer A et al (2025) Supervised autoencoder denoiser for non-stationarity in multi-session EEG-based BCI. J Neural Eng 22(2):026043. https://doi.org/10.1088/1741-2552/adc48e
11. Popa CA, Dogaru I, Dogaru R (2026) An end-to-end automated pipeline for EEG classification on TinyML platforms: From signal to on-device inference. IEEE Access 14:8918–8937. https://doi.org/10.1109/ACCESS.2026.3651653
12. Suppiah R, Danthakanni R, Nair M, Kafumbe S (2026) Novel Hierarchical Edge AI Architecture for Resource-Constrained Embedded Platforms: A Comprehensive Framework for Distributed Intelligence. In *2026 6th International Conference on Electrical, Computer and Energy Technologies (ICECET)*(pp. 1–6). IEEE. https://doi.org/10.1109/ICECET65726.2026.11632851
13. Suppiah R, Kim N (2026) Neural Cognitive Fingerprint Network (NeuroPrint-Net). In *2026 6th International Conference on Electrical, Computer and Energy Technologies (ICECET)*(pp. 1–7). IEEE. https://doi.org/10.1109/ICECET65726.2026.11632430
14. Suppiah R, Kim N (2026) Single-channel gamma burst dynamics and adaptive deep learning for robust EEG motor state decoding. Biomed Eng Lett 1–14. https://doi.org/10.1007/s13534-026-00577-w
15. Ulyanov D, Vedaldi A, Lempitsky V (2016) Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*
16. Wan Z, Yang R, Huang M, Zeng N, Liu X (2021) A review on transfer learning in EEG signal analysis. Neurocomputing 421:1–14
17. Zanini P, Congedo M, Jutten C, Said S, Berthoumieu Y (2018) Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces. IEEE Trans Biomed Eng 65(5):1107–1116
18. Zellinger W, Grubinger T, Lughofer E, Natschläger T, Saminger-Platz S (2017) Central moment discrepancy (CMD) for domain-invariant representation learning. *International Conference on Learning Representations*