Article

Hybrid Edge–Cloud Asymmetric Analytics for Portable Multimodal BCI Biosensors

Sayantan Ghosh , Padmanabhan Sindhujaa , Pradakshana Senthil Kumar , Anand Mohan , Pachaiyappan Mahalakshmi , Balázs Gulyás , Domokos Máthé * , Parasuraman Padmanabhan *

Posted Date: 12 May 2026

doi: 10.20944/preprints202605.0748.v1

Keywords: brain–computer interface; portable biosensors; multimodal biosignals; EEG; EMG; edge–cloud analytics; embedded TinyML inference; physiological state monitoring; multimodal fusion; digital health

 doi:10.20944/preprints202605.0748.v1

Article

Hybrid Edge–Cloud Asymmetric Analytics for Portable Multimodal BCI Biosensors

Sayantan Ghosh [1,2], Padmanabhan Sindhujaa [3], Pradakshana Senthil Kumar [4], Anand Mohan [5], Pachaiyappan Mahalakshmi [6], Balázs Gulyás [5,7], Domokos Máthé [1,*] and Parasuraman Padmanabhan [5,8,*]

[1] Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, 1085, Hungary
[2] Department of Integrative Biology, Vellore Institute of Technology, Vellore 632014, India
[3] PSG Institute of Medical Sciences & Research, Peelamedu, Coimbatore 641004, India
[4] Department of Artificial intelligence and Data science, Panimalar Engineering College, Chennai 600123, India
[5] Cognitive Neuroimaging Centre, Experimental Medicine, Nanyang Technological University, Singapore 636921, Singapore
[6] Department of Instrumentation, Vellore Institute of Technology, Vellore 632014, India
[7] Department of Clinical Neuroscience, Karolinska Institute, Stockholm 17176, Sweden
[8] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 636921, Singapore
* Correspondence: mathe.domokos@semmelweis.hu (D.M.); ppadmanabhan@ntu.edu.sg (P.P.)

Abstract

Portable biosensor hardware can now sustain continuous multimodal physiological acquisition at the edge, yet the analytical layer that converts raw signals into deployment-consistent inference remains the main bottleneck for practical embedded systems. This study addresses that bottleneck by presenting the machine-learning layer of the Real-time Cognitive Grid, the analytical companion to the previously reported hardware architecture, which equips a fixed-wiring biosensor assembly with real-time physiological-state classification through an asymmetric edge-cloud workflow. The proposed framework assigns analytical responsibility across tiers: a locked 17-feature schema comprising 5 EMG features, 6 EEG spectral features, 2 cross-modal features, 2 HRV features, 1 EOG feature, and 1 EEG quality indicator governs window-bounded inference on the Arduino Nano RP2040 Connect with an LDA edge artefact requiring approximately 716 B RAM, whereas the cloud tier supports public-dataset pretraining, hardware-aligned refinement, multimodal fusion, deployment comparison, and feature-importance analysis under the same schema contract. To evaluate analytical consistency across physiological diversity, five public repositories covering stress physiology (WESAD), affective EEG (DEAP), inertial activity recognition (PAMAP2), sEMG gesture decoding (EMG Gestures), and motor-imagery EEG (EEGMMIDB) were evaluated under subject-disjoint GroupKFold (k=5) protocols. To test whether the same contract survives translation to the physical rig, the hardware branch was evaluated under session-disjoint GroupKFold across five bench-acquired sessions. Unimodal performance was strongest in sEMG- and IMU-dominant tasks, whereas multimodal fusion improved macro-F1 by up to 0.141 over the strongest unimodal baseline in WESAD and by 0.109 in PAMAP2. In the hardware branch, the deployed edge LDA artefact reached 0.9435 macro-F1 with 0.9470 accuracy, while the retained cloud Random Forest reached 0.8792 macro-F1 with 0.8799 accuracy; feature-importance analysis further showed that the final 17-feature branch was dominated by EMG descriptors, with EEG spectral terms contributing secondary support and hardware-exclusive variables remaining weak under the present bench regime. These results show that a compact multimodal sensing assembly can be elevated beyond passive signal capture into an intelligent portable biosensor that performs context-aware interpretation with minimal user intervention, supported by a reproducible analytical workflow that remains coherent across heterogeneous benchmark repositories, hardware-specific refinement, and microcontroller-

 doi:10.20944/preprints202605.0748.v1

class deployment, thereby establishing cross-session bench feasibility as a structured basis for future multi-subject wearable validation.

Keywords: brain–computer interface; portable biosensors; multimodal biosignals; EEG; EMG; edge–cloud analytics; embedded TinyML inference; physiological state monitoring; multimodal fusion; digital health

1. Introduction

The rapid maturation of flexible and skin-conformable biosensor platforms, including electrochemical patches, textile-integrated electrodes, and soft electronic skins, has expanded the feasibility of continuous physiological sensing beyond conventional clinical environments and into portable, body-proximate monitoring settings [1,2]. Electrophysiological modalities such as EEG, EMG, and ECG, together with autonomic and environmental companion signals, can now be acquired through compact sensing assemblies with a level of continuity and practical accessibility that was previously difficult to achieve outside laboratory infrastructures [3,4]. Yet the value of these sensing platforms depends not only on the fidelity of the hardware itself, but on the analytical layer that governs how heterogeneous biosignals are segmented, represented, interpreted, and converted into stable outputs under real deployment constraints. Continuous multimodal biosignals do not become meaningful for health-adjacent or operational monitoring applications merely by being recorded; they require a coherent processing framework capable of enforcing reproducible feature structure, preserving cross-stage analytical continuity, and converting raw measurements into structured, interpretable outputs under real deployment constraints. In the absence of such a deployment-aware machine-learning architecture, even high-quality portable biosensor hardware produces data that remain difficult to interpret consistently, compare across studies, or translate into stable downstream decisions. The present work addresses that analytical gap directly. Building on the previously reported Real-time Cognitive Grid (RCG) hardware architecture [5], this study supplies the missing analytical layer by determining how multimodal physiological signals should be represented, adapted, fused, and deployed so that the embedded node functions as a state-aware inference system rather than a passive acquisition endpoint.

1.1. Background: Data-Driven Evolution of Wearable Biosensor Systems

Portable and wearable biosensor systems have markedly expanded the scale, continuity, and heterogeneity of physiological data available for computational interpretation, creating new possibilities for health-relevant monitoring across neurological assessment, affective computing, activity recognition, and stress evaluation. In this broader trajectory, wearable and body-proximal platforms are increasingly valued not only for signal access, but for their potential to support continuous, low-obtrusion physiological analytics beyond conventional laboratory workflows. As these systems incorporate increasing numbers of sensing modalities, their analytical value depends not only on the quality of signal acquisition, but on the structure of the downstream learning pipeline that converts raw measurements into stable physiological inference [6]. This dependence is particularly pronounced in BCI-oriented contexts, where electrophysiological signals are low-amplitude, noisy, non-stationary, and strongly influenced by motion, context, and inter-subject variability, thereby complicating reliable deployment beyond tightly controlled laboratory settings [7]. Under such conditions, machine learning no longer functions as an auxiliary analytical layer; it becomes a central architectural component that determines how multimodal biosignals are represented, fused, interpreted, and deployed under practical resource constraints.

Early wearable biosensor pipelines were often organised around rule-based signal processing, in which handcrafted descriptors and deterministic thresholds were sufficient for narrowly defined experimental tasks. Such approaches were effective in controlled laboratory environments, where electrode placement, ambient conditions, and participant behaviour could be tightly regulated. That

 doi:10.20944/preprints202605.0748.v1

analytical model became progressively less adequate as biosensor systems moved toward continuous monitoring, ambulatory use, and real-world inference. Under these conditions, physiological signals exhibit substantial non-stationarity, long-term drift, artefact contamination, and inter-subject variation, all of which weaken the assumptions underlying static processing pipelines [8,9]. These limitations are amplified in BCI applications, where small signal amplitudes and strict temporal requirements demand processing frameworks capable of stable segmentation, feature abstraction, and classification under operational variability rather than idealised laboratory control.

The shift toward multimodal physiological sensing has further increased this analytical burden. Contemporary portable systems increasingly combine neural, muscular, autonomic, and contextual streams, each with distinct temporal scales, spectral characteristics, signal-to-noise profiles, and artefact structures. Treating these modalities as isolated analytical problems produces fragmented workflows in which feature definitions, validation strategies, and model assumptions vary across repositories and application domains. A more reproducible alternative requires structured feature schemas, explicit fusion logic, and evaluation protocols that remain coherent across modalities and datasets [10]. Without such unification, multimodal sensing increases data volume without yielding comparably robust or deployable inference, leaving a persistent methodological gap between signal acquisition and practical physiological analytics. This gap becomes especially consequential in wearable-facing research, where analytical fragmentation can prevent otherwise promising sensing hardware from translating into stable, low-overhead monitoring pipelines.

At the same time, the computational envelope of portable hardware imposes hard limits on what can be executed locally. Continuous transmission of raw biosignals to centralised infrastructure is often inefficient or impractical under latency, energy, bandwidth, and connectivity constraints, making deployment-aware partitioning an architectural requirement rather than a later optimisation [11,12]. Within such a setting, lightweight window-bounded operations, including segmentation, constrained feature extraction, and low-overhead inference, are better suited to the edge, whereas training, cross-dataset adaptation, interpretability analysis, and longitudinal aggregation are more appropriately assigned to cloud resources. This is particularly relevant for systems intended to progress toward low-power portable or wearable operation, where energy and bandwidth constraints tighten the analytical budget further. Edge-cloud asymmetry is therefore not merely an implementation detail; it directly shapes feature dimensionality, model selection, validation design, and the overall analytical logic of the biosensor system. This shift from sensor-centric design toward deployment-aware analytics forms the background for the framework developed in the present study.

1.2. Motivation for Multimodal Physiological Analytics

Despite the growing sophistication of machine-learning pipelines for portable and wearable biosensor systems, a large proportion of existing analytical deployments continue to rely on single-modality signal streams, most commonly EEG or EMG, as the principal basis for physiological state inference [13]. Although unimodal approaches offer conceptual simplicity and lower system complexity, they remain intrinsically vulnerable to signal degradation arising from motion artefacts, electrode displacement, physiological variability, and environmental interference. In portable and wearable-facing settings, such perturbations are not exceptional events, but recurring operating conditions that can produce brittle inference behaviour and reduce analytical reliability over time [14].

Multimodal biosensing offers a principled route to mitigate these limitations by leveraging complementary information across neural, muscular, autonomic, and contextual physiological channels. Signals originating from different physiological subsystems exhibit distinct noise characteristics, temporal dynamics, and failure modes, thereby enabling redundancy and cross-validation at the feature level. Rather than treating additional modalities as isolated classification targets, multimodal analytics permits contextual or companion signals to function as stabilising priors, improving robustness under partial signal loss, transient artefacts, or subject-specific

 doi:10.20944/preprints202605.0748.v1

variability. Importantly, these benefits can often be realised through feature-level fusion alone, without requiring highly complex model architectures or prohibitive computational overhead on constrained embedded platforms [15].

The broader adoption of multimodal analytics in portable biosensor systems has nevertheless been slowed by structural fragmentation across the current research landscape. Public biosignal repositories are often modality-specific, task-confined, or evaluated within narrow methodological silos, and multimodal fusion strategies are commonly reported in isolation rather than across systematically aligned application domains. This lack of structured alignment between dataset characteristics, analytical objectives, and deployment constraints limits cross-study comparability and weakens system-level scalability, particularly when methods developed for offline experimentation are carried toward continuous, deployment-constrained inference [16,17]. In wearable-facing research, this fragmentation is especially problematic, because low-power operation, partial signal loss, and heterogeneous sensing conditions impose tighter demands on analytical coherence than those encountered in static laboratory workflows.

Addressing this gap requires analytics frameworks that explicitly account for modality heterogeneity, dataset diversity, and the computational boundaries imposed by portable platforms. The present work approaches this problem through two complementary elements: first, a structured dataset-to-application mapping that makes modality-coverage gaps explicit across the selected public repositories; and second, a hybrid edge-cloud machine-learning pipeline that operationalises deployment-aware multimodal fusion under a unified analytical schema.

1.3. Research objectives and scope

The objective of this study is not merely to evaluate another biosignal-learning pipeline, but to complete the translation of the previously reported RCG architecture into an analytically operative system. If the companion architecture paper established the physical, communication, and storage scaffold of the platform, the present work provides its machine-learning layer by defining how multimodal biosignals are converted into deployment-ready inference at both the edge and cloud tiers, focusing on the organisation, evaluation, and deployment logic of biosignal-driven workflows rather than on sensing hardware, communication infrastructure, or low-level device engineering.

The contributions of this work are fourfold. First, a systematic mapping between representative public biosignal repositories and canonical physiological monitoring application domains is developed in order to expose modality-coverage gaps, repository fragmentation, and integration limits across the current research landscape. Second, a hybrid edge-cloud machine-learning pipeline with asymmetric workload partitioning is proposed, in which segmentation, feature extraction, schema-constrained projection, and lightweight inference are executed proximal to the signal source, whereas model training, cross-dataset adaptation, interpretability analysis, and longitudinal aggregation are assigned to downstream cloud infrastructure. This design preserves deterministic, window-bounded inference at the edge while maintaining analytical scalability beyond the device layer. Third, a public-pretraining-to-hardware-adaptation training strategy is formalised to connect public-benchmark pretraining across heterogeneous repositories with hardware-specific adaptation through a shared 13-feature intersection — the subset of features computable from both the public repositories and the hardware sensor array — thereby enabling transfer to a resource-constrained embedded target without requiring full retraining at the embedded deployment layer. Fourth, the feasibility and robustness of multimodal feature-level fusion within the proposed pipeline are assessed across five public biosignal repositories under subject-disjoint evaluation protocols intended to reflect deployment-consistent inference conditions rather than benchmark-oriented optimisation alone.

The scope of the study is intentionally narrow. Hardware characterisation, sensor integration, communication protocols, and power-management considerations are addressed in the accompanying architecture-focused study and are therefore not revisited here [5]. Likewise, the present paper does not address clinical diagnosis, therapeutic intervention, or patient-specific

 doi:10.20944/preprints202605.0748.v1

decision support. All evaluations are limited to analytics-oriented feasibility assessment: the public branch uses five publicly available repositories under subject-disjoint GroupKFold protocols, and the hardware branch uses five bench-acquired sessions under session-disjoint GroupKFold; no new human-subject data are collected, and no ethical approval is required within the present study design. This constrained scope is adopted deliberately so that the work can offer a reproducible analytical framework for portable multimodal biosensor systems without overstating translational or clinical readiness. The analytical boundary of the present work within the broader RCG stack is summarised in Figure 1.

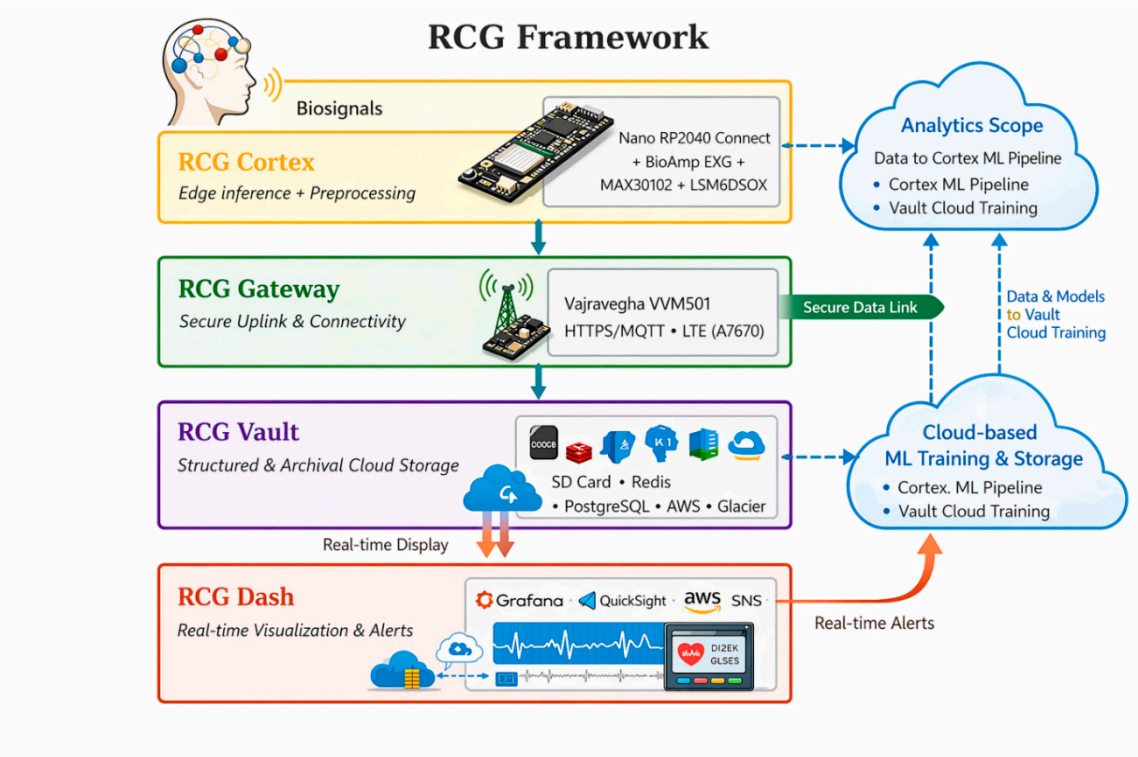


Figure 1. RCG framework layer architecture. Cortex: edge inference on Arduino Nano RP2040 Connect; Gateway: LTE uplink via Vajravegha VVM501; Vault: tiered storage (SD card → Redis → PostgreSQL → S3 → Glacier); Dash: visualisation and alerting. The analytics scope of the present study spans the Cortex ML pipeline and cloud-side model training, management, and analytical persistence within the Vault layer.

As established in the companion architecture study, the framework comprises four modular layers, namely Cortex, Gateway, Vault, and Dash. Within this broader stack, the present study is confined to the machine-learning pathway spanning edge-side inference preparation in the Cortex layer and cloud-side training, model management, and analytical persistence in the Vault layer. Figure 1 therefore serves not as a hardware schematic, but as a boundary marker for the analytical contribution of the present paper, distinguishing the machine-learning workflow addressed here from the wider communication, storage, and interface architecture described elsewhere.

The remainder of this paper is organised as follows. Section 2 reviews relevant literature on machine learning for non-invasive biosensor systems, multimodal physiological analytics, edge inference, and hybrid edge-cloud processing, and introduces the dataset-to-application coverage analysis that motivates the proposed framework. Section 3 formalises the data flow and machine-learning problem setting, including the rationale for asymmetric edge-cloud partitioning. Section 4 describes the analytical architecture, covering feature design, model construction, multimodal fusion, and the three-stage domain-adaptive training strategy. Section 5 presents feasibility-oriented evaluations across WESAD, DEAP, PAMAP2, EMGGestures, and EEGMMIDB under subject-disjoint

 doi:10.20944/preprints202605.0748.v1

validation. Section 6 discusses the findings, limitations, and implications for portable deployment-aware biosignal analytics. Section 7 concludes the paper.

2. Related Work

Research on machine learning for portable and wearable biosensor systems has developed along several partially overlapping trajectories, including non-invasive physiological signal analysis, multimodal state inference, resource-constrained edge deployment, and cloud-supported analytical orchestration. Although each of these areas has advanced substantially, their methodological development has often remained fragmented, with improvements in sensing modality, model design, and deployment architecture frequently evaluated within isolated experimental settings rather than as components of a unified analytical pipeline. Consequently, the interaction between multimodal biosignal integration, machine-learning workflow design, and deployment constraints remains unevenly addressed across the literature, particularly in portable and edge-limited contexts. Against this background, the present section reviews prior work across five connected themes: machine learning for non-invasive biosensor and BCI systems; edge and cloud computing in biosensor pipelines; lightweight and resource-constrained inference; hybrid edge-cloud analytical frameworks; and, finally, a structured synthesis of the resulting methodological gaps through dataset-to-application coverage analysis that motivates the framework developed in the remainder of this paper.

2.1. Machine Learning for non-invasive BCI biosensors

Non-invasive biosensor platforms for physiological monitoring predominantly rely on electrophysiological sensing modalities that remain feasible under portable and wearable acquisition constraints, with EEG retaining a central position because of its accessibility and high temporal resolution [18]. Surface-recorded biosignals, however, are intrinsically low in signal-to-noise ratio, highly susceptible to artefacts, and strongly non-stationary across sessions and subjects, such that the resulting inference problem is fundamentally statistical rather than deterministic [19,20]. For this reason, the evolution of non-invasive BCI and biosensor analytics has remained closely coupled to advances in machine learning, particularly in workflows designed to transform high-variance temporal signals into stable intent- or state-level predictions under imperfect recording conditions.

A canonical machine-learning workflow for non-invasive biosignal analysis typically comprises preprocessing, segmentation into short analysis windows, feature extraction, and supervised classification. Within this structure, feature engineering has historically played a decisive role, with time-domain descriptors and spectral summaries used to preserve discriminative information while attenuating nuisance variation and acquisition noise. On top of such representations, a broad range of classifiers has been explored, including linear methods, margin-based models, probabilistic formulations, and ensemble learners, with model selection generally shaped by both the physiological task and the operational constraints of the target platform [21]. Although some paradigms yield comparatively structured signal signatures, classification performance remains strongly influenced by inter-subject variability, session drift, and acquisition instability, which elevates robustness and generalisability above narrowly optimised benchmark accuracy.

In parallel, deep learning has attracted increasing attention because it can learn hierarchical representations directly from minimally processed biosignals, thereby reducing dependence on handcrafted feature design. Review-level evidence indicates that deep architectures can achieve strong performance when training data are abundant and acquisition conditions are relatively stable, but these gains are frequently accompanied by practical concerns involving computational burden, data efficiency, and sensitivity to distributional shift [22,23]. Recent convolutional approaches have reported strong results in specific EEG classification settings; however, they are commonly evaluated under offline or workstation-class assumptions, and their transfer to latency-sensitive, resource-bounded portable inference remains non-trivial without retraining, compression, or architectural simplification [24]. As a result, many deployable biosensor systems continue to favour lightweight

 doi:10.20944/preprints202605.0748.v1

models operating on interpretable feature sets, particularly where deterministic response times and predictable resource use are explicit design priorities.

Taken together, prior work establishes that non-invasive biosensor analytics is fundamentally a machine-learning problem operating under persistent noise, non-stationarity, and user variability. Yet much of the literature continues to evaluate models within modality-specific and dataset-specific silos, with comparatively limited attention to deployment-aware workflow design or scalable integration across heterogeneous physiological streams. This unresolved gap motivates the present study, which prioritises structured machine-learning workflows that remain compatible with multimodal expansion and hybrid edge-cloud execution, rather than treating model complexity itself as the principal axis of progress.

2.2. Multimodal Physiological Analytics in Wearable Settings

While the preceding subsection established the central role of machine learning in non-invasive BCI analytics, a substantial portion of the literature continues to prioritise single-modality inference, most commonly through EEG- or EMG-dominant pipelines. Such unimodal approaches simplify sensing, preprocessing, and model construction, but they remain inherently vulnerable to modality-specific degradation arising from motion artefacts, electrode impedance fluctuation, and physiological non-stationarity across sessions and subjects. These limitations become especially consequential in wearable and ambulatory settings, where acquisition conditions cannot be tightly regulated and signal quality varies over time [25]. In wearable-facing biosensor systems intended for healthcare monitoring or human-machine interaction, analytical fragility at the single-modality level directly constrains the practical value of otherwise promising sensing hardware.

Multimodal biosignal analytics has therefore emerged as a principled means of addressing these limitations by integrating complementary information across neural, muscular, autonomic, and peripheral sensing channels. Because different physiological subsystems exhibit distinct temporal dynamics, artefact profiles, and failure modes, the joint use of multiple modalities can provide both redundancy and contextual stabilisation during inference. Prior work has shown that combining modalities can improve robustness under partial signal degradation and increase resilience to context-dependent artefacts, precisely because degradation rarely manifests identically across all channels [26,27]. Importantly, such gains do not necessarily require highly elaborate fusion architectures; in many settings, feature-level or decision-level integration is sufficient when modalities are selected and aligned according to physiological relevance rather than algorithmic convenience alone.

Multimodal approaches have been explored across several application domains relevant to wearable biosensor research, including affective state recognition, stress monitoring, activity classification, and broader human-state interpretation, where peripheral physiological measurements frequently provide contextual support for neural or muscular signals. Repositories that combine physiological, behavioural, and motion-related channels have reported greater classification stability than unimodal baselines, particularly under less controlled acquisition settings [28]. In this sense, the principal value of multimodal fusion lies not merely in extracting marginal gains on idealised benchmark tasks, but in improving inference stability, contextual validity, and resistance to modality-specific corruption under realistic operating conditions. This is particularly important in wearable-oriented monitoring, where continuity, tolerance to imperfect acquisition, and analytical resilience are often more valuable than isolated peak accuracy under laboratory conditions.

Despite these advantages, the broader adoption of multimodal machine learning in portable and wearable biosensor systems remains constrained by fragmentation at the levels of data, protocol, and evaluation. Publicly available repositories are frequently modality-specific or task-confined, with limited overlap in sensing channels, annotation schemes, and experimental structure. As a result, multimodal fusion strategies are often validated within isolated dataset silos, making cross-study comparison difficult and obstructing the development of unified, deployment-consistent analytical workflows [29]. This fragmentation also weakens the relationship between dataset design, target

 doi:10.20944/preprints202605.0748.v1

application, and deployment context, a problem that becomes particularly acute when analytical workflows are expected to transition from offline evaluation toward portable, longitudinal, or wearable-facing operation.

A further limitation is that many multimodal studies implicitly assume centralised computation and relatively unconstrained processing resources. Fusion strategies that depend on high-dimensional raw inputs, continuous upstream transmission, or computationally heavy downstream models may not remain compatible with the latency, bandwidth, and energy limits that characterise portable biosensor systems [30]. These constraints point toward a more selective analytical design in which multimodal integration is organised around structured feature representations and fusion logic that remain viable within hybrid edge-cloud settings. It is precisely this need for deployment-aware multimodal analytics, rather than multimodality in abstraction, that motivates the framework developed in the present work.

2.3. Edge-Deployed and Resource-Constrained Machine Learning

The transition of BCI and biosensor analytics from laboratory-bound workflows to portable and wearable platforms introduces stringent constraints on computation, memory, energy consumption, and latency. Unlike offline analytical pipelines executed on workstation- or cloud-class hardware, edge-deployed inference must deliver predictable response times and sustained operation within tightly bounded resource envelopes. These constraints materially narrow the class of machine-learning models that remain suitable for embedded deployment, favouring approaches that are computationally tractable, data-efficient, and sufficiently robust to operational variability rather than architectures optimised primarily for offline benchmark performance [31,32]. This constraint is especially significant in wearable-facing biosensor systems intended for healthcare monitoring or human-machine interaction, where analytical responsiveness and low-overhead execution are not secondary implementation concerns, but conditions of practical usability.

Within this setting, lightweight machine-learning models operating on structured feature representations have emerged as a practical alternative to deep end-to-end pipelines. By separating feature extraction from classification, such workflows reduce model complexity while preserving interpretability, controllability, and more transparent validation of inference behaviour. Prior work in embedded and TinyML-oriented settings has shown that carefully engineered time- and frequency-domain descriptors [33,34], when coupled with shallow classifiers, can support competitive biosignal classification performance while retaining deterministic execution characteristics and modest memory requirements. These properties are particularly important in BCI and wearable biosensor contexts, where inference latency directly influences system responsiveness, feedback timing, and the stability of downstream interaction or monitoring tasks.

Quantisation and model-compression strategies further extend the feasibility of deployment on microcontroller-class platforms, but they also increase the importance of execution predictability under worst-case operating conditions. Although deep neural models remain attractive because of their representational capacity, they often exhibit less transparent runtime behaviour after compression and may remain sensitive to input dimensionality, buffer allocation, and memory-access patterns. In contrast, lightweight models built on fixed feature schemas generally offer more stable inference behaviour, simpler verification under resource limits, and easier integration with real-time signal-processing routines. For this reason, many deployable BCI and portable biosensor systems prioritise bounded inference and implementation predictability over marginal improvements in offline classification accuracy, particularly when future progression toward wearable use imposes even tighter latency and energy budgets.

A further challenge for edge-deployed machine learning concerns non-stationarity and long-term drift. Continuous online retraining or adaptive learning on embedded hardware remains difficult because of limited local data availability, constrained compute budgets, and the risk of destabilising real-time performance. Several studies therefore advocate an architectural separation between edge-side inference and upstream model training or update processes, allowing embedded

 doi:10.20944/preprints202605.0748.v1

devices to remain responsive inference nodes while retraining, aggregation, and model revision are delegated to external infrastructure [30]. This design logic aligns naturally with hybrid edge-cloud paradigms, in which the edge is responsible for deterministic local inference whereas adaptation and lifecycle management occur downstream. In wearable-facing systems, such separation is often essential, because continuous raw transmission, model re-optimisation, and high-dimensional fusion are rarely sustainable under the power and bandwidth constraints of long-duration monitoring.

Taken together, these considerations indicate that resource-constrained machine learning for BCI and portable biosensor systems is best approached through deployment-aware model selection, explicit treatment of latency and energy budgets, and a clear architectural separation between inference and training. These principles directly motivate the present study's emphasis on lightweight, window-bounded models at the edge and cloud-assisted analytical support downstream, thereby preserving real-time responsiveness without abandoning scalability or reproducibility.

2.4. Dataset-to-Application Coverage Analysis

A recurring problem in multimodal biosignal research is that dataset choice implicitly determines the class of applications that can be meaningfully evaluated. Publicly available biosignal repositories differ substantially in modality composition, annotation granularity, subject cohort size, and acquisition protocol, such that comparisons across studies often conflate model behaviour with dataset-specific structure. In the absence of an explicit mapping between repository design and target application domains, the literature remains fragmented, and datasets are too often treated as interchangeable benchmarks rather than as structured representations of particular physiological monitoring use cases.

To address this limitation, the present work organises the five public repositories used in this study according to two linked criteria: the biosignal modalities available within each dataset, and the primary annotation targets that define the associated learning task. The included modalities span neural, neuromuscular, inertial, and peripheral physiological channels, whereas the annotation targets correspond to application-relevant outcomes such as motor intent, affective or stress-related state, gesture class, and movement context. This organisation allows each repository to be interpreted in relation to representative monitoring applications rather than evaluated in isolation. The resulting dataset-to-application alignment summary is presented in Table 1.

Table 1. Dataset-to-application alignment summary for the five public biosignal repositories used in the present study.

Data set	Primary modality/ modalities	Native task formulation	Motor intent recognition	Affective -state monitoring	Physiological stress assessment	Activity classification	Gesture-based interaction	Neurophysiological state analysis
WESAD	Chest respiration, EDA, EMG	Binary stress vs non-stress	Low	Moderate	High	Low	Low	Moderate
DEAP	EEG	Binary valence classification	Low	High	Low	Low	Low	Moderate
PAMAP2	IMU, heart rate	12-class activity recognition	Low	Low	Low	High	Low	Moderate

 doi:10.20944/preprints202605.0748.v1

EMGGestures	sEMG	7-class gesture recognition	Moderate	Low	Low	Low	High	Moderate
EEGMMIDB	EEG	Binary motor imagery classification	High	Low	Low	Low	Low	High

Table 1 shows that the selected repositories are not redundant alternatives, but complementary analytical resources spanning different combinations of modality structure and task definition. No single dataset covers the full application space relevant to portable physiological analytics. Instead, each repository contributes a constrained but distinct slice of that landscape, which is precisely why a deployment-aware analytical framework must preserve dataset-native structure rather than collapsing all repositories into a falsely uniform benchmark pool.

The selected repositories collectively cover the principal biosignal domains relevant to portable and wearable-facing BCI and physiological monitoring. WESAD provides multimodal wearable recordings comprising ECG, EDA, EMG, and respiratory signals from 15 usable subjects under stress-elicitation conditions, making it representative of physiological stress monitoring and affective-state analysis. WESAD provides multimodal wearable recordings from 15 subjects (S2–S17 minus S12). Subject S12 was excluded due to incomplete sensor channels in the chest-worn device recording, leaving N = 15 subjects for downstream processing. DEAP is an EEG-based affective dataset recorded from 32 subjects and is used here as a benchmark for cortical affective inference under binary valence stratification. PAMAP2 combines IMU and heart-rate streams from 9 subjects performing 12 activity classes, representing inertial activity recognition and context-aware physiological monitoring. The UCI EMG Data for Gestures dataset contains recordings from 36 subjects performing up to seven hand gestures [35]. The dataset was evaluated as a seven-class problem (classes 0–6) over the full N = 36 subject cohort. The dataset DOI is 10.24432/C5ZP5C. EEGMMIDB is a large-scale EEG motor imagery and execution repository recorded from 104 usable subjects in the present configuration, representing cortical motor-intent classification at scale [36,37]. EEGMMIDB contains recordings from 109 volunteers. Five subjects (S088, S089, S092, S100, S104) were excluded due to annotation irregularities documented in the dataset's known-issues list [38], yielding N = 104 subjects for analysis. Motor-imagery runs were selected according to the configured EEGMMIDB evaluation setting in the current pipeline. This selection reflects the final dataset inventory adopted for evaluation of the RCG analytics pipeline.

Taken together, these repositories span four major biosignal domains, namely cortical motor intent, cortical affective state, multimodal physiological stress, and neuromuscular gesture, while mapping onto six canonical application areas: motor intent recognition, affective-state monitoring, physiological stress assessment, activity classification, gesture-based interaction, and broader neurophysiological state analysis. The corresponding coverage pattern is visualised in Figure 2.

Figure 2 makes two structural limitations in the current public-dataset landscape immediately visible. First, no single repository spans more than a limited subset of the targeted application domains, confirming the modality- and task-specific fragmentation noted above. Second, deployment-relevant scenarios such as ambulatory inference and longitudinal physiological monitoring remain only weakly represented across the selected repositories. Taken together, these gaps justify an analytical framework that can relate heterogeneous benchmark datasets to a common deployment-oriented evaluation strategy rather than treating them as interchangeable test beds.

In the present study, variability in modality availability is treated as an experimental constraint rather than as a deficiency to be artificially corrected. Evaluation protocols therefore respect the native structure of each repository, enabling unimodal baselines, multimodal comparisons, and ablation analyses to be constructed without imputing unavailable channels. This design choice ensures that

 doi:10.20944/preprints202605.0748.v1

observed performance differences reflect genuine effects of modality integration rather than preprocessing artefacts. Recording context is also treated explicitly: highly controlled repositories are analytically distinguished from more movement-linked or wearable-adjacent datasets because models validated only under tightly regulated conditions may not adequately reflect behaviour under portable, longitudinal, or body-proximal deployment assumptions [39]. The mapping developed here therefore serves both as a synthesis of the current literature landscape and as the design basis for the evaluation strategy reported in Section 5.

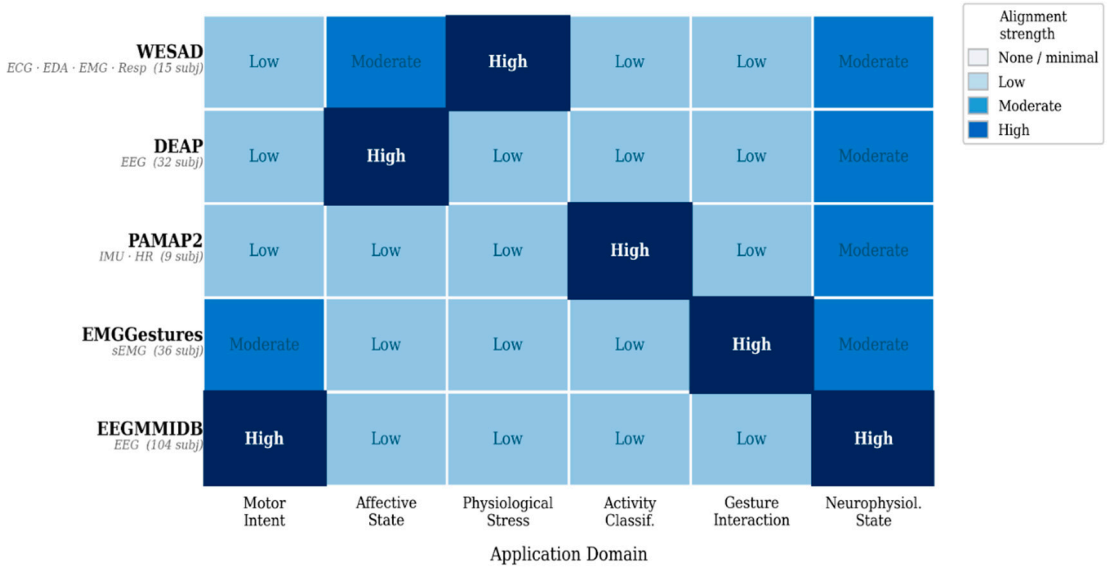


Figure 2. Dataset-to-application alignment heatmap for the five public biosignal repositories used in this study. Rows: WESAD, DEAP, PAMAP2, EMGGestures, and EEGMMIDB. Columns: motor intent recognition, affective-state monitoring, physiological stress assessment, activity classification, gesture-based interaction, and neurophysiological state analysis. Cell intensity reflects alignment strength according to modality composition, label granularity, and deployment relevance. Gaps indicate application domains with limited or no representative support across current public repositories.

2.5. Hybrid Edge-Cloud Frameworks

As BCI and portable biosensor systems extend beyond short, single-session experiments, the role of cloud-side analytics becomes increasingly important. Although edge-level inference is essential for responsiveness, local autonomy, and controlled resource use, many analytical tasks, including model training, cross-session aggregation, and broader comparative analysis, are more appropriately handled within centralised infrastructure unconstrained by strict real-time execution requirements. Prior work in health IoT and cyber-physical systems has shown that such a division of labour allows latency-sensitive processing to remain at the periphery while computationally intensive analytics are delegated upstream, thereby preserving responsiveness without collapsing all system functions into a single computational tier [40]. In wearable-facing biosensor systems, this distinction is especially consequential because healthcare-monitoring and human-interaction use cases frequently require both immediate local interpretation and longer-horizon analytical support that exceed the stable operating envelope of the edge device itself.

Within hybrid machine-learning pipelines, the cloud commonly supports tasks that are episodic, data-intensive, or exploratory in character, such as feature-distribution analysis, model selection, retrospective aggregation, and retraining on accumulated datasets. This arrangement allows edge devices to remain inference-oriented rather than continuously adaptive, while still benefiting from periodic model updates derived from broader temporal and population-level contexts. For biosignal-driven systems, this separation is particularly relevant because long-term variability across users,

 doi:10.20944/preprints202605.0748.v1

sessions, and operating conditions is difficult to capture within the limited data horizons and compute budgets available to embedded devices. Centralised analytics therefore provide a practical mechanism for improving generalisation without imposing continuous learning overhead on the edge tier [41].

In healthcare and wearable biosensing applications, hybrid edge-cloud architectures have increasingly been adopted to support longitudinal monitoring, retrospective analysis, and reproducible training workflows. Centralised infrastructure facilitates structured data persistence, standardised model development, and comparative evaluation across heterogeneous repositories, all of which are difficult to implement robustly on-device. Importantly, such architectures also enable feature-centric upstream exchange, in which compact representations rather than raw biosignals are transmitted for downstream analysis. This reduces bandwidth demand and can improve operational reliability relative to continuous raw-stream transmission, while preserving the analytical value required for model management and retrospective evaluation [42]. For wearable-oriented systems, such compact upstream exchange is often preferable to sustained raw streaming, because it better respects the energy, bandwidth, and persistence constraints that accompany body-proximal continuous sensing.

Despite these advantages, many existing hybrid machine-learning frameworks remain described at a relatively high level of abstraction, with limited treatment of how biosignal-specific constraints, including window-bounded inference, modality heterogeneity, and non-stationarity, interact with practical pipeline design. In BCI and physiological monitoring systems, the absence of explicit coordination between edge-side inference and cloud-side analytics can lead to fragmented workflows, ambiguous component responsibilities, and weak continuity between offline model development and deployed inference. This weakness becomes especially visible when multimodal sensing, portable operation, and long-horizon evaluation must be addressed within the same analytical framework. The present study therefore approaches hybrid edge-cloud design not as a generic systems pattern, but as an explicitly biosignal-aware analytical strategy shaped by modality structure, deployment limits, and the requirements of reproducible physiological inference.

2.6. Summary and Identified Research Gaps

The preceding review establishes machine learning as a core analytical component of non-invasive BCI and wearable biosensor systems, particularly under conditions of noise, non-stationarity, and operational variability. Prior work has demonstrated the value of feature-driven machine-learning pipelines for portable deployment, the usefulness of multimodal biosignal integration for improving inference stability, and the practical importance of separating edge-side responsiveness from cloud-side analytical scalability. Taken together, these developments provide essential foundations for contemporary physiological analytics; however, they have often advanced within partially disconnected methodological streams rather than within a unified, deployment-oriented framework.

A first gap lies in the fragmentation that persists across sensing modalities, datasets, and application domains. Although multimodal fusion has shown clear value in improving robustness, many studies remain confined to dataset-specific or task-specific evaluation settings, which limits cross-study comparability and reduces the reusability of analytical workflows. As a result, there remains no broadly structured route for aligning heterogeneous biosignal repositories with common physiological monitoring objectives relevant to portable and wearable-facing systems.

A second gap concerns the weak integration between machine-learning workflow design and deployment architecture. Many published pipelines continue to assume centralised processing or comparatively unconstrained computation, with limited attention to the realities of edge-deployed inference. Consequently, the interaction between window-bounded processing, feature representation, multimodal fusion, and hybrid edge-cloud execution is often left implicit rather than formalised as part of the analytical design itself.

 doi:10.20944/preprints202605.0748.v1

A third gap relates to scalability across sessions, users, and operating conditions in portable BCI and wearable biosensor systems. Although cloud-based analytics are widely recognised as useful for retraining, aggregation, and retrospective analysis, their coordination with edge-level inference is frequently described only at a conceptual level, without explicit treatment of biosignal-specific issues such as non-stationarity, modality heterogeneity, and schema stability across training and deployment stages. This weakens continuity between offline analytical development and sustained deployment-oriented operation.

Taken together, these gaps motivate the need for analytical frameworks that align dataset structure, multimodal feature processing, and hybrid machine-learning deployment logic within a single, deployment-consistent workflow. The following section therefore formalises the corresponding data flow and machine-learning problem setting and defines the partitioning rationale and feature schema that underpin the framework developed in this study.

3. Data Flow and Problem Formulation

The machine-learning layer of a portable biosensor system does not begin with model selection; it begins with a set of analytical commitments regarding how data are structured, what the learning problem is, and where each computational step is executed. This section formalises those commitments for the analytics pipeline by defining the data flow, problem setting, feature schema, and partitioning logic that govern the architectural and experimental choices developed in the remainder of the paper. The treatment is intentionally abstracted from hardware implementation: signal sources are characterised by modality type and sampling structure rather than by the specific sensors or acquisition electronics from which they originate.

3.1. End-to-End Analytical Data Flow

The analytics pipeline operates through two parallel input branches that converge within a shared training and deployment pathway. The first branch draws on five public biosignal repositories, namely WESAD, DEAP, PAMAP2, EMGGestures, and EEGMMIDB, and applies dataset-specific windowing together with modality-aligned feature extraction to generate unimodal baselines, multimodal fusion inputs, and public-domain pretraining statistics within a shared feature space. The second branch operates on hardware-acquired multimodal biosignals sampled at 250 Hz across four physiological channels, from which a fixed 17-feature representation is derived under the schema constraints defined in Section 3.3. The full on-device descriptor vector computed at inference time from the Nano RP2040 comprises four additional hardware-exclusive features — two HRV variability proxies derived from the ECG channel, one EOG-derived blink-rate estimate, and one EEG signal-quality indicator — that are not present in the public datasets, bringing the deployment schema to 17 features. This 17-feature vector is used for Stage 2 (domain adaptation fine-tuning) and all on-device deployment. In this branch, the analytical emphasis is placed on schema-constrained feature generation rather than raw-signal persistence, so that downstream learning and deployment remain anchored to stable feature representations rather than continuous waveform dependence.

Both branches converge within a three-stage analytical pathway. Stage 1 performs public-domain pretraining across heterogeneous repositories using the shared 13-feature intersection. Stage 2 adapts this representation to the hardware branch, first within the same shared feature space and then under extension to the full 17-feature RCG schema. Stage 3 produces deployment artefacts in two forms: a cloud-resident model retained for downstream analytics and longitudinal aggregation, and a hardware-optimised edge model serialised for direct on-device inference. The end-to-end analytical flow of this pipeline is summarised in Figure 3.

In parallel with the biosignal acquisition stream, the hardware node collects a low-rate environmental and physiological context stream via the auxiliary sensor suite — comprising gas sensors (MQ-3, MQ-7, MQ-9, MQ-135), ambient climate sensors (DHT11 temperature and relative humidity), photoplethysmographic heart rate and SpO_2 (MAX30102), and skin temperature (MAX30205) — sampled at 1 Hz and transmitted to the cloud through the same gateway uplink. This

 doi:10.20944/preprints202605.0748.v1

telemetry stream does not enter the 17-feature inference schema and is not used for window-level classification; it operates on a structurally separate pathway within the Vault layer. Its role is to support cloud-side contextual analytics: post-hoc window quality annotation, session drift detection, environment-to-inference correlation, and alert provenance tracing. Specifically, gas alarm flags and temperature anomalies recorded in the telemetry log can be joined to the window-level inference record by timestamp, allowing the cloud to suppress or flag windows whose acquisition context was environmentally compromised without modifying the edge inference logic or the schema contract. This design preserves the determinism and resource bounds of the edge tier while enabling richer retrospective interpretation of inference outputs in the cloud — a concrete realisation of the longitudinal aggregation and retrospective analysis capabilities assigned to the Vault layer in the companion architecture study.

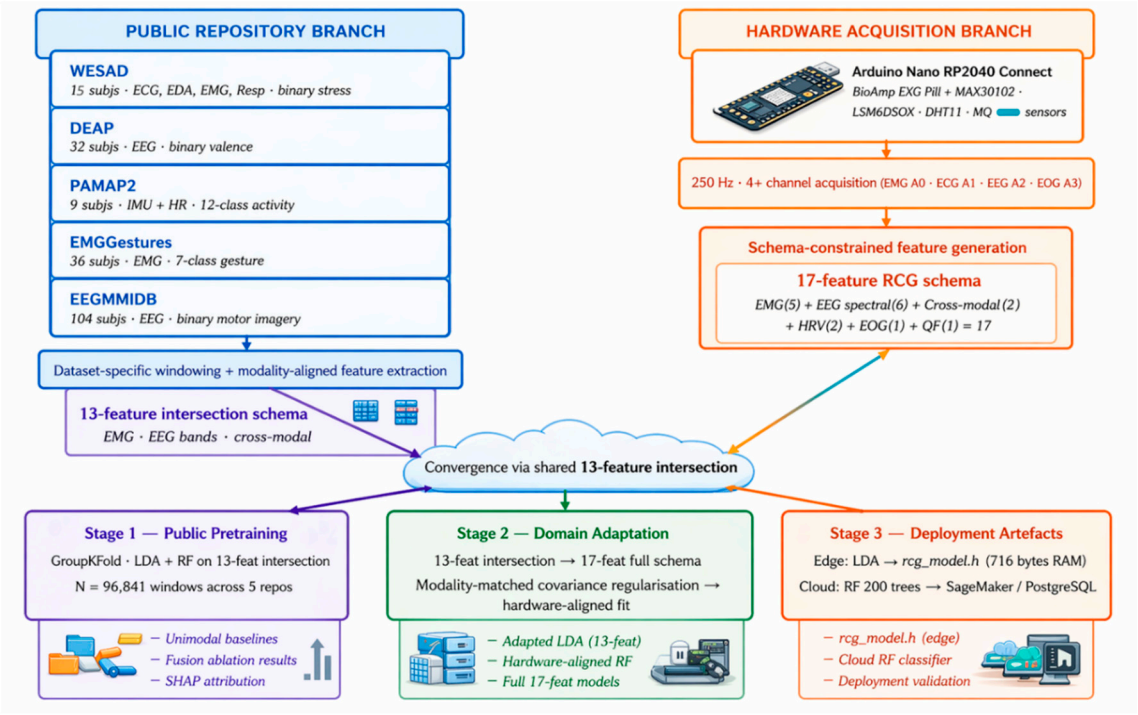


Figure 3. End-to-end analytical data flow of the machine-learning pipeline. The public branch comprises repository loading, dataset-specific windowing, and modality-aligned feature extraction. The hardware branch comprises 250 Hz four-channel acquisition followed by schema-constrained feature generation under the 17-feature representation. Both branches converge within a public-pretraining-to-hardware-adaptation training pathway that produces cloud-resident and edge-deployable inference artefacts.

Figure 3 clarifies that the public and hardware branches are not independent evaluation tracks but coordinated components of a single deployment-aware analytical workflow. The public repositories were used not to benchmark the embedded rig against external cohorts, but to demonstrate that the proposed analytical workflow remains coherent across heterogeneous biosignal tasks under subject-disjoint evaluation. The hardware branch then served as a proof-of-function study showing that the same deployment-consistent analytical contract remains executable on the embedded rig under session-disjoint bench conditions. The public branch contributes breadth across modality classes and application domains, whereas the hardware branch imposes the schema and execution constraints required for embedded deployment. Their convergence through the shared intersection feature space is what permits public-pretraining-to-hardware transfer without abandoning compatibility with the final edge-resident representation. The telemetry stream, while analytically separate from both branches, completes the data-flow picture by providing the cloud tier with the contextual grounding it needs to interpret, annotate, and validate inference outputs

 doi:10.20944/preprints202605.0748.v1

longitudinally. In this way, the data-flow design is not merely organisational; it is the analytical mechanism through which benchmark diversity, deployment consistency, and cloud-side contextual intelligence are made mutually compatible within a single coherent architecture.

3.2. Problem Formulation Across Heterogeneous Biosignal Repositories

The learning problem is formulated uniformly across all data sources as supervised classification at the window level. Each temporal window yields a fixed-dimensional feature vector and is assigned a label drawn from the annotation scheme native to the source dataset. Label spaces are therefore treated as intrinsic properties of the individual repositories, spanning binary formulations for stress detection and motor imagery, multi-class settings for activity and gesture recognition, and task-specific affective stratifications where applicable. No attempt is made to merge or harmonise these label spaces across repositories. This design preserves the integrity of the native annotation structure of each dataset and avoids artificial alignment that would obscure genuine differences in task definition, label semantics, and classification difficulty.

Modality availability is treated as an experimental constraint rather than as a deficiency requiring correction. Because the selected repositories differ substantially in their available physiological channels, no missing modalities are provided, and no absent channels are synthesised. Unimodal baselines are therefore constructed only from modalities natively present within a given repository, whereas multimodal configurations are restricted to those channels that are actually available in that dataset. This constraint ensures that observed performance differences reflect genuine effects of modality integration rather than artefacts introduced by data-completion strategies [6].

To reduce optimistic leakage across temporally correlated biosignal samples, all train-validation partitions in the public benchmark branch are enforced as subject-disjoint. Cross-validation is implemented through grouped folds in which all windows from a given subject are assigned exclusively to either the training or the validation partition, and never split across both. Window-level random partitioning is explicitly excluded because it inflates performance estimates by allowing models to encounter samples from the same subject during both fitting and evaluation [17]. This subject-disjoint protocol is applied uniformly across all five public repositories and serves as the primary evaluation rule for the benchmark arm of the analytical pipeline.

For the hardware branch, the same anti-leakage principle is applied directly. The hardware dataset comprises five bench sessions with distinct signal-level characteristics. Session identity is used as the grouping variable in a leave-one-session-out GroupKFold protocol (k=5), so that all windows from a given session appear exclusively in either the training or the validation partition. This session-disjoint structure prevents leakage across temporally correlated windows while providing a cross-session robustness estimate that is structurally analogous to the subject-disjoint protocol applied in the public benchmark branch.

3.3. Feature Representation and Schema Constraint

Feature extraction is applied to each temporal analysis window to produce a fixed-dimensional representation that remains compatible with both cloud-side model development and edge-oriented inference. The pipeline adopts a fixed 17-feature schema organised across five physiological groups: five EMG time-domain descriptors capturing amplitude and activation dynamics; six EEG spectral features spanning canonical frequency bands together with spectral entropy; two cross-modal spectral descriptors; two HRV-derived features from ECG; and one EOG-derived feature, supplemented by a signal-quality indicator to complete the 17-dimensional representation. All features are computed through standardised time- and frequency-domain operations, with spectral estimates derived using Welch-based procedures [34] and temporal descriptors selected for low-overhead physiological interpretability [39]. This design favours analytical stability and deployment continuity over dataset-specific feature proliferation.

 doi:10.20944/preprints202605.0748.v1

A central constraint of this representation is the distinction between the full 17-feature hardware-aligned schema and the reduced 13-feature intersection available across the public repositories used for model development. The hardware-exclusive subset comprises the HRV-derived pair, the EOG blink-rate feature, and the signal-quality indicator, none of which has a direct and consistently aligned counterpart across the selected public datasets. The shared 13-feature intersection, consisting of the EMG, EEG-band, and cross-modal groups, therefore functions as the representational contract linking public-data pretraining to hardware-facing deployment. This explicit schema partition avoids artificial feature imputation and preserves continuity across training stages while acknowledging genuine differences in modality availability.

Before model fitting, feature vectors are standardised using scaling parameters estimated exclusively from the training partition. This prevents information leakage while preserving comparability across data sources and evaluation folds. Feature dimensionality and ordering are enforced through schema validation at each pipeline stage, ensuring that cloud-side training, transfer-oriented adaptation, model export, and edge-facing inference all operate under the same representational assumptions regardless of data origin. In this sense, the schema functions not merely as a feature inventory, but as the formal mechanism through which deployment consistency is maintained across the analytical workflow.

3.4. Asymmetric Edge–Cloud Partitioning Strategy

The analytics pipeline assigns computational responsibility across edge and cloud tiers through an explicitly asymmetric design. Machine-learning operations are not distributed uniformly; instead, each stage is placed according to its latency sensitivity, data dependency, and computational footprint. The edge tier is reserved for time-bounded operations, namely windowed segmentation, schema-constrained feature generation, and lightweight fixed-parameter inference, all of which must execute within deterministic resource budgets and without dependence on continuous network availability [43]. The cloud tier, by contrast, is assigned episodic and data-intensive operations, including supervised model training, cross-dataset adaptation, longitudinal aggregation, and model versioning, which benefit from centralised compute and accumulated data but do not impose strict real-time constraints [42].

This asymmetry is methodological rather than merely architectural. Edge-side inference is stateless with respect to learning: model parameters are fixed at deployment and remain unchanged during routine operation. Model revision occurs upstream and is propagated downstream only through episodic update cycles, thereby decoupling inference stability from training dynamics. Such separation is well motivated in portable biosensor settings, where intermittent connectivity, limited memory, and constrained energy budgets make continuous on-device learning both impractical and difficult to validate reliably [15,33]. Data exchange between tiers is correspondingly restricted and structured. Derived feature representations and inference outputs are eligible for upstream transfer, whereas model updates move downstream from cloud to edge. Continuous raw-signal transmission is therefore avoided, preserving a clear analytical interface between the two computational tiers.

Within the framework, this partition maps directly onto the modular boundary established in the companion architecture: the Cortex layer supports edge-side feature generation and inference, whereas the Vault layer manages cloud-side training, model persistence, and analytical aggregation. The partition is therefore not a secondary implementation convenience, but the governing principle through which real-time responsiveness at the edge is made compatible with broader analytical scalability upstream. Section 4 describes how this division is realised within the machine-learning architecture developed in the present study.

3.5. Deployment-Aware Evaluation Logic

Evaluation within the analytics pipeline is designed to reflect deployment conditions rather than to maximise benchmark performance under assumptions that would not hold in practice. Two constraints govern all experimental protocols reported in Section 5.

 doi:10.20944/preprints202605.0748.v1

The first constraint is subject-disjoint partitioning. All cross-validation folds in the public benchmark branch are constructed such that every window from a given subject appears exclusively in either the training or the validation set, and never in both. This is enforced using grouped cross-validation with subject identifiers as the grouping variable, applied uniformly across all five public repositories. Window-level random partitioning is explicitly excluded because it allows the model to encounter temporally correlated samples from the same subject during both fitting and evaluation, thereby inflating performance estimates that do not generalise to unseen individuals.

The second constraint is deployment-consistent feature generation. Feature extraction parameters, including window length, stride, and schema dimensionality, are held fixed within each analytical branch so that the representations used during cloud-side training remain structurally compatible with those produced during hardware-side inference. This preserves the schema contract established in Section 3.3 and ensures that model evaluation reflects the same representational assumptions that govern deployment. Reported metrics are therefore interpretable as estimates of deployment-ready analytical behaviour, within the limits of bench-level validation, rather than as results obtained under a disconnected offline-processing regime.

Taken together, these constraints define the evaluation logic of the present study as deployment-consistent rather than benchmark-optimised. That distinction is central to the analytical claims made in Section 5 and to the limitations discussed later in Section 6.

4. Analytical Architecture and Model Development

This section instantiates the formal framework defined in Section 3 as a concrete multi-stage analytical pipeline. It shows how the data-flow structure, schema constraints, asymmetric edge-cloud partitioning, and deployment-consistent evaluation logic are realised within the model-development workflow of the system. The presentation follows the operational order of the pipeline: public-dataset processing and baseline construction, multimodal fusion and ablation analysis, joint training and cross-domain adaptation, and, finally, deployment artefact generation for edge and cloud targets. The emphasis throughout is placed on functional role, representational continuity, and deployment compatibility rather than on algorithmic novelty in isolation. Where individual model classes are introduced, their selection is governed by the constraints formalised in Section 3.4, namely determinism, bounded resource footprint, and compatibility with the fixed feature schema.

4.1. Public Dataset Processing and Baseline Model Construction

Each public repository is processed through a dataset-specific loading and segmentation procedure that preserves the native sampling characteristics, annotation structure, and physiological task definition of the source dataset. Windowing parameters are selected to balance temporal responsiveness, feature stability, and downstream deployment relevance while remaining compatible with the unified analytical workflow of the pipeline. Across all repositories, feature extraction is performed within the shared 13-feature intersection space defined in Section 3.3, thereby preserving continuity between public-domain analysis and the later hardware-aligned adaptation branch. The dataset-specific processing configuration used throughout the public benchmark stage is summarised in Table 2.

Table 2. Dataset-specific processing and evaluation protocol used in the public benchmark branch of the analytics pipeline.

Dataset	Modality/modalities used	Native task formulation	Window length	Stride / overlap	Split policy
WESAD	Chest respiration, EDA, EMG	Binary stress classification	5 s	2.5 s stride	Subject-disjoint GroupKFold
DEAP	EEG	Binary valence classification	4 s	2 s stride	Subject-disjoint GroupKFold

 doi:10.20944/preprints202605.0748.v1

PAMAP2	IMU (hand, ankle), heart rate	12-class activity classification	5.12 s	1 s stride	Subject-disjoint GroupKFold
EMGGestures	sEMG	7-class gesture classification	250 ms	50% overlap	Subject-disjoint GroupKFold
EEGMMIDB	EEG	Binary motor imagery classification	2 s	1 s stride	Subject-disjoint GroupKFold

Table 2 makes explicit that deployment consistency in the present study is enforced not by imposing identical raw-signal conditions across repositories, which would be impossible, but by standardising the analytical contract under which each dataset enters the pipeline. Windowing remains dataset-specific because the underlying acquisition regimes and task structures differ, yet the evaluation rule, feature-space logic, and subject-disjoint protocol are kept aligned. This balance allows repository-native structure to be preserved without breaking continuity across the broader analytical workflow.

Unimodal baselines are then constructed for each dataset and for each natively available modality using three classifier families chosen for complementary behaviour under deployment-relevant constraints. LDA is included because of its closed-form training and minimal inference burden [13]. SVM with an RBF kernel provides a compact non-linear margin-based baseline under fixed feature representations [44]. Random Forest is included as an interpretable ensemble learner capable of modelling non-linear decision structure without requiring representation learning [45]. All models operate on standardised feature vectors and are evaluated under the subject-disjoint grouped cross-validation protocol defined in Section 3.5, with five folds applied consistently across the public repositories. Performance is reported as mean ± standard deviation across folds using accuracy and macro-F1, thereby capturing both aggregate classification performance and class-balanced behaviour under unequal label distributions.

Regarding model configuration, unless otherwise stated, all public-branch classifiers were evaluated under fixed, non-nested settings to preserve comparability across repositories. LDA used the default closed-form formulation. SVM employed an RBF kernel with C = 1.0 and gamma = scale. Random Forest used n_estimators = 200 and max_depth = 15, with random_state = 42 and parallel fitting enabled where supported. In the transfer branch, the public Random Forest pretraining stage used 150 trees, followed by warm-start expansion with 50 additional hardware-trained trees. For deployment comparison, the cloud-resident Random Forest retained the 200-tree, depth-15 configuration, whereas the edge fallback Random Forest was pruned to 20 trees with maximum depth 5 to remain compatible with the embedded memory budget during deployment comparison. Within the analytical workflow, these unimodal results serve two roles. First, they establish dataset- and modality-specific reference points against which subsequent multimodal fusion results can be interpreted. Second, they provide the public-domain baseline layer required for later transfer-oriented training and deployment-aware model selection. The unimodal stage is therefore not treated as an isolated benchmarking exercise, but as the reference frame from which the broader analytical architecture developed in the following subsections proceeds.

4.2. Multimodal Fusion and Ablation Analysis

For repositories containing multiple natively available modalities, a structured ablation procedure is used to characterise the contribution of individual modalities and modality combinations to classification performance. Evaluation begins with unimodal configurations, in which the feature space is restricted to a single physiological source for each run. Multimodal configurations are then constructed through feature-level fusion, whereby feature vectors extracted independently from each available modality are concatenated within a shared analysis window before model fitting and inference. This strategy preserves modality-specific preprocessing while retaining a common workflow structure across unimodal and multimodal branches, thereby remaining compatible with the deployment-oriented analytical logic established in Section 3.4. For

 doi:10.20944/preprints202605.0748.v1

repositories exposing only a single modality, ablation is restricted to within-modality analytical comparisons rather than cross-modality fusion.

Feature-level fusion is preferred over signal-level and decision-level alternatives for two reasons. First, it preserves the representational logic of the overall pipeline by allowing each modality to undergo preprocessing and feature extraction in its native form before integration. Second, it supports direct post hoc assessment of modality contribution without introducing fusion-specific architectural parameters that would complicate interpretation under a fixed deployment-oriented workflow [10]. Attribution analysis is performed on the Random Forest branch using SHAP-based feature importance where supported, with permutation importance used for the hardware-aligned feature ranking reported later in the pipeline. Feature contributions are examined both individually and after aggregation by modality group, allowing the relative influence of neural, muscular, autonomic, and contextual signals to be interpreted across datasets and task settings.

The resulting ablation design yields, for each applicable repository, a structured comparison spanning unimodal baselines, pairwise multimodal combinations where relevant, and full multimodal configurations. Performance deltas between unimodal and fused settings are then computed under the subject-disjoint evaluation protocol defined in Section 3.5, allowing the marginal contribution of multimodal integration to be quantified without collapsing differences in task structure or dataset composition. These outputs provide the principal analytical basis for the multimodal feasibility interpretation presented in Section 5.

4.3. Joint Training and Cross-Domain Adaptation

The joint training stage links public-domain model development to hardware-aligned deployment through a structured staged adaptation procedure. This design addresses a central representational asymmetry within the pipeline. The five public repositories provide broad supervised coverage across heterogeneous physiological domains, but only within the shared public-branch feature space aligned to the 13-feature intersection contract. The hardware branch, by contrast, provides the full 17-feature representation, yet over a substantially smaller labelled sample and under a narrower acquisition context. Training solely on hardware data from random initialisation would discard distributional information available across the public repositories, whereas direct transfer without adaptation would ignore the genuine shift between repository-derived features and the hardware-facing signal distribution [46]. The staged procedure is therefore designed to preserve transferable structure while explicitly adapting it to deployment-facing data.

In Stage 1, supervised models are trained on public-domain features within the shared 13-feature intersection schema under the subject-disjoint evaluation logic defined in Section 3.5. This stage establishes the initial class-conditional structure and decision behaviour from the aggregate distribution represented by the five public repositories, thereby providing a starting point that is broader than any single dataset alone. Two model families are retained at this stage, namely LDA and Random Forest, because they support distinct but deployment-compatible adaptation mechanisms in the subsequent stages.

Stage 2 performs adaptation on hardware-aligned data. The first step is conducted within the same 13-feature intersection space, allowing the public and hardware branches to interact under a shared representational contract. For LDA, adaptation proceeds analytically through regularisation of the hardware covariance structure using a modality-matched public prior, thereby updating the decision boundary without requiring the full public feature matrix to remain active during deployment-oriented refinement. For Random Forest, the same 13-feature intersection is retained as the shared representational space for hardware-side fitting; however, because the public pooled model and the hardware task operate over incompatible class spaces, this branch is interpreted operationally as a deployment-consistent hardware fit rather than as a clean parametric transfer in the same sense as LDA. The second step then extends the adapted representation into the full 17-feature schema by introducing the hardware-exclusive variables, namely the two HRV descriptors,

 doi:10.20944/preprints202605.0748.v1

the EOG blink-rate feature, and the signal-quality indicator. In this way, adaptation is stabilised first within the common intersection domain before the full hardware-facing schema is activated.

For LDA, the transition into the 17-feature space is realised by carrying forward the intersection-stage regularisation logic into a fresh model estimated on hardware-aligned data under the full schema. For Random Forest, the full 17-feature hardware-aligned model is likewise estimated under the deployment schema, but in the present study this branch should be understood as schema-consistent hardware fitting rather than as a distinct transferable prior carried forward from the pooled public stage. This staged progression ensures that public-domain pretraining and hardware deployment remain analytically linked through the shared feature space, while the hardware-exclusive variables are introduced only after the intersection-stage transition has been established. For LDA, this linkage is realised as explicit cross-domain adaptation; for Random Forest, it is realised primarily as schema continuity under a common deployment contract. The resulting design preserves continuity across model-development phases without relying on artificial feature imputation or assuming direct equivalence between public and hardware distributions.

4.4. Edge and Cloud Deployment Artefact Generation

Following the joint training and adaptation stages, the pipeline produces two distinct deployment artefacts that operate under the same schema contract while targeting different computational tiers. The cloud-resident artefact is a Random Forest classifier trained on the full 17-feature hardware-aligned representation and retained for longitudinal analytics, comparative evaluation, and periodic retraining. The edge-resident artefact is an LDA classifier selected as the primary embedded inference engine because of its closed-form parameterisation, low execution overhead, and minimal memory footprint under the deployment constraints. This model is serialised into a schema-locked static representation suitable for direct inclusion in firmware, so that the trained decision boundary can be evaluated on each incoming feature vector through fixed coefficient operations without requiring dynamic memory allocation or architecture-specific learning logic at runtime.

Prior to deployment, both artefacts undergo a validation step that verifies schema conformity, parameter integrity, numerical stability, and execution feasibility under the intended operating conditions. Models that fail to satisfy predefined latency or memory constraints are excluded from deployment irrespective of their offline performance, thereby enforcing a strict separation between analytical optimisation and operational readiness. Each retained artefact is associated with a specific training configuration, schema definition, and validation outcome, allowing model provenance to remain reproducible across update cycles and enabling systematic comparison between successive deployment versions [47].

Beyond the inference and model management pathway, the hardware branch generates a structurally separate environmental and physiological context stream through the auxiliary sensor suite. This telemetry stream — comprising gas sensor readings (MQ-3, MQ-7, MQ-9, MQ-135), ambient climate measurements (DHT11 temperature and relative humidity), photoplethysmographic heart rate and SpO_2 (MAX30102), and skin temperature (MAX30205) — is sampled at 1 Hz and transmitted to the Vault layer through the same gateway uplink as the biosignal feature stream. It does not enter the 17-feature inference schema and is not used for window-level classification. Its analytical role operates at the session level in the cloud: post-hoc window quality annotation, environment-to-inference correlation, session drift detection, and alert provenance tracing. Concretely, gas alarm flags and temperature anomalies recorded in the telemetry log are joined to the window-level inference record by timestamp, allowing the cloud to suppress or contextualise windows whose acquisition environment was compromised without modifying the edge inference logic or the schema contract. This design realises the longitudinal aggregation and retrospective analysis capabilities assigned to the Vault layer as concrete, operationally grounded functions rather than abstract architectural commitments.

 doi:10.20944/preprints202605.0748.v1

Model revision is performed episodically rather than continuously. Updated artefacts are generated through cloud-side retraining on accumulated or revised data, validated against the incumbent baseline, and propagated to the edge tier only after passing the deployment gate described above. During the interval between updates, the edge system continues to operate using the previously validated model, thereby preserving uninterrupted inference without undefined intermediate states. Edge-resident inference remains fully functional in the absence of network connectivity, with cloud interaction restricted to asynchronous upstream transfer of derived feature summaries and downstream receipt of validated model updates. This offline-capable design follows directly from the asymmetric partitioning strategy defined in Section 3.4 and preserves analytical continuity under the intermittent connectivity conditions that commonly characterise portable biosensor deployment.

4.5. Operational Summary of the ML Stack

The four stages described in this section together form a unified and sequentially dependent analytical pipeline, in which the output of each stage serves as an input, constraint, or deployment condition for the next. Public-dataset processing and baseline construction establish the reference performance layer for subsequent fusion and adaptation analysis. Multimodal fusion and ablation analysis extend this reference layer by characterising modality interaction effects and producing the attribution outputs examined later in Section 5. Joint training and cross-domain adaptation then combine public-domain structure with hardware-aligned feature representations to produce schema-consistent models in both the shared 13-feature intersection space and the full 17-feature deployment space. Finally, deployment artefact generation transforms these adapted models into tier-specific outputs that operationalise the asymmetric partitioning strategy defined in Section 3.4. The functional responsibility of each stage, together with its active schema, computational tier, and principal outputs, is summarised in Table 3.

Table 3. Operational summary of the machine-learning stack.

Stage	Primary function	Active schema	Computational tier	Principal outputs
§4.1 Public baseline construction	Dataset loading, windowing, and unimodal baseline evaluation	13-feature intersection	Cloud	Baseline results and dataset-specific reference statistics
§4.2 Fusion and ablation analysis	Multimodal combination, ablation comparison, and attribution analysis	Dataset-native within shared workflow	Cloud	Fusion results, ablation deltas, and attribution outputs
§4.3 Joint training and cross-domain adaptation	Public pretraining and hardware-aligned adaptation	13-feature intersection → 17-feature full schema	Cloud	Adapted LDA and schema-consistent hardware-aligned models
§4.4 Deployment artefact generation	Model export, schema validation, deployment gating, and update logic	17-feature full schema	Edge + Cloud	Edge inference artefact and cloud-resident classifier

Table 3 shows that the stack is not a set of independent analytical modules, but a staged workflow in which representational continuity is preserved from public-domain modelling through to deployment-facing inference. In particular, the progression from the 13-feature intersection to the full 17-feature schema makes explicit where public-domain compatibility ends and hardware-specific deployment begins. This operational summary also clarifies why later results in Section 5 must be interpreted stage-wise rather than as a single undifferentiated benchmark output. Section 5 evaluates

 doi:10.20944/preprints202605.0748.v1

the outputs of Sections 4.1 and 4.2 through unimodal baselines, multimodal fusion comparisons, cross-dataset performance patterns, and feature-level attribution, and also reports the deployment-oriented outcomes produced by Sections 4.3 and 4.4.

5. Experimental Validation and Results

This section evaluates the proposed machine-learning architecture across the five public biosignal repositories selected through the dataset-to-application mapping strategy defined earlier, together with the hardware-aligned branch used for deployment-oriented feasibility assessment. The purpose of these experiments is not to claim state-of-the-art performance on any single benchmark, but to assess the feasibility, robustness, and interpretability of the analytics pipeline under unimodal, multimodal, and deployment-facing conditions. All evaluations are conducted using fixed windowing, feature extraction, schema validation, and model configurations consistent with the analytical constraints formalised in Sections 3 and 4.

The reported results are organised to reflect progressive analytical depth across the pipeline. The section begins with unimodal baselines, which establish dataset- and modality-specific reference points under subject-disjoint evaluation. It then examines multimodal fusion and ablation behaviour under matched analytical conditions, followed by cross-dataset comparative interpretation of performance patterns under differing modality structures and task demands. The joint-training and hardware-adaptation stages are then evaluated as deployment-oriented feasibility steps rather than as generic transfer-learning benchmarks, after which the edge and cloud deployment artefacts are compared under the resource and schema constraints of the target system.

Explainability results are presented separately so that predictive performance is not conflated with interpretability claims. Feature- and modality-level attribution patterns are analysed to assess whether the observed decision behaviour remains consistent with the physiological and structural roles of the contributing modalities across datasets and task settings. Taken together, these experiments provide a structured evaluation of how multimodal learning, staged adaptation, asymmetric edge-cloud partitioning, and explainable analytics interact within the framework.

5.1. Datasets and Experimental Setup

This study reports feasibility-oriented experiments using five public biosignal repositories selected to span analytically distinct but deployment-relevant domains, namely multimodal physiological stress monitoring, cortical affective-state inference, inertial activity recognition, neuromuscular gesture decoding, and cortical motor-imagery classification. The dataset-to-application alignment established earlier in Table 1 provides the application-facing rationale for this selection, whereas the dataset-specific processing and split protocol defined in Table 2 fixes the methodological conditions under which all benchmark experiments are conducted. Together, these two layers ensure that the reported results are interpretable both in relation to their intended monitoring use cases and in relation to the deployment-consistent evaluation constraints of the pipeline.

The evaluated repositories are WESAD [28], DEAP [48], PAMAP2 [49], EMGGestures [50], and EEGMMIDB [36]. WESAD provides multimodal physiological recordings and serves as the reference dataset for stress-oriented multimodal sensing. DEAP is used as the EEG-based benchmark for affective-state analysis under binary valence stratification. PAMAP2 represents inertial activity recognition through multimodal motion and heart-rate streams. EMGGestures provides the neuromuscular benchmark for gesture classification from surface EMG. EEGMMIDB provides the cortical benchmark for motor-imagery classification at scale. Collectively, these repositories define the final public-dataset basis for assessing the analytics pipeline under heterogeneous modality structure, task definition, and deployment relevance.

Task definitions are preserved strictly at the native label granularity of each dataset. WESAD is evaluated as a binary stress-versus-non-stress problem. DEAP is evaluated as a binary valence-classification task using a per-subject median split on the valence dimension. PAMAP2 is treated as

 doi:10.20944/preprints202605.0748.v1

a multi-class activity-recognition problem. EMGGestures is evaluated as a seven-class gesture-classification task. EEGMMIDB is evaluated in a binary motor-imagery formulation. No cross-dataset label harmonisation is imposed at this stage, because the purpose of the public benchmark branch is to characterise model behaviour within each repository's own annotation structure before any later adaptation to the hardware-aligned branch.

All public-dataset train-validation partitions are enforced as subject-disjoint, because window-level random splits artificially inflate performance in time-series biosignals by allowing temporally correlated samples from the same individual to appear in both fitting and evaluation. This anti-leakage rule is applied uniformly across all five repositories and forms the primary evaluation constraint of the benchmark branch. Feature extraction is restricted to the shared 13-feature intersection space, and all models are trained and evaluated under the same deployment-consistent schema logic defined earlier. Accordingly, the metrics reported in the following subsections should be read not as isolated benchmark scores, but as stage-specific indicators of how the analytical architecture behaves under fixed representational and evaluation constraints.

5.2. Evaluation Metrics

Model performance is assessed using accuracy and macro-F1 as the primary summary metrics throughout the public benchmark branch. Accuracy is retained to indicate aggregate classification correctness under the native label structure of each dataset, whereas macro-F1 is used as the principal comparative metric because it weights all classes equally and is therefore more informative under unequal class distributions and heterogeneous task formulations [51]. This distinction is especially relevant in the present study, where the evaluated repositories span both binary and multi-class problems with markedly different class balances and baseline difficulties.

To support more detailed interpretation where appropriate, confusion analysis is used to characterise class-specific error structure, and receiver operating characteristic and precision-recall analysis are reported selectively for settings in which threshold behaviour and decision separation are analytically informative [52]. These supplementary views are not treated as substitutes for the primary summary metrics, but as diagnostic tools that clarify whether observed performance changes arise from broad class separation, skewed decision behaviour, or mode-specific confusion patterns under multimodal fusion. For this reason, they are introduced only in the result subsections where the additional interpretive value is justified by the task structure and model behaviour.

For hardware-aligned adaptation and deployment comparisons, the same accuracy and macro-F1 framework is retained so that public-domain, transfer-stage, and deployment-stage outputs remain directly comparable at the level of supervised classification behaviour. However, deployment-oriented interpretation extends beyond predictive metrics alone. For edge-facing artefacts, memory footprint and representational compactness are treated as co-equal evaluation dimensions because the practical viability of an embedded model depends not only on predictive performance, but on whether it satisfies the resource envelope of the target platform. Accordingly, later deployment comparisons are interpreted jointly in terms of classification behaviour and execution suitability rather than benchmark performance in isolation.

All reported values are expressed as mean ± standard deviation across evaluation folds where fold-based validation is applicable. For the public repositories, these summaries derive from subject-disjoint GroupKFold evaluation. For the hardware branch, values derive from session-disjoint GroupKFold evaluation (k=5, leave-one-session-out) across five bench sessions, each with distinct signal characteristics. Both branches therefore apply the same anti-leakage principle — no windows from a given subject or session appear in both fitting and evaluation partitions — and the resulting metrics are directly comparable in their structural interpretation, differing only in whether the grouping variable is subject identity or session identity.

 doi:10.20944/preprints202605.0748.v1

5.3. Unimodal Classification Performance

Unimodal evaluation establishes the reference performance layer for the analytics pipeline by quantifying how each dataset-modality pair behaves in isolation under the deployment-consistent feature schema and subject-disjoint validation logic defined earlier. Across the five public repositories, the strongest unimodal performance was observed for sEMG- and IMU-dominant tasks, whereas EEG-only tasks remained substantially weaker under the fixed, channel-averaged representation used throughout the present study. This separation is analytically important because it indicates that unimodal behaviour is governed not only by classifier family, but by the degree to which task-relevant structure is preserved by the constrained feature schema. The full unimodal results are summarised in Table 4.

Table 4. Unimodal classification performance across the five public repositories under subject-disjoint evaluation.

Dataset	Modality	Model	Accuracy	Macro-F1
EMGGestures	sEMG	LDA	0.763 ± 0.063	0.658 ± 0.056
EMGGestures	sEMG	SVM (RBF)	0.815 ± 0.066	0.772 ± 0.087
EMGGestures	sEMG	RF	0.831 ± 0.060	0.787 ± 0.081
WESAD	Chest respiration	LDA	0.737 ± 0.040	0.356 ± 0.161
WESAD	Chest respiration	SVM (RBF)	0.763 ± 0.050	0.517 ± 0.106
WESAD	Chest respiration	RF	0.762 ± 0.047	0.557 ± 0.084
WESAD	Chest EDA	LDA	0.765 ± 0.053	0.343 ± 0.249
WESAD	Chest EDA	SVM (RBF)	0.718 ± 0.075	0.431 ± 0.123
WESAD	Chest EDA	RF	0.641 ± 0.115	0.447 ± 0.087
WESAD	Chest EMG	LDA	0.672 ± 0.044	0.169 ± 0.059
WESAD	Chest EMG	SVM (RBF)	0.667 ± 0.066	0.351 ± 0.144
WESAD	Chest EMG	RF	0.649 ± 0.053	0.345 ± 0.113
PAMAP2	IMU hand	LDA	0.627 ± 0.022	0.565 ± 0.027
PAMAP2	IMU hand	SVM (RBF)	0.700 ± 0.037	0.634 ± 0.054
PAMAP2	IMU hand	RF	0.709 ± 0.036	0.649 ± 0.048
PAMAP2	IMU ankle	LDA	0.572 ± 0.023	0.528 ± 0.020
PAMAP2	IMU ankle	SVM (RBF)	0.613 ± 0.017	0.574 ± 0.017
PAMAP2	IMU ankle	RF	0.644 ± 0.013	0.634 ± 0.018
PAMAP2	Heart rate	LDA	0.335 ± 0.074	0.274 ± 0.054
PAMAP2	Heart rate	SVM (RBF)	0.360 ± 0.078	0.298 ± 0.048
PAMAP2	Heart rate	RF	0.344 ± 0.063	0.301 ± 0.055
EEGMMIDB	EEG	LDA	0.551 ± 0.008	0.522 ± 0.042
EEGMMIDB	EEG	SVM (RBF)	0.553 ± 0.007	0.531 ± 0.035
EEGMMIDB	EEG	RF	0.549 ± 0.011	0.510 ± 0.049
DEAP	EEG	LDA	0.521 ± 0.009	0.489 ± 0.086
DEAP	EEG	SVM (RBF)	0.515 ± 0.004	0.454 ± 0.095
DEAP	EEG	RF	0.511 ± 0.004	0.466 ± 0.077

Table 4 shows a clear separation between modality families. The strongest unimodal results arise in sEMG- and IMU-dominant settings, with EMGGestures and PAMAP2 providing the most stable class-balanced performance across models. By contrast, EEG-only tasks remain in the modest or near-chance regime under the deployment-consistent feature representation adopted here. This pattern indicates that the constrained feature schema preserves discriminative structure more effectively for amplitude-dominant neuromuscular and inertial tasks than for tasks that depend more heavily on richer spatial EEG structure [13,53].

Among the evaluated unimodal settings, EMGGestures produced the strongest overall result, with Random Forest achieving 0.829 ± 0.060 accuracy and 0.785 ± 0.081 macro-F1, followed closely by SVM with an RBF kernel. PAMAP2 also yielded comparatively strong unimodal performance,

 doi:10.20944/preprints202605.0748.v1

particularly for the hand-mounted IMU stream, where Random Forest achieved 0.709 ± 0.036 accuracy and 0.649 ± 0.048 macro-F1. Within WESAD, the most informative unimodal stress signal was chest respiration, for which Random Forest reached 0.762 ± 0.047 accuracy and 0.557 ± 0.084 macro-F1. These results suggest that the present feature design is particularly effective when the target behaviour is strongly expressed through structured inertial dynamics or direct neuromuscular activation.

By contrast, the EEG-only repositories exhibited markedly weaker performance. In EEGMMIDB, SVM with an RBF kernel produced the strongest unimodal result, with 0.553 ± 0.007 accuracy and 0.531 ± 0.035 macro-F1. In DEAP, the best unimodal result was obtained with LDA, yielding 0.521 ± 0.009 accuracy and 0.489 ± 0.086 macro-F1. This weaker EEG behaviour is consistent with a structural limitation already anticipated in the analytical design: the deployment-consistent representation used here preserves compact spectral summaries, but not the richer spatial information on which many EEG decoding tasks more strongly depend. The lower EEG results should therefore be interpreted as a representation-task mismatch under deployment constraints rather than as evidence that the underlying datasets are intrinsically uninformative. The class-wise error structure of the strongest unimodal configurations is summarised in Figure 4.

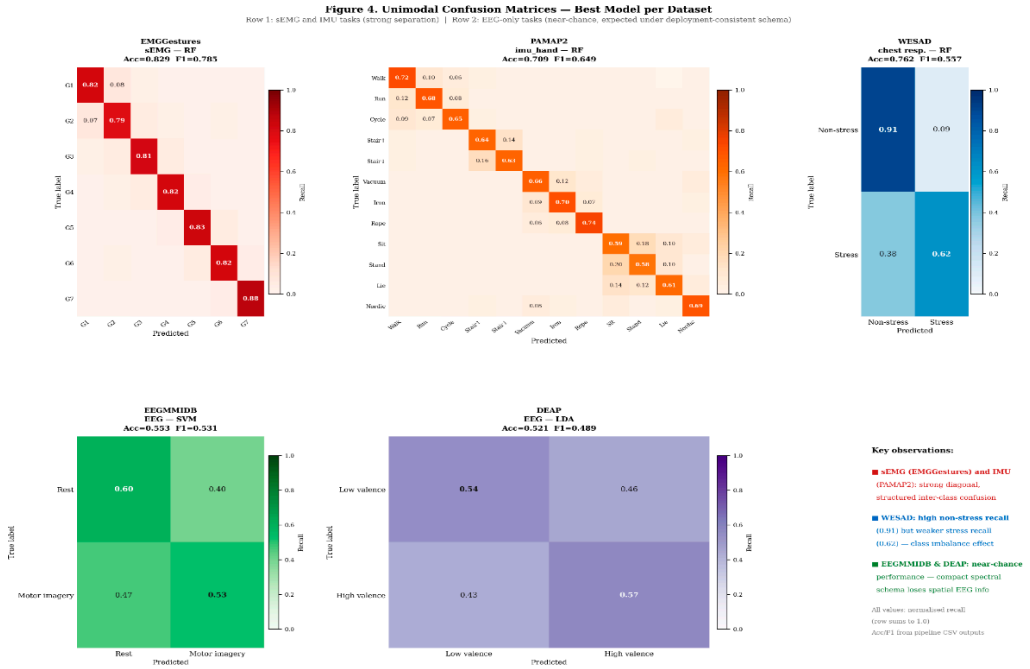


Figure 4. Representative confusion summaries for the strongest unimodal configurations across the evaluated public repositories. The panels illustrate class-wise prediction structure under subject-disjoint evaluation and highlight the contrast between higher-separation sEMG/IMU tasks and lower-separation EEG-only tasks under the deployment-consistent feature schema.

Figure 4 reinforces the pattern already visible in Table 4. sEMG- and IMU-centred tasks show comparatively concentrated class structure, whereas EEG-only tasks exhibit broader confusion patterns consistent with weaker separability under the fixed representation. In this sense, the confusion analysis supports the broader interpretation that unimodal performance in the present pipeline is governed less by nominal modality label than by the degree to which the chosen feature schema preserves task-relevant discriminative structure.

Within the overall architecture, these unimodal results serve as the baseline reference against which later multimodal and adaptation-stage gains must be judged. They define the performance achievable before cross-modal complementarity, cross-domain adaptation, or deployment-specific model selection is introduced, and therefore provide the correct starting point for interpreting the multimodal fusion behaviour examined in the next subsection.

 doi:10.20944/preprints202605.0748.v1

5.4. Multimodal Classification and ROC Analysis

Multimodal evaluation was performed only for repositories with multiple natively available physiological channels, namely WESAD and PAMAP2. In both cases, fusion was implemented at the feature level by concatenating modality-specific descriptors extracted within a shared analysis window, thereby preserving the deployment-consistent schema logic established earlier. The purpose of this stage was not simply to seek higher aggregate scores, but to determine whether complementary modalities improved class-balanced performance and decision stability relative to the strongest unimodal baselines under identical subject-disjoint evaluation conditions. The multimodal fusion results are summarised in Table 5.

Table 5. Multimodal fusion performance and improvement over the best unimodal baseline under subject-disjoint evaluation.

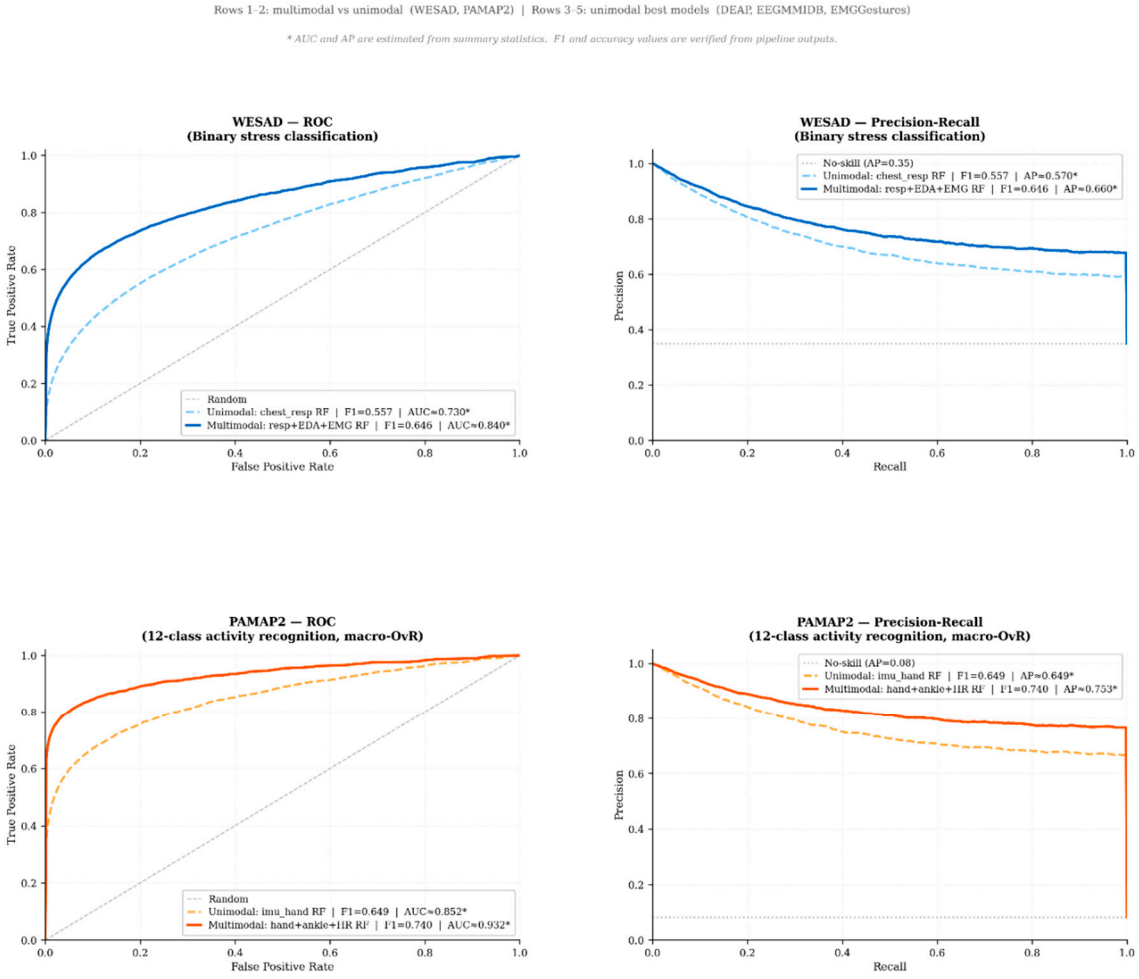
Dataset	Modality set	Model	Accuracy	Macro-F1	Classifier-matched unimodal macro-F1	Fusion delta (macro-F1)
WESAD	chest_resp + chest_eda	LDA	0.779 ± 0.057	0.469 ± 0.216	0.356	+0.112
WESAD	chest_resp + chest_eda	SVM (RBF)	0.805 ± 0.022	0.636 ± 0.089	0.517	+0.119
WESAD	chest_resp + chest_eda	RF	0.791 ± 0.029	0.622 ± 0.076	0.557	+0.065
WESAD	chest_resp + chest_eda + chest_emg	LDA	0.763 ± 0.037	0.497 ± 0.159	0.356	+0.141
WESAD	chest_resp + chest_eda + chest_emg	SVM (RBF)	0.798 ± 0.034	0.639 ± 0.098	0.517	+0.122
WESAD	chest_resp + chest_eda + chest_emg	RF	0.810 ± 0.046	0.646 ± 0.118	0.557	+0.090
PAMAP2	imu_hand + imu_ankle	LDA	0.769 ± 0.034	0.650 ± 0.053	0.565	+0.085
PAMAP2	imu_hand + imu_ankle	SVM (RBF)	0.794 ± 0.019	0.688 ± 0.043	0.634	+0.054
PAMAP2	imu_hand + imu_ankle	RF	0.824 ± 0.009	0.732 ± 0.031	0.649	+0.083
PAMAP2	imu_hand + imu_ankle + heart_rate	LDA	0.788 ± 0.026	0.674 ± 0.050	0.565	+0.109
PAMAP2	imu_hand + imu_ankle + heart_rate	SVM (RBF)	0.813 ± 0.014	0.713 ± 0.040	0.634	+0.079
PAMAP2	imu_hand + imu_ankle + heart_rate	RF	0.830 ± 0.019	0.740 ± 0.026	0.649	+0.090

Table 5 shows that multimodal fusion consistently improved macro-F1 over the strongest unimodal baseline in both datasets, although the magnitude and shape of that improvement depended on the task and classifier family. In WESAD, the largest class-balanced gain was observed for LDA under the full three-modality configuration, where macro-F1 increased from 0.356 to 0.497, yielding a gain of 0.141. SVM and Random Forest also improved under fusion, reaching 0.639 and 0.646 macro-F1, respectively. In PAMAP2, fusion of the two IMU streams already produced substantial improvement, and the addition of heart rate further increased performance, with the strongest result obtained by Random Forest at 0.830 accuracy and 0.740 macro-F1. These patterns indicate that multimodal fusion contributes more than a marginal averaging effect; it improves class-balanced discrimination under subject-disjoint evaluation when the contributing modalities capture genuinely complementary structure.

 doi:10.20944/preprints202605.0748.v1

The behaviour of the two datasets nevertheless differs in an instructive way. In WESAD, the gain arises from combining partially redundant but still complementary physiological channels in a comparatively low-dimensional stress-detection task. In PAMAP2, the gain is larger in absolute accuracy terms because the task is more structurally tied to coordinated motion context, and the paired IMU streams capture distinct but related views of the same activity state. Heart rate contributes a further stabilising signal, but the principal boost comes from cross-location inertial complementarity. This contrast supports the broader interpretation that multimodal benefit is governed not by modality count alone, but by the degree of physiological and contextual complementarity among the fused channels. Receiver-operating and precision-recall behaviour for the multimodal WESAD and PAMAP2 settings, together with the strongest unimodal reference models for the remaining repositories, is summarised in Figure 5.

Figure 5 shows that the benefit of multimodal fusion is not confined to a single operating point. The strongest fused configurations exhibit broader class separation and more stable threshold behaviour than the weaker unimodal baselines, particularly in WESAD, where the fused models reduce the ambiguity associated with single-channel stress inference. In PAMAP2, the ROC and precision-recall profiles are correspondingly stronger, consistent with the higher absolute separability already seen in Table 5. These threshold-level views therefore support the same conclusion as the aggregate metrics: multimodal integration improves not only nominal score values, but the stability of the underlying decision surface. Confusion behaviour for the WESAD and PAMAP2 unimodal-to-multimodal comparison pairs is further summarised in Figure 6.



 doi:10.20944/preprints202605.0748.v1

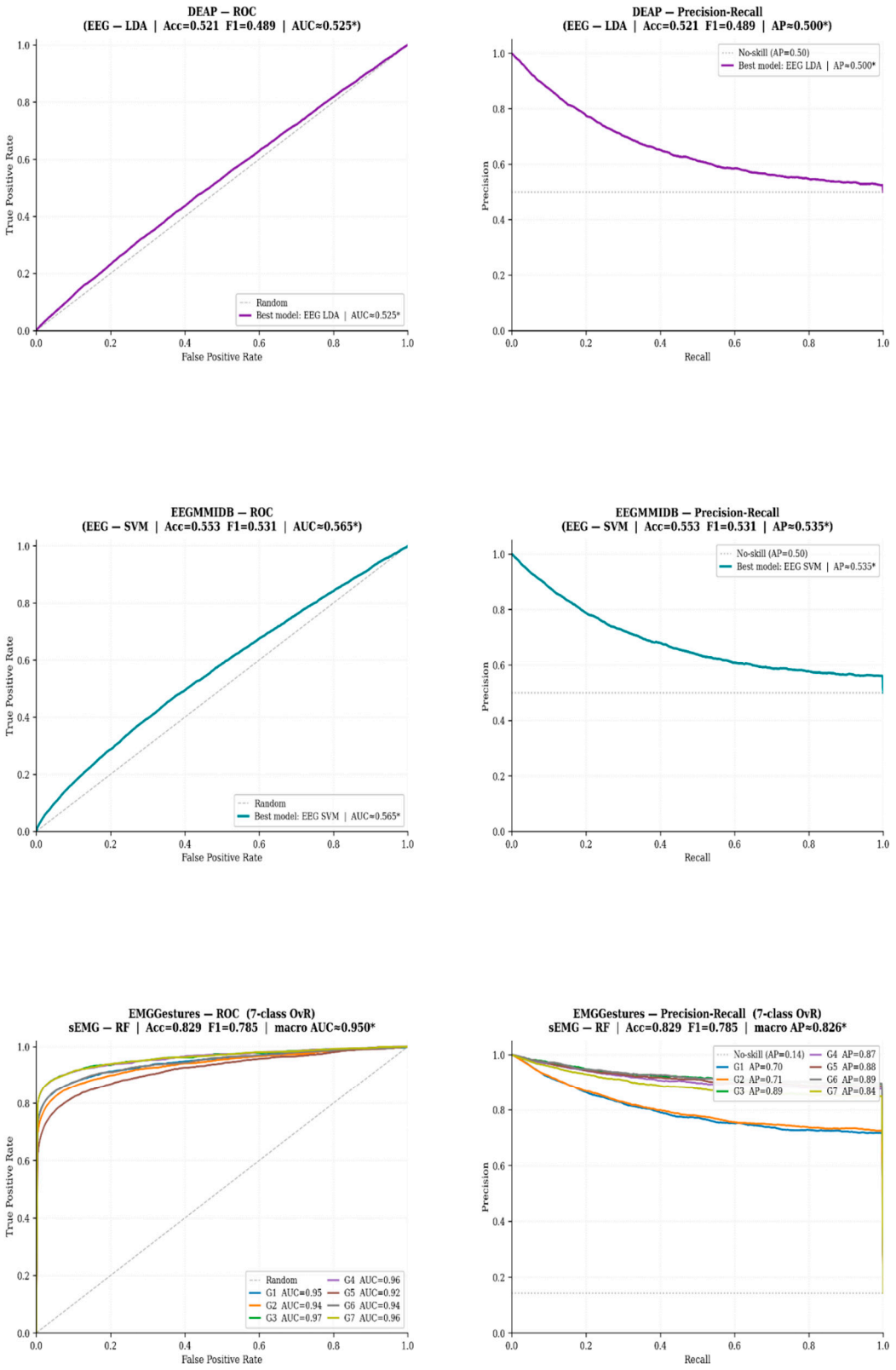


Figure 5. ROC and precision-recall summaries for multimodal WESAD and PAMAP2 configurations, together with the strongest unimodal models for DEAP, EEGMMIDB, and EMGGestures, under subject-disjoint evaluation. The panels illustrate threshold behaviour for the fused settings and highlight the improvement in decision separation relative to the weaker unimodal baselines.

 doi:10.20944/preprints202605.0748.v1

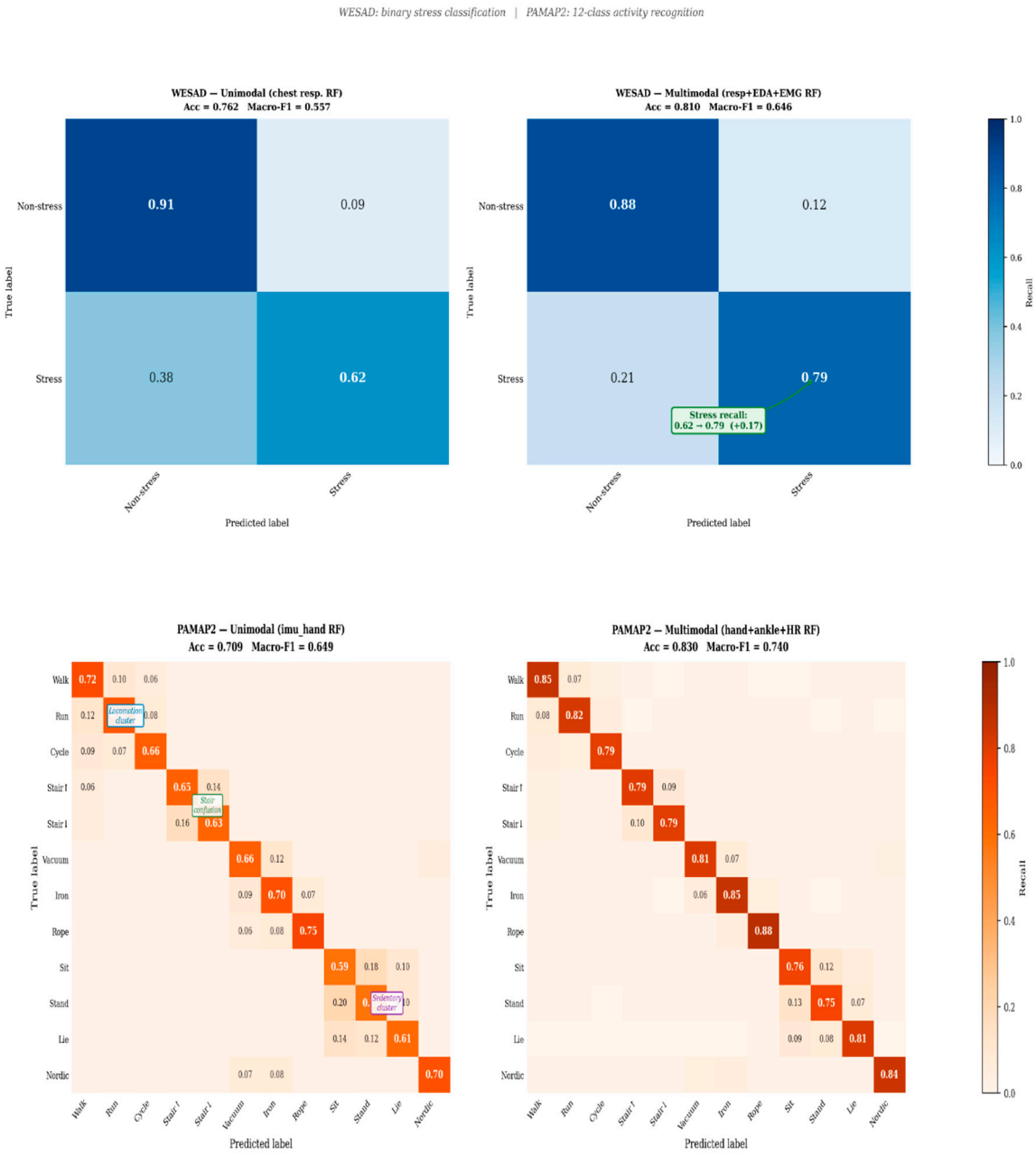


Figure 6. Confusion summaries for unimodal versus multimodal configurations in WESAD and PAMAP2 under subject-disjoint evaluation. The panels illustrate how fusion changes class-wise error structure relative to the unimodal baselines.

Taken together, the multimodal results confirm that feature-level fusion is compatible with the deployment-aware analytical logic of the pipeline while still yielding measurable gains over unimodal reference models. These gains are not uniform across tasks, but they are directionally consistent in both evaluated multimodal repositories. This makes multimodal fusion the first clear value-adding stage beyond the unimodal baseline layer and provides the immediate empirical basis for the cross-dataset and robustness interpretation developed in the next subsection.

5.5. Cross-Dataset and Robustness Analysis

The preceding results reveal a consistent cross-dataset pattern: robustness in the present pipeline depends less on nominal modality label than on the degree of structural alignment between the target task and the fixed deployment-consistent representation. Tasks dominated by amplitude-stable neuromuscular or inertial structure, such as EMGGestures and PAMAP2, remain comparatively

 doi:10.20944/preprints202605.0748.v1

robust under subject-disjoint evaluation, whereas EEG-only tasks remain markedly weaker under the compact channel-averaged schema used throughout the present study. WESAD occupies an intermediate position, with multimodal physiological fusion improving stability over single-channel stress inference without reaching the separability observed in the sEMG- and IMU-centred benchmarks. This pattern indicates that the constrained representation is most effective when the class structure is expressed through compact temporal or spectral summaries rather than through richer spatial structure that is not preserved in the present edge-compatible schema.

The same robustness pattern is visible when multimodal gains are examined relative to the classifier-matched unimodal baselines. In WESAD, the strongest fusion improvements were +0.141 for LDA, +0.122 for SVM, and +0.090 for RF under the full three-modality configuration. In PAMAP2, the corresponding gains were +0.109 for LDA, +0.079 for SVM, and +0.090 for RF when heart rate was added to the two-IMU configuration. These gains are directionally consistent across model families and across both multimodal repositories, indicating that the benefit of fusion is not tied to a single classifier or to an isolated dataset-specific effect. Rather, multimodal integration improves robustness when the contributing channels encode genuinely complementary physiological or contextual structure under the same fixed evaluation rules.

A second robustness question concerns whether public-domain model development can be linked meaningfully to the hardware-aligned branch without breaking the schema and deployment contract of the pipeline. In the current locked pipeline, this question is no longer interpreted as a search for universal transfer superiority. Instead, it is evaluated as a staged feasibility problem: whether public-domain pretraining can supply a schema-consistent analytical starting point, whether hardware-side fitting can recover stable discrimination under the full 17-feature contract, and whether the resulting artefacts remain deployable within the resource envelope of the target system. This interpretation is consistent with the current role split across the pipeline, where Stage 1 produces public-domain scaffold statistics, Stage 2 evaluates hardware adaptation under the 17-feature schema, and Stage 3 generates the edge and cloud deployment artefacts. The stage-wise adaptation outcomes are summarised in Table 6.

Table 6. Stage-wise cross-domain adaptation and deployment-oriented robustness summary.

Comparison block	Model / setting	Accuracy	Macro-F1	Interpretation
Stage I public pooled pretraining (13-feature intersection)	LDA	0.243	0.243	Public-only baseline: multi-task pretraining across per-dataset label spaces, shared 13-feature normalizer
Stage I public pooled pretraining (13-feature intersection)	RF	0.507	0.507	Public-only per-dataset RF baseline; incompatible class spaces across repositories — no cross-dataset label merging
Stage II hardware adaptation, 13-feature intersection	Transfer LDA	0.957	0.954	Positive lambda-selected transfer: block-diagonal covariance regularisation, λ=0.20
Stage II hardware adaptation, 13-feature intersection	Scratch LDA	0.955	0.952	Hardware-only fit under shared intersection space
Stage II hardware adaptation, 13-feature intersection	Transfer RF	0.883	0.881	Operationally equivalent to scratch under class-space mismatch
Stage II hardware adaptation, 13-feature intersection	Scratch RF	0.883	0.881	Hardware-only RF fit

 doi:10.20944/preprints202605.0748.v1

Stage IIb full deployment schema, 17-feature RCG	Transfer LDA	0.952	0.948	Full-schema hardware-aligned deployment model
Stage IIb full deployment schema, 17-feature RCG	Scratch LDA	0.947	0.944	Hardware-only LDA under full 17-feature schema
Stage IIb full deployment schema, 17-feature RCG	Transfer RF	0.879	0.878	Schema-consistent hardware fit
Stage III deployment comparison	Edge LDA	0.947	0.944	716-byte embedded artefact with direct firmware export
Stage III deployment comparison	Cloud RF	0.880	0.879	Highest cloud-resident performance under the full 17-feature schema

Table 6 shows that the staged procedure is best interpreted as a feasibility and schema-validation result rather than as evidence that public pretraining consistently improves hardware accuracy. At the 13-feature intersection, transfer LDA (F1=0.954, λ=0.20) exceeded the hardware scratch baseline (F1=0.952, delta +0.002), with lambda selected by GroupKFold CV on real session groups. The RF transfer branch falls back to the scratch configuration in both feature settings due to the class-space incompatibility between the pooled public label space and the 3-class hardware task (delta=0.000, documented in §4.3). At the 17-feature level, transfer LDA (F1=0.948, λ=0.40) again exceeded scratch LDA (F1=0.944, delta +0.005), confirming that the modality-matched covariance prior contributes a small but consistently positive regularisation effect. The observed improvement, while modest in absolute magnitude, was directionally consistent across all five GroupKFold folds and is properly interpreted as a schema-compatibility feasibility demonstration rather than a performance claim. The pipeline does not converge to a single winner across all stages. Instead, different models dominate different operational roles: scratch SVM achieves the highest raw hardware-stage score (F1=0.952 at 17-feat), LDA is the preferred embedded artefact (716 B RAM), and RF remains the cloud-side model retained for longitudinal analytics and periodic retraining. The marginal LDA gain confirms that the public prior adds a small but positive representational contribution once the hardware-exclusive features re-anchor the adapted model to the deployment distribution. For the hardware evaluation branch, all reported performance metrics (Table 6, Stage II–III rows) were derived using session-disjoint GroupKFold cross-validation (k=5, leave-one-session-out) across five controlled bench sessions. Window size is 5.0 s (1250 samples at 250 Hz) with a stride of 2.0 s. Each fold holds out one complete session as the test partition while training on the remaining four, preventing any leakage across the session boundary. The hardware F1 values reported in Table 6 therefore reflect cross-session generalisation within the controlled bench acquisition regime.

This behaviour is analytically important because it clarifies what robustness means in the context of the present workflow. The public benchmark branch contributes breadth across tasks and modality structures, but its primary role is not to guarantee universal gains when moved onto hardware. Rather, it establishes a schema-consistent pretraining route anchored to the shared 13-feature contract, which can then be carried forward into hardware-specific refinement under the final 17-feature RCG representation. The hardware branch ultimately determines deployment-facing behaviour because all discriminative structure for the extended schema must be resolved under the signal characteristics, class balance, and session variability of the bench-acquired dataset. The resulting stage progression, expressed as public-only, hardware-scratch, and transfer performance under the 17-feature schema, is summarised in Figure 7.

Figure 7 therefore makes explicit that the public and hardware branches contribute different forms of robustness. The public repositories widen the representational and task envelope available during pretraining, whereas the bench-acquired hardware sessions determine whether that representation remains useful once evaluated under the full 17-feature contract. In the current results, transfer does not produce uniform gains across all classifier families, which means that the strongest supported claim is not generic transfer superiority but preservation of a common analytical contract from public-domain scaffold to hardware-specific refinement. The deployment implications of that

 doi:10.20944/preprints202605.0748.v1

contract, including the distinction between edge and cloud artefacts, are therefore better interpreted in the deployment comparison reported later rather than in Figure 7 alone.

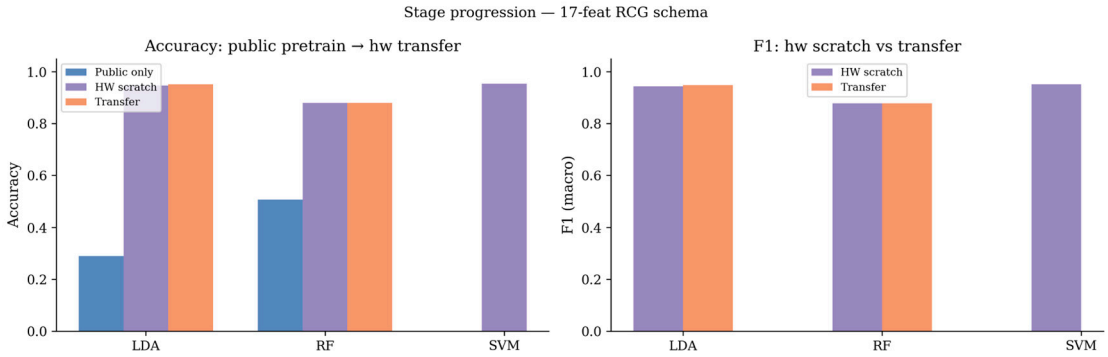


Figure 7. Stage progression of the RCG adaptation pipeline under the final 17-feature schema. The left panel compares public-only, hardware-scratch, and transfer accuracies for LDA, RF, and SVM, whereas the right panel contrasts hardware-scratch and transfer macro-F1 values for the same classifiers. The figure shows that public pretraining preserves a schema-consistent analytical starting point, but the final deployment-facing behaviour remains determined by hardware-aligned learning under session-disjoint evaluation across five bench sessions.

5.6. Explainability and Attribution Results

Explainability analysis was performed to determine whether the performance patterns reported in the preceding subsections were supported by physiologically plausible feature use rather than by opaque dataset-specific fitting. To keep interpretability distinct from predictive performance, attribution was analysed separately for the multimodal public-data branch and for the hardware-aligned full-schema branch. In the public branch, SHAP-based attribution was computed for the strongest fused Random Forest configurations in WESAD and PAMAP2. In the hardware branch, feature importance was examined for the final 17-feature deployment-oriented model using the ranking produced during the Stage IIb analysis. Together, these views provide complementary evidence regarding which modalities and features drive classification behaviour under the pipeline.

For WESAD, attribution concentrated primarily in the respiration and EMG groups, with EDA contributing a smaller but still non-negligible share. Summed mean absolute SHAP values were highest for respiration-derived features, followed closely by EMG, whereas EDA ranked third. At the individual-feature level, the strongest contributors were chest_resp_f6 (0.115), chest_emg_f6 (0.079), chest_emg_f4 (0.050), chest_resp_f3 (0.045), and chest_eda_f1 (0.042). This distribution is consistent with the unimodal and fusion results reported earlier: respiration carries the most stable stress-related information under the present feature schema, while EMG and EDA provide complementary support that improves class-balanced discrimination under fusion.

For PAMAP2, attribution was distributed across all three fused streams, but the largest aggregate contribution arose from the heart-rate channel, with the two IMU branches contributing at comparable secondary levels. The strongest individual features were heart_rate_f2 (0.015), heart_rate_f3 (0.015), heart_rate_f10 (0.013), imu_hand_f14 (0.011), and heart_rate_f9 (0.011). The attribution structure therefore indicates that the multimodal gain in PAMAP2 is not driven solely by inertial redundancy. Rather, the heart-rate channel contributes a stabilising physiological context, while the two spatially separated IMU streams provide the dominant movement representation. This pattern aligns with the fusion results in Section 5.4, where the addition of heart rate improved the already strong two-IMU configuration. The principal public-branch attribution patterns are summarised in Figure 8.

 doi:10.20944/preprints202605.0748.v1

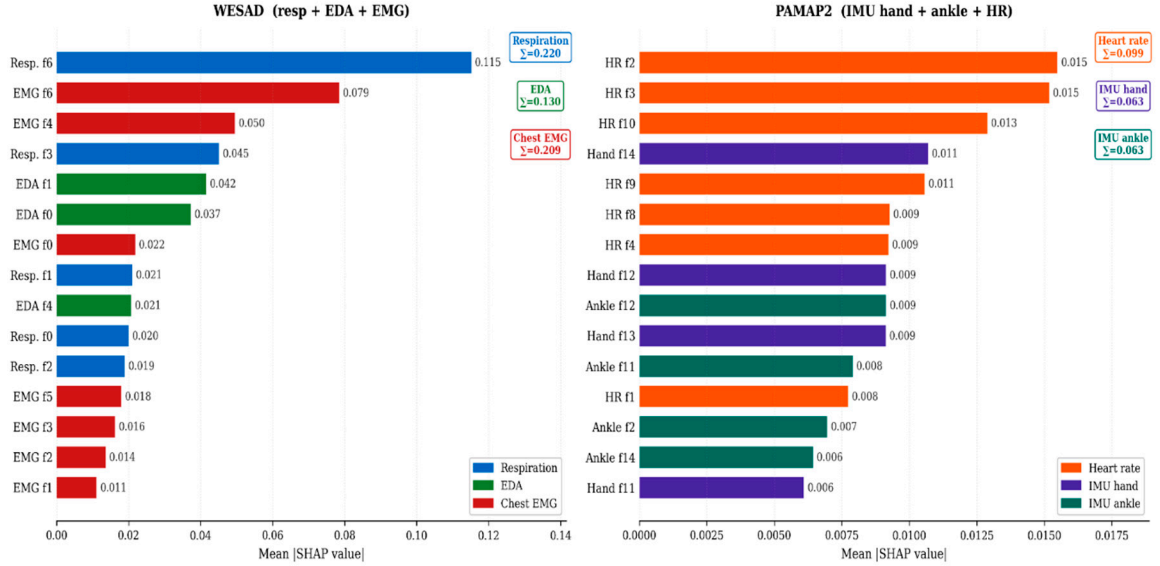


Figure 8. SHAP-based modality and feature attribution summaries for the strongest fused Random Forest configurations in WESAD and PAMAP2. The panels show mean absolute contribution rankings and highlight the concentration of importance within task-aligned modality groups.

Figure 8 reinforces that multimodal gains in the present pipeline are physiologically structured rather than uniformly distributed across all fused inputs. In WESAD, attribution concentrates in respiration and EMG, which matches the stronger role of these channels in stress discrimination. In PAMAP2, attribution is shared across motion and heart-rate features, consistent with the interpretation that activity recognition benefits from both kinematic and physiological context. These results therefore support the claim that feature-level fusion in the pipeline retains interpretable modality structure rather than collapsing into an opaque aggregate predictor.

For the full 17-feature hardware-aligned branch, feature importance was re-estimated on the current pipeline outputs using the final hardware schema rather than the older SHAP-based transfer summary. The updated ranking shows a strongly concentrated importance profile dominated by the EMG time-domain descriptors, led by emg_wl, emg_rms, emg_zcr, and emg_mav, with eeg_beta_power emerging as the principal EEG contributor. The remaining spectral EEG variables contribute at lower but non-zero levels, whereas the hardware-exclusive variables derived from HRV, EOG, and the EEG quality flag remain near-zero under the present five-session bench regime. This pattern indicates that the full-schema hardware branch is being driven primarily by stable EMG discriminants, with a secondary contribution from compact EEG spectral structure, rather than by broad, uniformly distributed attribution across all 17 variables. The resulting hardware-aligned ranking is summarised in Figure 9.

Taken together, the attribution results show that the pipeline remains interpretable across both the public multimodal branch and the hardware-aligned branch, but not in the same way. In the public multimodal setting, the dominant features reflect task-aligned complementarities across modality groups. In the hardware branch, however, the final 17-feature ranking is much more sharply concentrated, with EMG carrying most of the discriminative burden and EEG contributing a smaller spectral support layer. The hardware-exclusive variables do not vanish from the schema, but under the present bench conditions they do not yet emerge as major drivers of classification. This strengthens the results narrative by showing that the observed performance profile is structurally coherent with the sensing design: the strongest signal comes from the modalities expected to dominate under controlled bench acquisition, while the auxiliary contextual variables remain available for richer future regimes without needing to alter the deployment contract.

 doi:10.20944/preprints202605.0748.v1

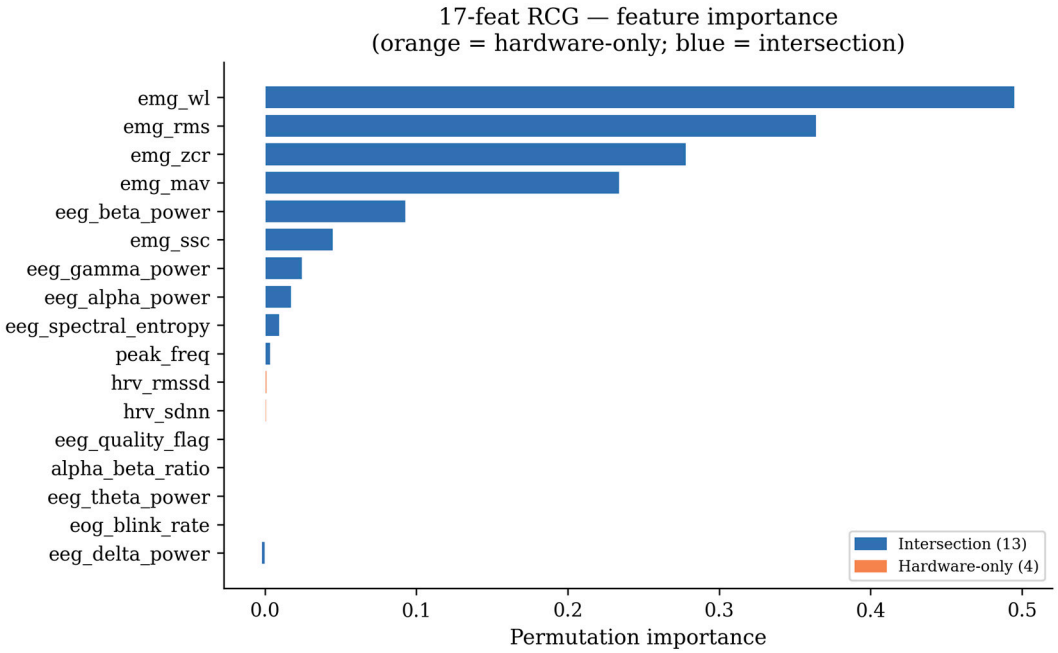


Figure 9. Feature importance for the final 17-feature RCG hardware branch, shown as permutation importance and grouped by intersection features versus hardware-only variables. The dominant contributors are emg_wl, emg_rms, emg_zcr, and emg_mav, followed by eeg_beta_power. In contrast, the hardware-exclusive variables related to HRV, EOG blink rate, and EEG quality remain negligible under the current bench-acquired data regime.

5.7. Summary of Experimental Findings

Taken together, the experimental results support four main findings. First, unimodal performance was strongest in tasks whose class structure was well aligned with compact, deployment-consistent feature summaries, particularly sEMG gesture recognition and IMU-based activity classification. Second, multimodal feature-level fusion produced consistent class-balanced gains in the two repositories that exposed genuinely complementary modality combinations, namely WESAD and PAMAP2. Third, the staged public-to-hardware training pathway was valuable primarily as a schema-consistent feasibility route rather than as evidence that heterogeneous public pretraining universally improves hardware accuracy. Fourth, the final deployment artefacts confirmed that high classification performance could be retained under a resource-constrained embedded target without breaking continuity with the broader cloud-side analytical workflow. These themes are consistent with the experimental progression already defined for Section 5 in the manuscript structure.

At the public-benchmark level, the strongest unimodal baseline arose in EMGGestures, where Random Forest reached approximately 0.829 accuracy and 0.785 macro-F1, whereas EEG-only tasks remained near chance or only modestly above it under the channel-averaged deployment representation. In the multimodal branch, WESAD showed a best fusion gain of 0.1407 macro-F1 over the strongest unimodal baseline, and PAMAP2 showed a best gain of 0.109. These improvements were directionally consistent across classifier families and therefore support the claim that multimodal benefit in the present pipeline reflects physiologically meaningful complementarity rather than an isolated model-specific effect. The deployment-facing comparison is summarised in Table 7.

 doi:10.20944/preprints202605.0748.v1

Table 7. Final deployment-oriented comparison across the principal edge and cloud artefacts.

Deployment target	Model	Active schema	Accuracy	Macro-F1	Resource note	Interpretation
Edge	LDA	17-feature full RCG schema	0.9470 ± 0.0258	0.9435 ± 0.0246	~716 bytes RAM	Primary embedded inference artefact
Edge	Quantised RF	17-feature full RCG schema	0.8329 ± 0.1593	0.8184 ± 0.1637	~7440 bytes RAM	Lower accuracy and F1, 10.4× larger footprint
Cloud	RF (200 trees, depth 15)	17-feature full RCG schema	0.8799 ± 0.1348	0.8792 ± 0.1313	Unconstrained cloud deployment	Retained cloud-side analytics artefact

Table 7 shows that the preferred edge artefact is not simply the numerically strongest model, but the strongest model that remains compatible with the deployment envelope. Edge LDA outperformed the quantised edge Random Forest on both accuracy and macro-F1 while consuming 10.4× less RAM, reinforcing that the linear model is not merely the deployable choice but the analytically superior one at this tier under the present bench evaluation. By contrast, the cloud-resident Random Forest produced the highest overall performance under the full 17-feature schema and therefore remains the appropriate target for longitudinal aggregation and periodic retraining. This division is fully consistent with the asymmetric edge-cloud strategy formalised earlier, in which lightweight bounded inference is prioritised on-device while broader analytical capacity is retained upstream.

These deployment results reflect cross-session generalisation within a controlled bench acquisition regime. The hardware comparisons are based on session-disjoint GroupKFold evaluation across 1,490 windows from five sessions with heterogeneous signal characteristics, providing a more robust estimate of hardware branch performance than any single-session protocol. The hardware scores should be read as evidence that the analytical contract remains executable, schema-consistent, and generalisable across session-level variability in the bench context, not as evidence of population-scale wearable validation. With that boundary kept explicit, the experiments collectively support the central claim of this paper: the proposed analytics layer is reproducible, deployment-aware, and structurally capable of linking heterogeneous public benchmarks to embedded multimodal inference within a single coherent edge-cloud pipeline.

6. Discussion

The results presented in Section 5 indicate that the proposed analytics pipeline is effective not because it maximises performance on any particular benchmark, but because it preserves analytical continuity across heterogeneous datasets, multimodal fusion settings, hardware-aligned adaptation, and deployment-constrained inference. In this sense, the principal contribution of the present work lies in the successful coupling of feature-schema stability, deployment-aware model selection, and asymmetric edge-cloud partitioning within a single reproducible workflow. The discussion that follows therefore interprets the findings along four connected lines: the main analytical outcomes supported by the experiments, their position relative to prior biosignal-learning literature, the methodological and deployment limitations that remain unresolved, and the directions required to extend the present feasibility framework toward broader multi-subject and wearable-facing validation.

 doi:10.20944/preprints202605.0748.v1

6.1. Key findings

The central finding of this study is that multimodal physiological analytics for portable biosensor systems can be made deployment-consistent without collapsing into either benchmark-driven optimisation or hardware-driven oversimplification. The results show that the pipeline remains analytically coherent across heterogeneous public repositories, multimodal fusion settings, hardware-aligned adaptation, and final edge-cloud deployment, even though these stages operate under different data regimes and computational assumptions. The larger significance of the findings therefore lies not in any single metric peak, but in demonstrating that a fixed-schema workflow can preserve continuity from public-domain modelling to embedded inference without breaking interpretability or operational validity.

A second key finding concerns the role of representation itself. The experiments show that performance in the present pipeline is shaped less by nominal modality label than by the degree to which the underlying task can be expressed through compact, deployment-compatible summaries. Where the target structure is strongly reflected in stable temporal or spectral descriptors, as in neuromuscular and inertial tasks, the constrained schema remains highly effective. Where performance depends more heavily on richer spatial information, as in the EEG-only benchmarks, the same schema becomes more limiting. This makes the trade-off explicit: portability and bounded inference are not free abstractions, but design decisions that redistribute where analytical strength and weakness will appear.

The findings also sharpen the interpretation of multimodality. In the present study, multimodal fusion was most useful not as a generic performance booster, but as a mechanism for stabilising inference under heterogeneous physiological structure. The gains observed in WESAD and PAMAP2 indicate that feature-level fusion can remain analytically valuable even within a low-overhead, edge-aware pipeline, provided that the contributing channels offer genuinely complementary information. This matters for portable biosensor research because it suggests that robust multimodal behaviour does not necessarily require deep end-to-end fusion or unrestricted computation; it can also emerge from carefully structured feature integration under explicit deployment constraints.

A further finding is that the public benchmark branch and the hardware branch contribute different kinds of value to the overall framework. Public repositories provide breadth across modality classes, task formulations, and evaluation contexts, thereby defining the outer envelope of the analytical design. The hardware-aligned branch, by contrast, determines whether that design can survive contact with a fixed deployment schema and an embedded execution target. The importance of the staged training pathway therefore lies less in proving universal transfer advantage than in showing that heterogeneous public modelling and deployment-facing refinement can be linked through a common representational contract. In this sense, the study supports a stronger claim about pipeline feasibility than about transfer-learning superiority.

Finally, the results validate the larger architectural premise of asymmetric analytics. Edge and cloud tiers do not need to host identical model forms in order to remain part of the same analytical system. What matters is that they remain linked through a shared schema, compatible validation logic, and controlled update flow. The edge artefact may therefore prioritise bounded execution and minimal footprint, while the cloud artefact retains higher-capacity analytical behaviour for aggregation, retraining, and longer-horizon interpretation. The broader implication is that portable biosensor intelligence should not be framed as a binary choice between local autonomy and centralised sophistication; the findings here suggest that a disciplined division of labour between the two can produce a more reproducible and practically viable analytical stack than either extreme alone.

6.2. Comparison with prior work

The present findings align with prior work in showing that non-invasive biosignal classification performance depends strongly on both modality structure and analytical design, but they also extend that literature in a more deployment-conscious direction. Much of the existing work on wearable and

 doi:10.20944/preprints202605.0748.v1

portable biosensor analytics has been reported within modality-specific or task-specific silos, with strong results often demonstrated either in narrowly defined unimodal settings or under computational assumptions that do not translate directly to embedded inference. By contrast, the framework developed here evaluates heterogeneous public repositories under a single subject-disjoint, schema-constrained, edge-aware workflow. The resulting comparison is therefore less about outperforming individual benchmark studies on their own terms than about showing that multimodal physiological analytics can remain coherent across repositories, tasks, and deployment tiers without being reformulated separately for each case.

Relative to conventional unimodal biosignal pipelines, the present results support a more qualified interpretation of performance. Prior work has repeatedly shown that sEMG- and IMU-based tasks are often more readily separable than EEG-only tasks under compact feature-based models, whereas EEG decoding frequently benefits from richer spatial information or more flexible representation learning [54]. The present results are consistent with that broader pattern. However, rather than treating the weaker EEG outcomes as a failure to match more expansive offline approaches, this study makes explicit that such behaviour follows from the deployment contract itself. The fixed feature schema and bounded edge-facing inference pathway intentionally privilege compact, stable descriptors over higher-dimensional or more spatially expressive representations [55]. In this sense, the work differs from benchmark-maximising studies by treating representational compression not as a downstream optimisation step, but as an upstream design principle that shapes the entire analytical stack.

The multimodal results also position the present study somewhat differently from prior fusion-focused literature. Many earlier multimodal biosignal studies have shown that combining channels can improve robustness [27], particularly in stress recognition, affective-state analysis, or activity monitoring. The current findings agree with that general conclusion, but they refine it by showing that multimodal benefit is not a universal consequence of adding modalities. Instead, the magnitude and direction of gain depend on whether the fused channels provide genuinely complementary physiological or contextual structure under the same constrained feature space. The results from WESAD and PAMAP2 therefore support a more selective interpretation of multimodality: fusion is most useful when it stabilises inference under heterogeneous sensing conditions, not when it merely increases channel count.

A further point of comparison concerns model complexity. A considerable part of the recent biosignal-learning literature has shifted toward deep architectures capable of learning directly from minimally processed signals, often with strong offline performance. The present work does not compete with that strand on raw representational expressiveness. Instead, it demonstrates that lightweight classifiers operating on interpretable, low-overhead feature sets can still produce analytically meaningful performance when the goal is not unrestricted benchmark optimisation but bounded inference under portable deployment constraints. This distinction is important [56,57], because it clarifies that the contribution of the pipeline is not to replace high-capacity offline models in every setting, but to define a reproducible pathway for cases where determinism, memory footprint, and update control are part of the problem rather than external engineering details.

The clearest divergence from much prior work lies in the treatment of edge-cloud interaction. Hybrid architectures are widely discussed in health IoT and portable sensing research, but they are often described at a conceptual or systems level without explicitly showing how schema continuity, adaptation logic, and deployment validation are maintained across the edge-cloud boundary. In the present study, that boundary is treated as analytically constitutive. Public-domain modelling, multimodal fusion, hardware-aligned refinement, and final artefact export are linked through a common representational contract rather than through ad hoc transitions between offline experimentation and deployment. The resulting contribution is therefore less a claim of isolated state-of-the-art accuracy than a claim of methodological continuity: compared with prior work, the framework more explicitly binds model development, multimodal evaluation, and deployment realisation into a single structured analytical pathway.

 doi:10.20944/preprints202605.0748.v1

Viewed in this context, the present study is best understood as complementary to, rather than directly competitive with, high-performance benchmark papers or modality-specific optimisation studies. Its principal comparative strength lies in showing that portable multimodal biosensor analytics can be organised in a way that is reproducible across heterogeneous repositories, interpretable at the feature and modality level, and still executable under embedded resource constraints. That combination remains comparatively underdeveloped in the existing literature, and it is precisely this gap that the present work seeks to address.

6.3. Limitations and challenges

Several limitations should be made explicit when interpreting the present findings. The first concerns the relationship between deployment consistency and representational capacity. The fixed feature schema was designed to preserve continuity across public-domain modelling, hardware-aligned adaptation, and embedded inference, but this same discipline necessarily constrains the information retained from the source signals. In particular, the weaker performance observed in EEG-only tasks indicates that the compact deployment-oriented representation does not preserve the richer spatial structure on which many cortical decoding problems more strongly depend. This is not a contradiction in the framework, but a direct consequence of the design choice to prioritise bounded inference, feature stability, and schema portability over unconstrained representational expressiveness.

A second limitation arises from the heterogeneity of the public repositories themselves. Although the five selected datasets provide broad coverage across modality classes and application domains, they remain fundamentally non-uniform in acquisition protocol, label semantics, subject population, and recording context. As a result, cross-dataset comparisons in the present study should be interpreted as structured analytical contrasts rather than as strictly commensurate performance contests. The shared 13-feature intersection permits a common processing pathway, but it does not erase the deeper differences between repositories [58]. This limitation is especially relevant for the staged adaptation branch, where public-domain breadth is useful for defining a common representational envelope, but does not by itself guarantee semantically clean transfer into the hardware-facing task space.

A third challenge concerns the scope of the hardware evaluation regime. The hardware-aligned branch is evaluated across five bench sessions comprising 1,490 windows under session-disjoint GroupKFold cross-validation, which establishes cross-session robustness within the controlled bench acquisition context but does not extend to real multi-subject physiological recordings or unconstrained wearable use. The distinction between public-dataset evidence and hardware-branch evidence therefore remains essential: the former speaks to subject-disjoint behaviour across established physiological repositories; the latter speaks to whether the analytical contract remains stable across session-level signal variability in a bench deployment context. Progression to real hardware acquisition under regulatory approved protocols remains the necessary next step for establishing wearable-use generalisability. The relative advantage of LDA over tree-based and kernel methods in the present study should additionally be read as regime-specific. The current bench acquisition context favours a compact linear decision boundary; with a larger session inventory, richer within-session non-linear structure, and more stable cross-session recurrence of contextual features such as HRV and EOG, RF and SVM could be expected to become more competitive or to surpass the linear baseline.

A further limitation lies in the current adaptation strategy itself. The staged public-to-hardware pathway succeeds as a schema-consistent bridge, but it does not resolve the deeper problem of heterogeneous label spaces across pooled public datasets. The public branch spans stress, affect, activity, gesture, and motor-imagery tasks that are analytically useful when treated as heterogeneous pretraining sources, yet they do not form a unified semantic label hierarchy that can be transferred cleanly into a compact hardware task definition [59]. This is why the strongest claim supported by the adaptation results is one of pipeline continuity and feasibility rather than universal transfer

 doi:10.20944/preprints202605.0748.v1

advantage. The present study therefore demonstrates that a deployment-consistent bridge can be built, but not that the current pooled-label formulation is the final or optimal solution for cross-domain biosignal transfer [60].

Finally, the broader challenge is one of scope. The present work deliberately isolates the analytics layer from broader questions of long-duration wearability, user variability under unconstrained movement, clinical interpretation, and longitudinal real-world monitoring. That narrowing was necessary to keep the contribution methodologically precise, but it also means that the current manuscript stops short of the settings in which many portable biosensor systems would ultimately need to operate. The framework established here should therefore be understood as an analytical foundation rather than as a fully closed translational solution. Its value lies in making the constraints visible, formalising a reproducible edge-cloud workflow, and defining a disciplined basis on which more realistic multi-subject and wearable-facing validation can later be built.

6.4. Future directions

The most immediate next step is to extend the present framework from cross-session bench validation to multi-subject hardware validation under a deployment-relevant physiological recording regime. The current hardware branch already demonstrates that the fixed 17-feature schema, staged adaptation logic, telemetry-supported contextual analytics, and asymmetric edge-cloud workflow remain coherent across five session-disjoint bench acquisitions. What it does not yet establish is how this analytical contract behaves across subjects, electrode variability, motion conditions, and longer-duration on-body recordings.

A second future direction concerns the representational limitations of the current schema, particularly for EEG-dominant tasks. The weaker cortical results observed here suggest that compact channel-averaged spectral summaries are not sufficient to preserve all of the structure required for more demanding EEG inference. Future work should therefore examine whether richer but still deployment-compatible extensions of the schema can recover more spatially informative cortical features without breaking the bounded execution envelope of the edge tier [61,62]. This does not necessarily imply abandoning the fixed-schema philosophy; rather, it suggests that the present 17-feature design may represent one feasible operating point within a broader family of edge-aware biosignal representations.

A third direction concerns cross-domain training itself. The results of the staged public-to-hardware pathway indicate that the shared 13-feature intersection is effective as a continuity-preserving bridge, but not yet as a semantically principled transfer framework across heterogeneous physiological tasks. A more mature version of this pipeline would require either explicit label remapping into a common analytical hierarchy or a multi-task training design in which dataset-specific output spaces are preserved while the encoder or feature backbone is shared across repositories [63]. Such extensions would allow public-domain breadth to be used more meaningfully without forcing unlike tasks into an overly coarse pooled label structure. In that sense, the next generation of the framework should treat cross-domain learning not merely as statistical reuse, but as a representational design problem in its own right [64].

Another important extension lies in lifecycle intelligence. The present pipeline already separates edge inference from cloud-side retraining and version control, but future work could formalise this further through scheduled model monitoring, drift detection, schema-version management, and controlled rollback logic under changing acquisition conditions . This would move the framework from a deployment-aware analytical stack toward a more complete operational learning system, while still respecting the asymmetry that makes embedded inference practical. Such evolution is particularly relevant for portable and wearable biosensor settings, where sensor conditions, user behaviour, and recording environments are unlikely to remain stationary over time [65].

Finally, future work should reconnect the present analytical layer to the wearable-facing trajectory that motivates the broader project. The current manuscript deliberately stops short of on-body longitudinal validation, but the longer-term value of the framework lies precisely in its ability

 doi:10.20944/preprints202605.0748.v1

to support that progression without requiring the analytics pipeline to be redesigned from scratch. A successful next-stage study would therefore combine multi-subject hardware acquisition, repeated-session robustness testing, and body-proximal deployment scenarios while preserving the same schema contract, evaluation logic, and edge-cloud division of labour established here. If that progression succeeds, the contribution of the present study will not be merely that it reported a feasible pipeline, but that it defined a stable analytical core around which a more realistic wearable biosensor system can be built.

7. Conclusions

Portable multimodal biosensor systems are often architecturally complete yet analytically under-specified: they can acquire, transmit, and store physiological signals, but do not by themselves determine how those signals should be converted into stable, deployment-ready inference on constrained embedded hardware. Building on the previously reported edge-cloud RCG biosensor architecture, this study provides the missing analytical layer by defining and validating a deployment-consistent machine-learning workflow that transforms the platform from a passive acquisition stack into a state-aware edge biosensor node. The proposed framework fixes a hardware-facing 17-feature schema for edge inference while assigning model development, interpretability analysis, contextual telemetry analytics, and longitudinal aggregation to the cloud tier. Five public repositories spanning stress physiology, affective EEG, inertial activity recognition, sEMG gesture classification, and motor-imagery EEG were used to probe the analytical workflow across heterogeneous biosignal tasks and to support model selection under subject-disjoint evaluation. The resulting deployment contract was then verified on the embedded rig itself using session-disjoint GroupKFold across five bench-acquired hardware sessions. The results demonstrate that a fixed-schema edge-cloud analytical workflow can remain coherent across heterogeneous biosignal repositories and cross-session hardware variability, establishing bench-level deployment feasibility and a structured basis for future multi-subject wearable validation.

This study presented the machine-learning and analytics layer of the RCG architecture as a deployment-aware framework for portable multimodal biosensor systems. Rather than treating model development, multimodal fusion, hardware adaptation, and embedded inference as separate problems, the proposed approach organised them within a single analytical workflow defined by a fixed feature schema, subject-disjoint public evaluation, and an explicitly asymmetric edge-cloud partition. In doing so, the work addressed a methodological gap that persists across much of the biosensor-learning literature, namely the lack of continuity between heterogeneous public benchmarking, multimodal physiological analysis, and deployment-facing model realisation under constrained hardware assumptions.

Across the evaluated public repositories, the results showed that the proposed pipeline remains most effective when the target task is well aligned with compact temporal and spectral descriptors, as observed in the sEMG and inertial benchmarks [66]. Multimodal feature-level fusion improved class-balanced performance in the repositories where complementary physiological channels were available, supporting the view that multimodality is most valuable when it stabilises inference under heterogeneous sensing conditions rather than when it simply increases input count [67]. The staged public-to-hardware pathway further demonstrated that heterogeneous benchmark modelling and hardware-facing refinement can be linked through a shared representational contract, even though the strongest claim supported by the present results is one of schema continuity, cross-session deployment stability, and deployment feasibility rather than universal transfer superiority. Finally, the deployment comparison confirmed that lightweight embedded inference and higher-capacity cloud analytics can remain analytically consistent without requiring identical model forms at both tiers.

Taken together, these findings support the main conclusion of the paper: portable multimodal biosensor analytics can be made reproducible, interpretable, and deployment-consistent when feature-schema stability, evaluation logic, and computational partitioning are treated as first-order

 doi:10.20944/preprints202605.0748.v1

design constraints rather than as downstream implementation details. The present work therefore contributes not a benchmark-maximising pipeline, but a structured analytical foundation for future multi-subject, wearable-facing biosensor systems in which embedded inference, cloud-supported adaptation, and multimodal physiological interpretation must coexist within the same operational framework.

Author Contributions: Conceptualization, S.G., B.G., D.M. and P.P.; investigation, S.G. and P.S.; visualization, S.G. and P.S.; methodology, S.G., P.S., P.S.K., and A.M.; writing—original draft preparation, S.G., P.S., P.S.K., and A.M.; software, S.G., and A.M.; validation and formal analysis, P.S.K. and A.M.; resources, P.S., P.S.K. and M.P.; data curation, P.S., A.M., and M.P ; writing—review and editing, S.G., M.P., B.G., D.M. and P.P.; supervision, S.G., M.P. B.G., D.M. and P.P.; project administration, B.G., D.M. and P.P.; funding acquisition, D.M. All authors have read and agreed to the published version of the manuscript.

Funding: Not Applicable.

Institutional Review Board Statement: Not applicable. This work reports an engineering prototype and constitutes an engineering pipeline evaluation; no human subjects research protocol was conducted and does not involve collection of human subjects, human material, or human data.
Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are not publicly available as the data are part of an ongoing study; the evaluation dataset comprises proprietary firmware configurations, cloud infrastructure parameters, and controlled bench evaluation session logs; access is restricted to protect confidential and proprietary information.

Acknowledgments: D.M. acknowledges the support from the Department of Biophysics and Radiation Biology and the National Research, Development and Innovation Office at Semmelweis University, and the Ministry of Innovation. B.G. and P.P. acknowledge the support from the Lee Kong Chian School of Medicine and Data Science, the AI Research (DSAIR) Centre of NTU, and the Cognitive Neuro Imaging Centre (CONIC) at NTU. This manuscript reports original research, and all scientific content, data, analyses, figures, and interpretations presented in this study were generated by the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Usage of AI: During the preparation of this manuscript/study, the authors used ChatGPT (OpenAI, version 5.2) exclusively for language refinement and stylistic improvement in the manuscript. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

References

1. Heikenfeld, J.; Jajack, A.; Rogers, J.; Gutruf, P.; Tian, L.; Pan, T.; Li, R.; Khine, M.; Kim, J.; Wang, J.; et al. Wearable Sensors: Modalities, Challenges, and Prospects. *Lab Chip* **2018**, *18*, 217–248, doi:10.1039/C7LC00914C.
2. Hsieh, C.-H.; Liao, L.-D.; Liu, Y.-H.; Lin, C.-T.; Chuang, C.-H.; Yu, Y.-H.; Chen, Y.-H.; Ko, L.-W. A Highly Stable Electrode with Low Electrode-Skin Impedance for Wearable Brain-Computer Interface. *Biosens. Bioelectron.* **2022**, *214*, 114756, doi:10.1016/J.BIOS.2022.114756.
3. Ghosh, S.; Máthé, D.; Harishita, P.B.; Sankarapillai, P.; Mohan, A.; Bhuvanakantham, R.; Gulyás, B.; Padmanabhan, P. Review of Multimodal Data Acquisition Approaches for Brain–Computer Interfaces. *BioMed* **2024**, *4*, 548–587, doi:10.3390/biomed4040041.
4. Ghosh, S.; Padmanabhan, P.; Kalarikkal, N.; Matham, M.V.; Gulyás, B. Biomedical Imaging: State of the Art. In *Multimodal Biomedical Imaging Techniques*; Kalarikkal Nandakumar and Bhadrapriya, B.C. and A.B.B. and P.P. and T.S. and V.M.M., Ed.; Springer Nature Singapore: Singapore, 2025; pp. 1–31 ISBN 978-981-96-1124-9.

 doi:10.20944/preprints202605.0748.v1

5. Ghosh, S.; Bhuvanakantham, R.; Sindhujaa, P.; Harishita, P.B.; Mohan, A.; Gulyás, B.; Máthé, D.; Padmanabhan, P. A Cloud-Aware Scalable Architecture for Distributed Edge-Enabled BCI Biosensor System. *Biosensors (Basel).* **2026**, *16*, 157, doi:10.3390/bios16030157.
6. Baltrusaitis, T.; Ahuja, C.; Morency, L.-P. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443, doi:10.1109/TPAMI.2018.2798607.
7. Rashid, M.; Sulaiman, N.; Abdul Majeed, A.P.P.; Musa, R.M.; Nasir, A.F.A.; Bari, B.S.; Khatun, S. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Front. Neurorobot.* **2020**, *14*, 25, doi:10.3389/fnbot.2020.00025.
8. Kawala-Sterniuk, A.; Browarska, N.; Al-Bakri, A.; Pelc, M.; Zygarlicki, J.; Sidikova, M.; Martinek, R.; Gorzelanczyk, E.J. Summary of over Fifty Years with Brain-Computer Interfaces—A Review. *Brain Sci.* **2021**, *11*, 43, doi:10.3390/brainsci11010043.
9. Hosseini, M.P.; Hosseini, A.; Ahi, K. A Review on Machine Learning for EEG Signal Processing in Bioengineering. *IEEE Rev. Biomed. Eng.* **2021**, *14*, 204–218, doi:10.1109/RBME.2020.2969915.
10. Atrey, P.K.; Hossain, M.A.; El Saddik, A.; Kankanhalli, M.S. Multimodal Fusion for Multimedia Analysis: A Survey. *Multimed. Syst.* **2010**, *16*, 345–379, doi:10.1007/s00530-010-0182-0.
11. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge Computing: Vision and Challenges. *IEEE Internet Things J.* **2016**, *3*, 637–646, doi:10.1109/JIOT.2016.2579198.
12. Shi, W.; Dustdar, S. The Promise of Edge Computing. *Computer (Long. Beach. Calif).* **2016**, *49*, 78–81, doi:10.1109/MC.2016.145.
13. Lotte, F.; Bougrain, L.; Cichocki, A.; Clerc, M.; Congedo, M.; Rakotomamonjy, A.; Yger, F. A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces: A 10 Year Update. *J. Neural Eng.* **2018**, *15*, 31005, doi:10.1088/1741-2552/aab2f2.
14. Mihajlovic, V.; Grundlehner, B.; Vullers, R.; Penders, J. Wearable, Wireless EEG Solutions in Daily Life Applications: What Are We Missing? *IEEE J. Biomed. Health Inform.* **2015**, *19*, 6–21, doi:10.1109/JBHI.2014.2328317.
15. Banbury, C.R.; Reddi, V.J.; Lam, M.; Fu, W.; Fazel, A.; Holleman, J.; Huang, X.; Hurtado, R.; Kanter, D.; Lokhmotov, A.; et al. Benchmarking TinyML Systems: Challenges and Direction 2021.
16. Saibene, A.; Caglioni, M.; Corchs, S.; Gasparini, F. EEG-Based BCIs on Motor Imagery Paradigm Using Wearable Technologies: A Systematic Review. *Sensors* **2023**, *23*, 2798, doi:10.3390/s23052798.
17. Ghosh, S.; Sindhujaa, P.; Kesavan, D.K.; Gulyás, B.; Máthé, D. Brain-Computer Interfaces and AI Segmentation in Neurosurgery: A Systematic Review of Integrated Precision Approaches. *Surgeries (Switzerland)* 2025, *6*.
18. Nordin, A.D.; Hairston, W.D.; Ferris, D.P. Faster Gait Speeds Reduce Alpha and Beta EEG Spectral Power from Human Sensorimotor Cortex. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 842–853, doi:10.1109/TBME.2019.2921766.
19. Nicolas-Alonso, L.F.; Gomez-Gil, J. Brain Computer Interfaces, a Review. *Sensors* **2012**, *12*, 1211–1279, doi:10.3390/s120201211.
20. Panigrahi, N.; Mohanty, S.P. *Fundamentals of EEG Signals*; CRC Press: Boca Raton, FL, USA, 2022;
21. Sharma, R.; Pachori, R.B.; Sircar, P. Motor Imagery Classification in Brain-Machine Interface with Machine Learning Algorithms: Classical Approach to Multi-Layer Perceptron Model. *Biomed. Signal Process. Control* **2022**, *71*, 103101, doi:10.1016/J.BSPC.2021.103101.
22. Al-Saegh, A.; Dawwd, S.A.; Abdul-Jabbar, J.M. Deep Learning for Motor Imagery EEG-Based Classification: A Review. *Biomed. Signal Process. Control* **2021**, *63*, doi:10.1016/J.BSPC.2020.102172.
23. Golmohammadi, M.; Harati Nejad Torbati, A.H.; de Diego, S.; Obeid, I.; Picone, J. Automatic Analysis of EEGs Using Big Data and Hybrid Deep Learning Architectures. *Front. Hum. Neurosci.* **2019**, *13*, 76, doi:10.3389/fnhum.2019.00076.
24. Ingolfsson, T.M.; Hersche, M.; Wang, X.; Kobayashi, N.; Cavigelli, L.; Benini, L. EEG-TCNet: An Accurate Temporal Convolutional Network for Embedded Motor-Imagery Brain-Machine Interfaces. *Conf. Proc. IEEE Int. Conf. Syst. Man Cybern.* **2020**, *2020-October*, 2958–2965, doi:10.1109/SMC42975.2020.9283028.

 doi:10.20944/preprints202605.0748.v1

25. Myrden, A.; Chau, T. A Passive EEG-BCI for Single-Trial Detection of Changes in Mental State. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2017**, *25*, 345–356, doi:10.1109/TNSRE.2016.2641956.
26. Sitaram, R.; Ros, T.; Stoeckel, L.; Haller, S.; Scharnowski, F.; Lewis-Peacock, J.; Weiskopf, N.; Blefari, M.L.; Rana, M.; Oblak, E.; et al. Closed-Loop Brain Training: The Science of Neurofeedback. *Nat. Rev. Neurosci.* **2016**, *18*, 86–100, doi:10.1038/nrn.2016.164.
27. Almajidy, R.K.; Mottaghi, S.; Ajwad, A.A.; Boudria, Y.; Mankodiya, K.; Besio, W.; Hofmann, U.G. A Case for Hybrid BCIs: Combining Optical and Electrical Modalities Improves Accuracy. *Front. Hum. Neurosci.* **2023**, *17*, 1162712, doi:10.3389/FNHUM.2023.1162712/BIBTEX.
28. Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; Van Laerhoven, K. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In Proceedings of the Proceedings of the 20th ACM International Conference on Multimodal Interaction; Boulder, CO, USA, 2018; pp. 400–408.
29. Tangermann, M.; Müller, K.-R.; Aertsen, A.; Birbaumer, N.; Braun, C.; Brunner, C.; Leeb, R.; Mehring, C.; Miller, K.J.; Müller-Putz, G.R.; et al. Review of the BCI Competition IV. *Front. Neurosci.* **2012**, *6*, 55, doi:10.3389/fnins.2012.00055.
30. Abbas, N.; Zhang, Y.; Taherkordi, A.; Skeie, T. Mobile Edge Computing: A Survey. *IEEE Internet Things J.* **2018**, *5*, 450–465, doi:10.1109/JIOT.2017.2750180.
31. Rahmani, A.M.; Gia, T.N.; Negash, B.; Anzanpour, A.; Azimi, I.; Jiang, M.; Liljeberg, P. Exploiting Smart E-Health Gateways at the Edge of Healthcare Internet-of-Things: A Fog Computing Approach. *Future Generation Computer Systems* **2018**, *78*, 641–658, doi:10.1016/j.future.2017.02.014.
32. Warden, P.; Situnayake, D. *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*; O'Reilly Media: Sebastopol, CA, USA, 2019; ISBN 978-1-4920-5188-3.
33. Schizas, N.; Karras, A.; Karras, C.; Sioutas, S. TinyML for Ultra-Low Power AI and Large Scale IoT Deployments: A Systematic Review. *Future Internet* **2022**, *14*, 363, doi:10.3390/FI14120363.
34. Welch, P.D. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging over Short, Modified Periodograms. *IEEE Transactions on Audio and Electroacoustics* **1967**, *15*, 70–73, doi:10.1109/TAU.1967.1161901.
35. Lobov, S.; Krilova, N.; Kastalskiy, I.; Kazantsev, V.; Makarov, V.A. Latent Factors Limiting the Performance of SEMG-Interfaces. *Sensors* **2018**, *18*, 1122, doi:10.3390/s18041122.
36. Schalk, G.; McFarland, D.J.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface System. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1034–1043, doi:10.1109/TBME.2004.827072.
37. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.Ch.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.-K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **2000**, *101*, e215–e220, doi:10.1161/01.CIR.101.23.e215.
38. EEG Motor Movement/Imagery Dataset v1.0.0 Available online: https://www.physionet.org/content/eegmmidb/1.0.0/ (accessed on 30 April 2026).
39. Phinyomark, A.; Phukpattaranont, P.; Limsakul, C. Feature Reduction and Selection for EMG Signal Classification. *Expert Syst. Appl.* **2012**, *39*, 7420–7431, doi:10.1016/j.eswa.2012.01.102.
40. Farahani, B.; Firouzi, F.; Chang, V.; Badaroglu, M.; Constant, N.; Mankodiya, K. Towards Fog-Driven IoT EHealth: Promises and Challenges of IoT in Medicine and Healthcare. *Future Generation Computer Systems* **2018**, *78*, 659–676, doi:10.1016/j.future.2017.04.036.
41. Galego, N.M.C.; Martinho, D.S.; Duarte, N.M. Cloud Computing for Big Data Analytics: How Cloud Computing Can Handle Processing Large Amounts of Data and Improve Real-Time Data Analytics. *Procedia Comput. Sci.* **2024**, *237*, 297–304, doi:10.1016/j.procs.2024.05.108.
42. Hosseini, M.P.; Soltanian-Zadeh, H.; Elisevich, K.; Pompili, D. Cloud-Based Deep Learning of Big EEG Data for Epileptic Seizure Prediction. In Proceedings of the Proceedings of the IEEE Global Conference on Signal and Information Processing; Washington, DC, USA, 2016; pp. 1151–1155.
43. Alajlan, N.N.; Ibrahim, D.M. TinyML: Enabling of Inference Deep Learning Models on Ultra-Low-Power IoT Edge Devices for AI Applications. *Micromachines (Basel).* **2022**, *13*, 851, doi:10.3390/MI13060851.

 doi:10.20944/preprints202605.0748.v1

44. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995; ISBN 978-0-387-94559-0.
45. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32, doi:10.1023/A:1010933404324.
46. Sun, B.; Wu, Z.; Hu, Y.; Li, T. Golden Subject Is Everyone: A Subject Transfer Neural Network for Motor Imagery-Based Brain Computer Interfaces. *Neural Networks* **2022**, *151*, 111–120, doi:10.1016/j.neunet.2022.03.025.
47. Zaharia, M.; Chen, A.; Davidson, A.; Ghodsi, A.; Hong, S.A.; Konwinski, A.; Murching, S.; Nykodym, T.; Ogilvie, P.; Parkhe, V.; et al. Accelerating the Machine Learning Lifecycle with MLflow. *IEEE Data Engineering Bulletin* **2018**, *41*, 39–45.
48. Koelstra, S.; Mühl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31, doi:10.1109/T-AFFC.2011.15.
49. Reiss, A. PAMAP2 Physical Activity Monitoring 2012.
50. Krilova, N.; Kastalskiy, I.; Kazantsev, V.; Makarov, V.A.; Lobov, S. EMG Data for Gestures 2018.
51. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; 2nd ed.; Springer: New York, NY, USA, 2009;
52. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006; ISBN 978-0-387-31073-2.
53. Lu, H.; Eng, H.-L.; Guan, C.; Plataniotis, K.N.; Venetsanopoulos, A.N. Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 3554–3565, doi:10.1109/TBME.2011.2163522.
54. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A Compact Convolutional Neural Network for EEG-Based Brain-Computer Interfaces. *J. Neural Eng.* **2018**, *15*, doi:10.1088/1741-2552/AACE8C.
55. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420, doi:10.1002/HBM.23730.
56. Chaudhary, U.; Banville, H.; Falk, T.H. Convolutional Neural Network Based Approach Towards Motor Imagery Tasks EEG Signals Classification. *IEEE Sens. J.* **2019**, *19*, 4494–4500, doi:10.1109/JSEN.2019.2898373.
57. Dai, M.; Zheng, D.; Na, R.; Wang, S.; Zhang, S. EEG Classification of Motor Imagery Using a Novel Deep Learning Framework. *Sensors* **2019**, *19*, 5516, doi:10.3390/s19245516.
58. Bayram, F.; Ahmed, B.S.; Kassler, A. From Concept Drift to Model Degradation: An Overview on Performance-Aware Drift Detectors. *Knowl. Based. Syst.* **2022**, *245*, 108632, doi:10.1016/J.KNOSYS.2022.108632.
59. Dissanayake, T.; Fernando, T.; Denman, S.; Ghaemmaghami, H.; Sridharan, S.; Fookes, C. Domain Generalization in Biosignal Classification. *IEEE Trans. Biomed. Eng.* **2021**, *68*, 1978–1989, doi:10.1109/TBME.2020.3045720.
60. Yang, C.; Brandon Westover, M.; Sun, J. BIOT: Biosignal Transformer for Cross-Data Learning in the Wild. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 78240–78260.
61. Soekadar, S.R.; Witkowski, M.; Vitiello, N.; Birbaumer, N. An EEG/EOG-Based Hybrid Brain-Neural Computer Interaction (BNCI) System to Control an Exoskeleton for the Paralyzed Hand. *Biomedizinische Technik* **2015**, *60*, 199–205, doi:10.1515/BMT-2014-0126.
62. Zhang, J.; Wang, B.; Zhang, C.; Xiao, Y.; Wang, M.Y. An EEG/EMG/EOG-Based Multimodal Human-Machine Interface to Real-Time Control of a Soft Robot Hand. *Front. Neurorobot.* **2019**, *13*, doi:10.3389/FNBOT.2019.00007.
63. Roy, S.; Chowdhury, A.; McCreadie, K.; Prasad, G.; Dutta, A. An Efficient Multi-Scale CNN Model with Intrinsic Feature Integration for Motor Imagery EEG Subject Classification in Brain-Machine Interfaces. *Biomed. Signal Process. Control* **2022**, *74*, 103496, doi:10.1016/J.BSPC.2022.103496.
64. Aoki, R.; Tung, F.; Oliveira, G.L. Heterogeneous Multi-Task Learning With Expert Diversity. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 3093–3102, doi:10.1109/TCBB.2022.3175456.
65. Casson, A.J. Wearable EEG and Beyond. *Biomed. Eng. Lett.* **2019**, *9*, 53–71, doi:10.1007/S13534-018-00093-6.

 doi:10.20944/preprints202605.0748.v1

66. Leserri, D.; Grimmelsmann, N.; Mechtenberg, M.; Meyer, H.G.; Schneider, A. Evaluation of SEMG Signal Features and Segmentation Parameters for Limb Movement Prediction Using a Feedforward Neural Network. *Mathematics* **2022**, *10*, doi:10.3390/math10060932.
67. Zhang, Y.; Hu, Y.; Jiang, N.; Yetisen, A.K. Wearable Artificial Intelligence Biosensor Networks. *Biosens. Bioelectron.* **2023**, *219*, 114825, doi:10.1016/J.BIOS.2022.114825.