Journal of Universal Computer Science, vol. 32, no. 11 (2026), 1671-1689
submitted: 27/10/2025, accepted: 11/2/2026, appeared: 28/10/2026 CC BY 4.0

Multimodal EEG–Wearable Integration for Generalizable Pain Classification

Cuneyt Ozdemir
(Siirt University, Siirt, Turkey
https://orcid.org/0000-0002-9252-5888, cozdemir@siirt.edu.tr)

Abstract: Objective pain assessment is challenging due to the inherently subjective and multidimensional nature of pain, compounded by intersubject variability and noise in naturalistic settings. These factors limit the development of reliable digital biomarkers for clinical decision support and remote monitoring. In this work, we propose a hybrid deep learning architecture that performs early fusion of single channel EEG band power streams with peripheral biosignals recorded by a wristband (electrodermal activity, photoplethysmography/BVP, triaxial accelerometry, and skin temperature). The model uses residual 1D CNN encoders to learn EEG representations and a multilayer perceptron to encode wearable features; the two embeddings are fused in a unified layer to produce predictions for four classes under a strict subject independent evaluation protocol. On PhysioPain, the proposed approach achieves 86.98% accuracy and 87.46% weighted F1, surpassing widely used machine learning baselines, while maintaining an average inference latency of 0.1495 ms per sample for real time operation. In addition to widely used machine learning baselines, we benchmark contemporary deep 1D CNN baselines for a fair comparison: the proposed model improves over 1D DenseNet by 4.49 percentage points in accuracy and 4.70 percentage points in weighted F1, and over 1D ResNet by 7.90 percentage points in accuracy and 7.93 percentage points in weighted F1. To our knowledge, this is the first study to perform subject independent multiclass classification of natural pain phenotypes with explicit handling of class imbalance by early fusion of MindWave EEG with Empatica E4 signals, representing a step toward generalizable and clinically useful digital biomarkers of pain.

Keywords: Multimodal Fusion; Electroencephalography; Wearable Biosignals; Residual 1D CNN; Pain Classification
Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1
DOI: 10.3897/jucs.176195

1 Introduction

Pain is a complex, subjective, and multidimensional sensory and emotional experience that profoundly affects individuals' quality of life. According to the revised definition by the International Association for the Study of Pain (IASP) in 2020, pain is described as "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" [Raja et al. 2020]. This definition emphasizes that pain is not merely a physical sensation but rather a complex phenomenon influenced by an individual's cognitive, emotional, and social state [Treede et al. 2015]. Furthermore, pain is generally classified as either acute (short-term and typically reversible) or chronic (lasting longer than three months). Chronic pain, in particular, is associated with substantial physical and psychosocial consequences such as depression, anxiety, sleep disorders, and loss of productivity [Treede et al. 2019, IASP Subcommittee on Taxonomy 1986]. Data from the World Health Organization (WHO) further confirm

that pain represents a global health issue, adversely affecting individuals' daily activities, psychological well-being, and economic productivity [World Health Organization 2018].

Global prevalence statistics clearly highlight the widespread burden of pain. Global Burden of Disease (GBD) studies report that chronic pain affects approximately 20–35% of adults worldwide, with higher rates observed in low- and middle-income countries [Jackson et al. 2015, Breivik et al. 2006]. For instance, the prevalence of chronic pain in Europe ranges between 19–31%, while globally more than 1.5 billion individuals are diagnosed with chronic pain each year [GBD 2017 Collaborators 2018, Henschke et al. 2015]. Headache disorders rank among the leading causes of disability worldwide according to GBD data [Stovner et al. 2022], with migraine prevalence ranging between 14–16% and occurring more frequently in women [Stovner et al. 2007]. Low back pain remains the most prevalent musculoskeletal complaint globally and is associated with substantial socioeconomic costs. Menstrual pain (primary dysmenorrhea) affects the majority of women of reproductive age and constitutes a major factor in reduced quality of life. Chronic pain—particularly in forms such as low back pain and menstrual pain—results in annual economic losses amounting to billions of dollars, with costs in the United States alone exceeding $635 billion per year [Gaskin and Richard 2012]. During the COVID-19 pandemic, pain prevalence was reported at 33.3% in 2019, 32.8% in 2020, 32.5% in 2021, and 34.1% in 2022; however, these year-to-year variations were not found to be statistically significant compared to 2019 [Macchia et al. 2024]. Taken together, these findings underscore the urgent need for robust, objective, and generalizable assessment methods that can account for the prevalence and functional impact of everyday pain categories.

In this context, subjective pain assessment scales commonly used in clinical practice—such as the Visual Analog Scale (VAS) and the Numerical Rating Scale (NRS)—although practical, exhibit limitations in certain populations (e.g., neonates, intensive care patients, or individuals with neurological impairments) due to communication difficulties. Moreover, these scales may not always accurately capture inter-individual variability [Williamson and Hoggart 2005]. Such limitations have fueled increasing interest in objective biomarkers that could complement, or in some cases even replace, subjective reporting.

Motivated by this need, the present study focuses on four prevalent clinical pain phenotypes: headache, low back pain, menstrual pain, and the no-pain state (control). These categories represent substantial diversity in both prevalence and clinical outcomes; however, their heterogeneous phenotypes, multifactorial etiologies, and temporally dynamic biological processes pose unique methodological challenges for classification. Headache disorders, particularly migraine and tension-type headache, rank among the leading causes of disability worldwide according to the World Health Organization (WHO) and Global Burden of Disease (GBD) studies [Stovner et al. 2022, Stovner et al. 2007]. Migraine occurs approximately three times more frequently in women than in men and is closely associated with loss of productivity, reduced quality of life, and psychiatric comorbidities [Stovner et al. 2022, Stovner et al. 2007]. Low back pain, which can manifest in both acute and chronic forms, is recognized as the most prevalent musculoskeletal complaint on a global scale. According to the GBD 2017 report, low back pain ranks highest among all conditions in terms of lifetime disability burden [GBD 2017 Collaborators 2018, Henschke et al. 2015]. The multifactorial nature of low back pain—including mechanical, neurological, and psychosocial contributors—further complicates both assessment and treatment. Menstrual pain (primary dysmenorrhea) affects 50–90% of women of reproductive age and is recognized as one of the most common gynecological problems among young women [Iacovides et al. 2015]. This

condition contributes to school and work absenteeism, decreased quality of life, and increased use of analgesics. Owing to hormonal cycles and inter-individual variability, menstrual pain is considered a particularly challenging pain type to assess objectively. The no-pain condition, on the other hand, serves as a critical control group in comparative analyses, enabling the differentiation between pain-specific biomarkers and normal sensory–behavioral responses.

Nevertheless, the classification of these four pain categories entails several challenges. In headache, high clinical heterogeneity (e.g., migraine, tension-type, cluster headache) complicates the identification of discriminative biomarkers. In low back pain, multifactorial etiologies and the risk of chronicity render one-dimensional assessments insufficient. Menstrual pain exhibits temporal variability due to hormonal and individual differences, making consistent measurements difficult. Even the no-pain state, though essential as a baseline in experimental studies, is constrained by inter-individual physiological differences (e.g., stress, sleep patterns, emotional state), which can limit the homogeneity of the control group. Consequently, classification models built on these four categories must contend not only with achieving high accuracy but also with ensuring generalizability and clinical applicability.

Similarly, a variety of approaches have been developed in the literature to classify pain using physiological signals. Early studies primarily relied on unimodal biosignals such as electroencephalography (EEG) or heart rate variability (HRV). For example, Vijayakumar et al. (2017) used time–frequency representations of EEG data with random forest models to quantify tonic thermal pain across subjects, reporting a mean classification accuracy of 89.45% on independent test subjects and identifying the gamma band as the most informative frequency range [Vijayakumar et al. 2017]. Likewise, Gkikas et al. (2024) combined facial videos with heart rate signals in a transformer-based multimodal framework and achieved 82.74% accuracy in binary pain classification [Gkikas et al. 2024].

In recent years, deep learning architectures have increasingly gained prominence in this field. Lopez-Martinez and Picard (2017) proposed multi-task neural networks for personalized pain recognition from physiological signals, showing that modeling inter-individual differences improves recognition performance [Lopez-Martinez and Picard 2017]. Kächele et al. (2016), working with multimodal bio-physiological channels, developed person-centered methods for continuous pain intensity estimation [Kachele et al. 2016]. In the context of headache-specific studies, Bie et al. (2024) trained machine learning models on EEG power features to distinguish individuals with migraine from controls, reporting accuracies of 82–83% and increased occipital alpha-band power at 12 Hz in the migraine group [Bie et al. 2024]. Research on menstrual pain has been more limited; however, Wang et al. (2024) combined resting-state functional magnetic resonance imaging with machine learning and showed that dynamic functional connectivity of the sensorimotor network could accurately distinguish women with primary dysmenorrhea from healthy controls [Wang et al. 2024].

Electroencephalography (EEG) and peripheral physiological signals have emerged as promising modalities in pain assessment. Resting-state EEG studies report increased theta power in chronic pain, along with decreased power in the high-alpha/low-beta range in neuropathic pain [Mussigmann et al. 2022, Zebhauser et al. 2023]. In parallel, autonomic responses such as electrodermal activity (EDA), blood volume pulse (BVP/PPG), skin temperature, and accelerometry provide complementary information regarding the physiological and emotional dimensions of pain. Using multimodal datasets such as the BioVid Heat Pain Database, machine learning models that integrate multiple signal streams have been shown to achieve high accuracy in experimental pain classification

[Thiam et al. 2021, Walter et al. 2013].

Nevertheless, most existing studies focus solely on EEG or peripheral signals and remain limited to binary classification (pain vs. no pain) or pain intensity estimation [Fernandez Rojas et al. 2023, Ahmad and Barkana 2025]. Clinical phenotype-focused studies also tend to address a single condition, such as migraine [Khan et al. 2024], chronic low back pain [De et al. 2025], or primary dysmenorrhea [Rathi et al. 2025, Lee et al. 2017]. While multimodal deep learning approaches are becoming increasingly prevalent, recent examples such as transformer-based fusion of EDA and ECG signals [Farmani et al. 2025] and optimized ensemble deep learning on multimodal physiological signals [Gasmi et al. 2025] have been developed and evaluated on experimentally induced pain datasets such as BioVid, targeting pain detection or pain intensity levels rather than the classification of naturally occurring pain phenotypes. Moreover, only a few studies adopt subject-wise cross-validation or external validation protocols to properly evaluate generalizability across individuals [Petrušić et al. 2024, Mari et al. 2023, Albuquerque et al. 2022]. Recently, the PhysioPain dataset, which includes real-life phenotypes such as headache, low back pain, and menstrual pain, has been introduced [Patlar Akbulut et al. 2025]. However, multimodal and subject-wise deep learning models applied to this dataset remain limited.

EEG-based studies continue to report consistent band-power dynamics across acute and chronic pain states, with findings such as increased θ/β power and a slowed dominant peak frequency in the θ–α range during resting-state conditions drawing particular attention [Zebhauser et al. 2023, Mussigmann et al. 2022]. Phenotype-focused evidence has shown that increases in θ power correlate with affective–sensory components in primary dysmenorrhea [Lee et al. 2017], while chronic low back pain has been associated with regional alterations in alpha/θ/γ responses to acute pain [Kenefati et al. 2023]. Furthermore, studies on radiculopathy have demonstrated that resting-state EEG features can be meaningfully distinguished from healthy controls using SVM [Levitt et al. 2020].

Taken together, this heterogeneity and methodological variability highlight EEG and wearable/peripheral biosignals (EDA, BVP/PPG, skin temperature, accelerometry) as strong candidates for capturing objective components of pain. Evidence to date suggests that EDA, in particular, plays a prominent role. Using leave-one-subject-out (LOSO) protocols, pain vs. no-pain discrimination with EDA alone has achieved accuracies of around 80% [Kong et al. 2021], and in postoperative adults, EDA-based binary classification has yielded clinically meaningful performance [Aqajari et al. 2021]. For continuous intensity estimation and clinically meaningful multi-class scenarios, multimodal approaches are generally preferred over unimodal methods. On controlled experimental datasets (e.g., BioVid), multimodal deep learning approaches have produced strong results in both pain detection and intensity estimation [Thiam et al. 2021, Walter et al. 2013]. Similarly, on the X-ITE dataset, which integrates experimental electrical/thermal pain with behavioral and autonomic signals, methods have been developed for continuous pain monitoring [Othman et al. 2023]. Within the EmoPain framework, which targets chronic low back pain, multimodal motion and sEMG-based classification–regression tasks have been systematically evaluated with standard protocols [Egede et al. 2020].

Despite these advances, an important gap remains in the literature: (i) most studies continue to focus on experimental pain paradigms rather than addressing multiple naturally occurring clinical phenotypes within a single multi-class model; and (ii) generalizability metrics such as subject-wise evaluation and, ideally, external validation remain limited in application [Mari et al. 2022, Albuquerque et al. 2022]. Beyond these dataset and protocol gaps, recent trends in physiological machine learning increasingly emphasize representation learning, robustness under imperfect sensing, and responsible

deployment.

Recent work in physiological time-series learning has moved beyond feature-engineered pipelines and shallow CNNs toward sequence-aware modeling and self-supervised representation learning (SSL), where contrastive objectives can learn transferable temporal structure with reduced reliance on dense labels [Eldele et al. 2021, Wickstrøm et al. 2022]. In parallel, pain estimation studies have increasingly leveraged deep sequence models on autonomic signals, motivating representation learning that generalizes across subjects and contexts [Lopez-Martinez and Picard 2018]. Furthermore, the transition from controlled laboratory settings to real-world deployment has introduced new challenges regarding robustness and fairness. Recent literature emphasizes the need to handle missing modalities (e.g., sensor dropout) and noisy signals typical in ambulatory settings, and surveys have summarized strategies such as multimodal imputation and attention-based fusion for incomplete inputs [Wu et al. 2024]. Simultaneously, algorithmic fairness has gained traction; evidence suggests that AI models in biomedicine can exhibit sex- and gender-related biases, motivating fairness-aware evaluation and learning strategies to ensure equitable performance across demographics [Cirillo et al. 2020]. Finally, practical clinical integration demands addressing privacy and computational efficiency: federated learning (FL) has been proposed to support privacy-preserving training over distributed medical devices [Prasad et al. 2022], while Edge AI optimizations highlight the importance of compute- and energy-efficient inference on resource-constrained wearable platforms [Nahavandi et al. 2022].

While previous studies have explored EEG or autonomic signals independently, few have attempted multimodal integration under naturalistic and subject-wise conditions. The present study is designed to address these gaps. In particular, our main contributions are as follows:

- Novelty in problem formulation: Unlike prior studies focused on experimental pain paradigms [Fernandez Rojas et al. 2023, Walter et al. 2013, Lopez-Martinez and Picard 2017], this work formulates the problem of simultaneously classifying multiple naturally occurring clinical phenotypes—headache, low back pain, and menstrual pain—alongside the no-pain condition within a unified model. This approach addresses the pressing need to identify objective biomarkers that reflect the heterogeneous nature of pain in everyday life outside controlled laboratory settings.

- Early fusion with residual encoders: Representations extracted from EEG and wearable biosignals are integrated through an early fusion architecture that enables inter-modality correlations to be captured at earlier layers. Specifically, the EEG channel employs residual 1D convolutional encoders (Conv1D), allowing temporal dynamics to be extracted more effectively. Compared to late fusion or unimodal approaches [Thiam et al. 2019], this design offers the potential for learning a richer, more integrated feature space.

- Generalizability via subject-wise validation: To ensure clinical applicability and robustness, all evaluations were conducted under a subject-wise protocol. The proposed hybrid model was systematically compared against traditional machine learning baselines (RF, SVM, XGBoost) trained solely on wearable signals, quantitatively demonstrating its superiority. We further include deep 1D-CNN baselines (ResNet-1D and DenseNet-1D) as contemporary references within the same evaluation protocol. This evaluation methodology aligns closely with recent calls for improved

generalizability and external validation in the field [Petrušić et al. 2024, Mari et al. 2023].

– Imbalance-aware training: To address the class imbalance issue commonly encountered in clinical datasets (particularly the scarcity of menstrual pain samples), we designed a training pipeline that combines the focal loss function with the SMOTE oversampling technique. Model performance was reported using metrics that are more reliable under imbalanced scenarios, such as the macro-averaged F1-score, thereby ensuring fair evaluation of minority classes [Fernandez Rojas et al. 2023].

2 Materials and Methods

2.1 Dataset and Preprocessing

This study utilized version 4 of the PhysioPain dataset compiled by F.P. Akbulut et al. [Patlar Akbulut et al. 2025]. Although pain is inherently subjective, the dataset is motivated by the hypothesis that objective assessment is feasible through physiological measurements, including EEG and wearable biosignals. Recordings from 99 participants were acquired using the Empatica E4 wristband and the Mindwave Mobile 2 EEG device. The analysis targeted four categories: three naturally occurring clinical pain phenotypes (headache, low back pain, menstrual pain) and a no-pain (control) condition. EEG comprised ten channels (`Delta, Theta, Alpha1, Alpha2, Beta1, Beta2, Gamma1, Gamma2, Attention, Meditation`) sampled at 1 Hz. Signals were segmented into 30-second sliding windows with 80% overlap (6-second stride). Wearable/peripheral streams with varying sampling rates (`bvp, eda, x, y, z, temperature`) were standardized into 1-second windows. From each 1-second window, we extracted an 18-dimensional feature vector containing time- and frequency-domain statistics (e.g., dominant FFT magnitude for BVP/EDA; mean and variance for accelerometry and temperature). After modality-wise temporal synchronization, each sample consisted of an EEG tensor of size 30×10 and a corresponding 18-dimensional biosignal feature vector. The final preprocessed corpus contained 11,366 matched samples with class labels and participant identifiers, using the label set {`back_pain, headache, menstrual_pain, no_pain`}. To promote subject-level generalizability, the dataset was partitioned in a subject-wise manner: approximately 80% of the data was used for training and 20% for testing. To mitigate class imbalance in the training split (notably the lower prevalence of `menstrual_pain`), we applied the Synthetic Minority Over-sampling Technique (SMOTE) to the training set only, increasing its size to 14,340 samples. The test set remained unaltered to preserve the original distribution for unbiased evaluation.

Scope and practical constraints: The experiments in this study are restricted to the publicly available PhysioPain v4 dataset acquired with a fixed sensing configuration (MindWave Mobile 2 single-channel EEG band-power streams and Empatica E4 wearable biosignals). Accordingly, cross-dataset, cross-device, or cross-site validation is beyond the scope of this work, as no additional compatible datasets with the same label taxonomy and synchronized modalities are available to us. In addition, any fairness or demographic subgroup analysis is limited by the metadata provided within the dataset; where demographic attributes are incomplete or unavailable, we do not perform subgroup-stratified performance reporting and instead discuss fairness as an important direction for future work.

Table 1 summarizes the modalities, windowing strategies, and feature dimensions. Figure 1 depicts the original class distribution prior to SMOTE.

Modality	Channels	Window (overlap)	Sample / dimension
EEG (1 Hz)	Delta, Theta, Alpha1, Alpha2, Beta1, Beta2, Gamma1, Gamma2, Attention, Meditation	30 s (80%)	30×10
Biosignal	bvp, eda, x, y, z, temperature	1 s (sequential)	18

Table 1: Modalities, windowing strategies, and feature dimensions used in this study.

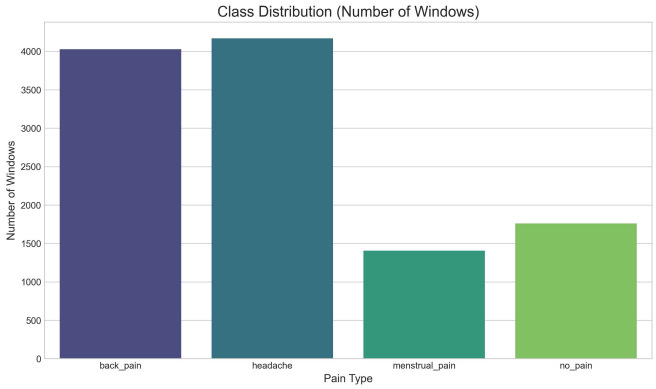

Figure 1: Class distribution (number of windows). Imbalance was addressed in the training set using SMOTE.

2.2 Proposed Hybrid Model Architecture

We propose a hybrid deep learning architecture based on early fusion that jointly exploits EEG band-power streams and wearable/peripheral features to classify four pain states (headache, low back pain, menstrual pain, no-pain) within a unified multi-class framework (Figure 2). The model comprises two parallel branches whose representations are fused and refined before the final classifier.

EEG branch (residual Conv1D encoders The EEG input is a 30×10 tensor per sample (time $\times$ channels). To capture local temporal structure and progressively deepen feature abstraction, the branch stacks four residual Conv1D blocks. Each block contains two $\mathrm{Conv1D}$ layers with kernel size $k{=}5$ and L_2 regularization of 10^{-4}, connected via a skip pathway to preserve gradient flow and stabilize training. A $\mathrm{GlobalAveragePooling1D}$ layer then aggregates temporal information, and a subsequent Dense layer yields a 128-dimensional EEG representation.

Wearable branch and fusion The wearable branch ingests the 18-dimensional feature vector and maps it through a two-layer multilayer perceptron ($\mathrm{Dense}(128) \rightarrow \mathrm{Dense}(128)$) to produce a 128-dimensional wearable representation. The EEG and wearable embeddings are concatenated into a 384-dimensional joint vector, which is passed through a fusion MLP ($128 \rightarrow 128$) equipped with dropout and batch normalization to improve regularization and training stability. A final $\mathrm{Dense}(4)$ layer with softmax outputs the class probabilities. All intermediate layers use the ReLU activation function.

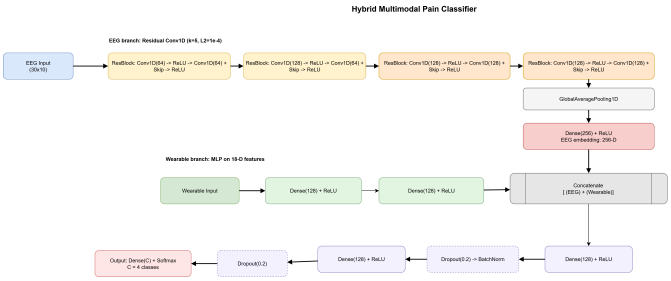

Figure 2: Hybrid multimodal pain classifier architecture.

2.3 Training Protocol and Baselines

To explicitly address class imbalance during optimization, we trained the hybrid model with the focal loss objective, using γ=1.5 and α=0.25. We employed the Adam optimizer with a learning rate of 3×10^{-4}. Early stopping monitored validation performance with a patience of 25 epochs, halting training upon convergence. In addition, the learning rate was adaptively reduced on validation plateaus by a factor of 0.2 with a patience of 7 epochs. Training proceeded with a batch size of 32 for up to 200 epochs, subject to early stopping. The same optimization settings (optimizer, learning rate schedule, batch size, and early-stopping criteria) were also used for the deep 1D-CNN baseline models to ensure a controlled comparison.

For benchmarking, we trained three widely used classical baselines—Random Forest (n=100 trees), probabilistic SVM, and XGBoost—exclusively on wearable biosignal features to isolate the contribution of EEG and multimodal fusion. All models followed the same subject-wise data partitioning as the hybrid approach. Evaluation metrics included overall accuracy, weighted F1-score, class-wise F1-scores, and confusion matrices. In addition, we implemented two contemporary deep-learning baselines, ResNet-1D and DenseNet-1D, and trained them under the same subject-wise split and training protocol.

3 Experimental Results

The proposed hybrid multimodal classifier was systematically evaluated under a strict subject-wise protocol. This section provides a comprehensive analysis of its performance relative to classical baselines, examines class-wise behavior and error patterns, and discusses the model's discriminative ability, robustness, and generalizability. In addition, training dynamics and implications for practical deployment are considered. In addition to classical baselines, we also report results for contemporary deep 1D-CNN baselines (ResNet-1D and DenseNet-1D) under the same subject-wise protocol to contextualize the proposed model's performance against representative neural architectures.

Overall results are summarized in Table 2 and Figure 3. The hybrid model achieved an accuracy of 86.98% and a weighted F1-score of 87.46%, outperforming all baselines by a substantial margin. Among deep-learning baselines, DenseNet-1D performed strongest, followed by ResNet-1D; nevertheless, the proposed hybrid model retained clear gains in both accuracy and weighted F1 while offering lower inference latency. Compared with deep 1D-CNN baselines, the proposed hybrid model improves over DenseNet-1D by +4.49 percentage points in accuracy and +4.70 percentage points in weighted F1, and over ResNet-1D by +7.90 percentage points in accuracy and +7.93 percentage points

in weighted F1. Random Forest, the strongest baseline, reached 75.36% accuracy and 75.98% F1, while SVM and XGBoost performed slightly worse. Thus, the proposed approach provided an absolute improvement of approximately 11.6 percentage points over Random Forest in both metrics. This difference is non-trivial: in a four-class classification setting with clinically heterogeneous categories, such an improvement indicates that multimodal integration with EEG contributes critical discriminatory information that cannot be captured by wearable signals alone.

Latency measurements further emphasize the practical relevance of the approach. The hybrid model requires only 0.1495 ms per-sample for inference, ensuring that even when applied to 30-second input windows, processing remains well within real-time constraints. In comparison, the deep 1D-CNN baselines incur higher latency (0.2186–0.2368 ms per sample), highlighting the proposed model's favorable accuracy–efficiency trade-off. This efficiency makes the architecture deployable in continuous monitoring scenarios, such as bedside systems or wearable healthcare applications.

Model	Accuracy	F1 (weighted)	Inference (ms/sample)
Hybrid (EEG + Wearables)	86.98%	87.46%	0.1495
DenseNet-1D	82.49%	82.76%	0.2186
ResNet-1D	79.08%	79.53%	0.2368
Random Forest	75.36%	75.98%	0
SVM	73.10%	72.97%	0
XGBoost	72.72%	73.77%	0

Table 2: Overall performance (subject-wise evaluation).

Class-wise performance and error analysis

Table 3 presents per-class F1-scores. The hybrid model demonstrated its strongest relative advantage in the more complex and heterogeneous categories. For `back_pain`, performance increased to 73.35%, compared to 59.65% for Random Forest. Similarly, for `headache`, the hybrid model achieved 86.38% F1, substantially outperforming all baselines (best baseline: 70.01%). These gains suggest that EEG contributes discriminative features for conditions with overlapping autonomic signatures, particularly back pain and headache, where wearable-only signals are insufficient.

In contrast, `menstrual_pain` classification reached near-perfect results across all models (F1 $\approx$ 100%). This outcome suggests that menstrual pain has distinctive biosignal patterns, possibly reflecting strong autonomic modulation linked to hormonal cycles. The `no_pain` class also exhibited very high F1-scores, with the hybrid model achieving 97.09%, confirming its ability to reliably identify baseline states.

The confusion matrix in Figure 4 highlights error patterns. The dominant source of misclassification was between `headache` and `back_pain`, where 239 headache samples were predicted as back pain. This overlap aligns with clinical evidence suggesting shared autonomic alterations between migraine and musculoskeletal pain syndromes, such as increased sympathetic activation. Importantly, `menstrual_pain` (246 correct) and `no_pain` (467 correct) were identified almost flawlessly, confirming that their biosignature profiles are robustly captured by the hybrid model.

Figure 3: Comparison of hybrid and baseline models in terms of accuracy and weighted F1-score.

Class	Hybrid	RF	SVM	XGBoost
back_pain	73.35%	59.65%	60.97%	54.75%
headache	86.38%	70.01%	65.96%	66.15%
menstrual_pain	99.60%	100.00%	99.39%	100.00%
no_pain	97.09%	93.74%	87.98%	97.02%

Table 3: Per-class F1-scores (test set, subject-wise).

Discriminative ability: PR and ROC analysis

Figures 5 and 6 provide a detailed view of the model's discriminative performance. Precision–Recall curves show high separability across all classes, with AP values of 82.2% for `back_pain`, 97.1% for `headache`, and nearly perfect values for `menstrual_pain` (100%) and `no_pain` (99.6%). These results confirm that even the most challenging class, back pain, retains sufficient discriminative signal when EEG is incorporated.

ROC curves corroborate these findings, with AUC values of 95.9% (`back_pain`), 97.1% (`headache`), 100% (`menstrual_pain`), and 99.9% (`no_pain`). Such high AUCs indicate that the model is robust to decision threshold variation, a desirable property for clinical applications where sensitivity vs. specificity trade-offs may vary.

Generalizability: per-subject variability

The per-subject box plot in Figure 7 illustrates generalizability. Accuracy values show a narrow interquartile range with a median near 100%, indicating consistently high recognition across individuals. Macro-F1, however, displays broader dispersion, with

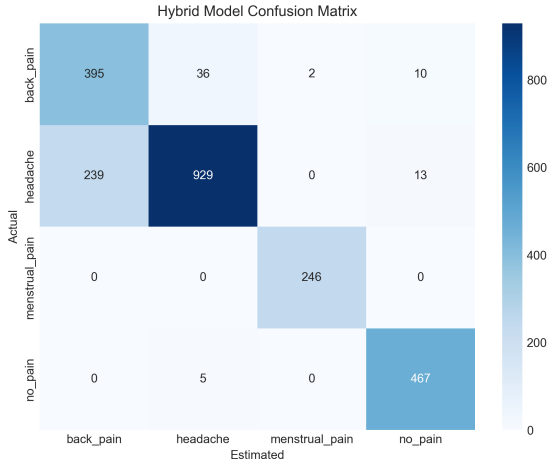

Figure 4: Confusion matrix of the hybrid model (test set).

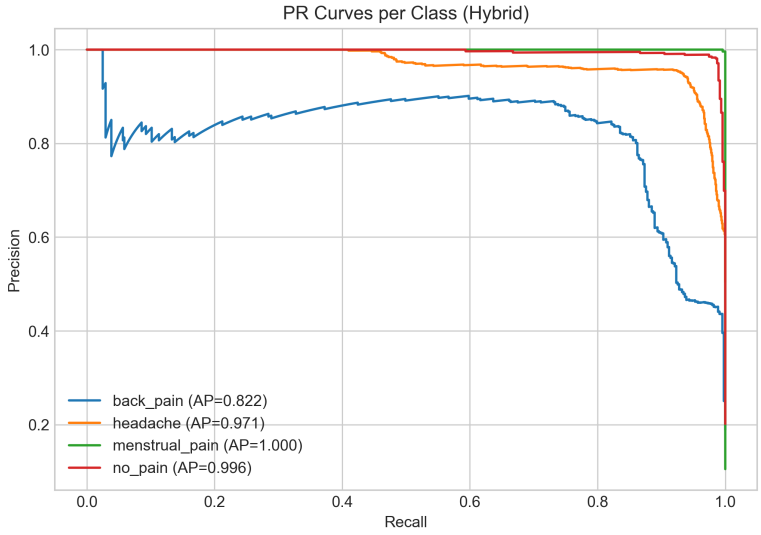

Figure 5: Class-wise PR curves and APs for the hybrid model (test set).

some subjects achieving markedly lower values. This divergence is attributable to inter-subject variability in class distribution (e.g., few menstrual pain samples for certain individuals) and phenotype heterogeneity. Such variability is expected in strict subject-wise evaluation and underscores the importance of including subject-level analysis in addition to aggregated metrics.

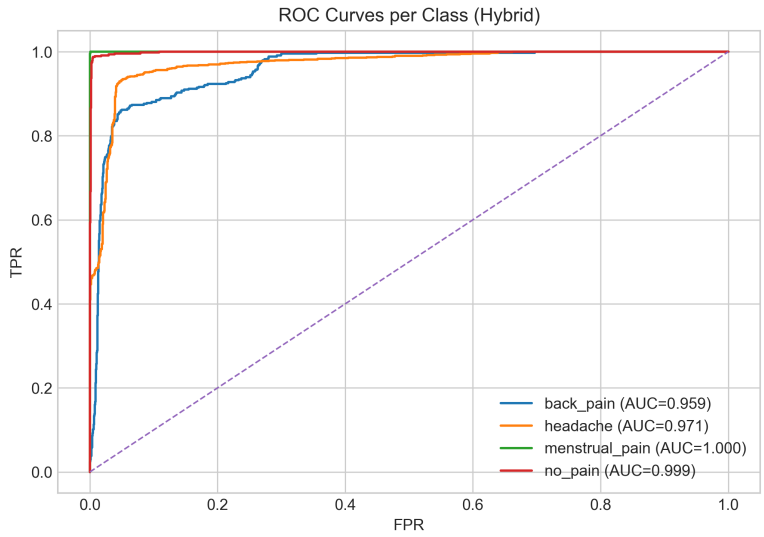

Figure 6: Class-wise ROC curves and AUCs for the hybrid model (test set).

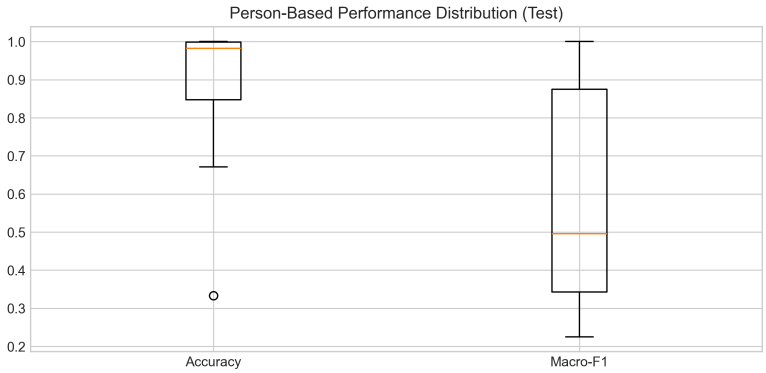

Figure 7: Per-subject distribution of accuracy and macro-F1 (test set).

Training dynamics and overfitting control

Training and validation curves (Figure 8) demonstrate stable convergence. Training loss decreases smoothly, while validation loss stabilizes without escalation, confirming the effectiveness of early stopping. Validation accuracy remains closely aligned with test performance, further confirming the model's robustness to overfitting. Notably, small oscillations in validation loss reflect the class imbalance and subject heterogeneity, yet these fluctuations did not translate into performance degradation.

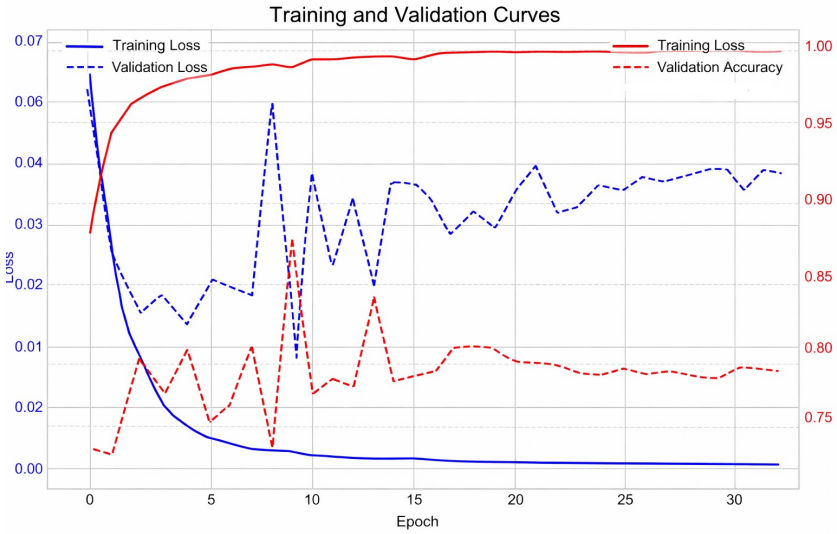

Figure 8: Training and validation curves for loss and accuracy.

Summary of findings

In summary, results demonstrate that the proposed hybrid model: (i) significantly outperforms unimodal baselines, particularly in heterogeneous phenotypes such as back pain and headache; (ii) achieves near-perfect discrimination for menstrual pain and no-pain conditions; (iii) maintains robust discriminative ability across evaluation metrics (F1, PR, ROC); (iv) generalizes well across individuals under a subject-wise protocol, despite expected variability in macro-F1; and (v) converges stably during training without overfitting.

Moreover, when compared to contemporary deep 1D-CNN baselines, the proposed model preserves its advantage while requiring less inference time, supporting its practical suitability for continuous monitoring. Together, these findings validate the hybrid multimodal fusion strategy as an effective and computationally efficient framework for clinically relevant, subject-wise pain classification.

4 Discussion and Conclusion

This study demonstrates that a hybrid deep learning model, integrating EEG band-power streams with wearable biosignals through early fusion, can objectively and subject-wise classify four clinically relevant pain states under natural conditions. The proposed architecture reached an accuracy of 86.98% and a weighted F1-score of 87.46%, significantly surpassing classical machine learning baselines trained solely on wearable features. In addition, benchmarking against contemporary deep 1D-CNN baselines further contextualized these gains: the proposed model improved over DenseNet-1D by +4.49 percentage points in accuracy and +4.70 percentage points in weighted F1, and over ResNet-1D

by +7.90 percentage points in accuracy and +7.93 percentage points in weighted F1 under the same subject-wise protocol. These findings provide compelling evidence that multimodal approaches leveraging both neural and autonomic signatures offer distinct advantages over unimodal strategies for pain phenotyping.

Beyond technical improvements, the results carry important implications for clinical translation. The model's ability to discriminate heterogeneous categories such as `headache` and `back_pain` highlights its promise for developing objective digital biomarkers for conditions traditionally assessed only through self-report. At the same time, the confusion matrix revealed frequent misclassifications between these two classes, suggesting that multimodal integration alone may not fully resolve overlaps in autonomic correlates or stress-related mechanisms. Additional modalities, such as electromyography (EMG), posture, or cardiovascular dynamics, may provide complementary signals to improve discrimination. In contrast, the near-perfect recognition of `menstrual_pain` indicates that this phenotype may be characterized by distinctive biosignal patterns, likely influenced by cyclical hormonal regulation. The `no_pain` condition was also detected with high fidelity, reflecting the model's robustness in capturing baseline states. Importantly, the inference time of approximately 0.15 ms per sample demonstrates that the system is computationally efficient and suitable for real-time deployment in bedside or edge-device applications.

Despite these encouraging results, several limitations must be acknowledged. First, evaluation was limited to a single dataset (PhysioPain), collected with a fixed sensing configuration (MindWave Mobile 2 and Empatica E4), constraining conclusions about cross-population generalizability. Therefore, cross-dataset, cross-device, and cross-site validation could not be performed within the scope of this study. External validation on larger, multi-center cohorts is essential to confirm robustness across demographics, devices, and acquisition contexts. Second, reliance on single-channel EEG and one-second wearable feature windows may not fully capture the spatiotemporal complexity of neural and autonomic dynamics. Richer modalities such as multi-channel EEG, ECG-derived HRV, or fNIRS could enhance sensitivity to subtle pain states. Third, although the hybrid model improved performance on challenging classes, the misclassification of headache versus back pain illustrates persistent challenges in disentangling overlapping physiological processes. Fourth, the present evaluation was conducted under the dataset's standard preprocessing assumptions; we did not perform explicit robustness stress tests under signal noise, motion artifacts, sensor dropout, or missing-modality conditions, which are common in ambulatory deployments. Finally, inter-subject variability in macro-F1 scores suggests that personalized calibration or adaptive learning strategies may be required for consistent performance in diverse populations. In addition, fairness and demographic variability (e.g., sex- and age-related differences in pain responses) could not be systematically assessed, as subgroup-stratified evaluation depends on the availability and completeness of demographic metadata in the dataset.

Future research directions can be structured along three main axes:

1. Model enrichment and robustness. Incorporating richer neural and peripheral modalities (e.g., multi-channel EEG, MEG, HRV, respiration) could expand discriminative capacity. Robustness to sensor dropout, missing modalities, and noisy inputs should be explicitly addressed through modality-dropout training, missing-data modeling, and stress testing under controlled artifact/noise injections.

2. Clinical generalizability. Rigorous external validation across sites and populations is necessary for clinical adoption. In addition, deploying the model in real-world contexts—such as perioperative monitoring, rehabilitation, or chronic pain management—would provide crucial evidence of practical utility. Cross-device studies (e.g.,

clinical-grade EEG vs. consumer-grade EEG) would further clarify performance trade-offs and transferability.

3. Interpretability and deployment. Integrating explainable AI (XAI) approaches to highlight relevant features or temporal segments would improve clinician trust and facilitate physiological insight. Parallel optimization for edge computing would enable deployment on portable or wearable devices, supporting continuous, unobtrusive monitoring. From a deployment perspective, privacy-preserving learning (e.g., federated learning) and practical constraints such as battery usage, long-term monitoring stability, and clinical workflow integration should be evaluated systematically; in this study, these aspects are discussed conceptually but were not experimentally assessed.

In conclusion, this work establishes a strong methodological foundation for multimodal, subject-wise classification of natural clinical pain phenotypes by combining EEG and wearable biosignals. The hybrid architecture not only surpasses unimodal baselines and representative deep 1D-CNN baselines but also provides novel insights into both the distinctiveness and the overlap of pain states. While challenges of generalizability and interpretability remain, the findings represent a foundational step toward clinically deployable, objective pain assessment systems, with the potential to transform both research and practice in pain evaluation.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest/Competing interests

The authors declare that they have no conflict of interest.

Ethics approval and consent to participate

Not applicable. The analyses in this study were conducted on the publicly available PhysioPain dataset. Ethical approval and informed consent were obtained by the original data collectors as stated in the dataset documentation.

Consent for publication

Not applicable. No individual person's data is presented in this manuscript.

Data availability

The dataset analyzed in this study PhysioPain v4 is publicly available through Mendeley Data [Patlar Akbulut et al. 2025]. Links to access the dataset are provided in the references.

Materials availability

Not applicable. This study did not involve the creation of new physical materials.

Code availability

The code used for model development and analysis is available from the corresponding author on reasonable request.

Author contribution

Cuneyt Ozdemir was responsible for the conception and design of the study, data preprocessing, model development, analysis, and manuscript writing. The author read and approved the final manuscript.

References

[Ahmad and Barkana 2025] Ahmad, B., & Barkana, B. D. (2025). Pain and the brain: A systematic review of methods, EEG biomarkers, limitations, and future directions. *Neurology International*, *17*(4), 46. https://doi.org/10.3390/neurolint17040046.

[Albuquerque et al. 2022] Albuquerque, I., Monteiro, J., Rosanne, O., & Falk, T. H. (2022). Estimating distribution shifts for predicting cross-subject generalization in electroencephalography-based mental workload assessment. *Frontiers in Artificial Intelligence*, *5*, 992732. https://doi.org/10.3389/frai.2022.992732.

[Aqajari et al. 2021] Aqajari, S. A. H., Cao, R., Kasaeyan Naeini, E., Calderon, M.-D., Zheng, K., Dutt, N., Liljeberg, P., Salanterä, S., Nelson, A. M., & Rahmani, A. M. (2021). Pain assessment tool with electrodermal activity for postoperative patients: Method validation study. *JMIR mHealth and uHealth*, *9*(5), e25258. https://doi.org/10.2196/25258.

[Bie et al. 2024] Bie, B., Ghosn, S., Sheikh, S. R., Araujo, M. L. D., Mehra, R., Mays, M., & Saab, C. Y. (2024). Electroencephalographic signatures of migraine in small prospective and large retrospective cohorts. *Scientific Reports*, *14*(1), 28673.

[Breivik et al. 2006] Breivik, H., Collett, B., Ventafridda, V., Cohen, R., & Gallacher, D. (2006). Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *European Journal of Pain*, *10*(4), 287–333. https://doi.org/10.1016/j.ejpain.2005.06.009.

[Cirillo et al. 2020] Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., Chadha, A. S., et al. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *npj Digital Medicine*, *3*(1), 81.

[De et al. 2025] De, S., Mukherjee, P., & Roy, A. H. (2025). GLEAM: A multimodal deep learning framework for chronic lower back pain detection using EEG and sEMG signals. *Computers in Biology and Medicine*, *189*, 109928. https://doi.org/10.1016/j.compbiomed.2025.109928.

[Egede et al. 2020] Egede, J., Wang, H., Venkatesh, S., Thiyagarajan, K., Kaltwang, S., Gilbert, A., et al. (2020). EMOPAIN Challenge 2020: Multimodal pain evaluation from facial and bodily expressions. *arXiv preprint arXiv:2001.07739*. https://discovery.ucl.ac.uk/id/eprint/10116032/1/307900a849.pdf.

[Eldele et al. 2021] Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C. K., Li, X., & Guan, C. (2021). Time-series representation learning via temporal and contextual contrasting. *arXiv preprint* arXiv:2106.14112. https://arxiv.org/abs/2106.14112.

[Farmani et al. 2025] Farmani, J., Bargshady, G., Gkikas, S., Tsiknakis, M., & Fernandez Rojas, R. (2025). A CrossMod-Transformer deep learning framework for multi-modal pain detection through EDA and ECG fusion. *Scientific Reports*, *15*(1), 29467. https://doi.org/10.1038/s41598-025-14238-y.

[Fernandez Rojas et al. 2023] Fernandez Rojas, R., Brown, N., Waddington, G., & Goecke, R. (2023). A systematic review of neurophysiological sensing for the assessment of acute pain. *npj Digital Medicine*, *6*(1), 76. https://doi.org/10.1038/s41746-023-00810-1.

[Gaskin and Richard 2012] Gaskin, D. J., & Richard, P. (2012). The economic costs of pain in the United States. *The Journal of Pain*, *13*(8), 715–724. https://doi.org/10.1016/j.jpain.2012.03.009.

[Gasmi et al. 2025] Gasmi, K., Hrizi, O., Ben Aoun, N., Alrashdi, I., Alqazzaz, A., Hamid, O., Altaieb, M. O., Abdalrahman, A. E. M., Ben Ammar, L., Mrabet, M., & Necibi, O. (2025). Enhanced multimodal physiological signal analysis for pain assessment using optimized ensemble deep learning. *Computer Modeling in Engineering & Sciences*, *143*(2), 2459–2489. https://doi.org/10.32604/cmes.2025.065817.

[GBD 2017 Collaborators 2018] GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. (2018). Global, regional, and national incidence, prevalence, and years lived with

disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7.

[Gkikas et al. 2024] Gkikas, S., Tachos, N. S., Andreadis, S., Pezoulas, V. C., Zaridis, D., Gkois, G., Matonaki, A., Stavropoulos, T. G., & Fotiadis, D. I. (2024). Multimodal automatic assessment of acute pain through facial videos and heart rate signals utilizing transformer-based architectures. *Frontiers in Pain Research*, *5*, 1372814.

[Henschke et al. 2015] Henschke, N., Kamper, S. J., & Maher, C. G. (2015). The epidemiology and economic consequences of pain. *Mayo Clinic Proceedings*, *90*(1), 139–147. https://doi.org/10.1016/j.mayocp.2014.09.010.

[Iacovides et al. 2015] Iacovides, S., Avidon, I., & Baker, F. C. (2015). What we know about primary dysmenorrhea today: a critical review. *Human Reproduction Update*, *21*(6), 762–778. https://doi.org/10.1093/humupd/dmv039.

[IASP Subcommittee on Taxonomy 1986] IASP Subcommittee on Taxonomy. (1986). Classification of chronic pain: descriptions of chronic pain syndromes and definitions of pain terms. *Pain*, *Suppl 3*, S1–S226.

[Jackson et al. 2015] Jackson, T., Thomas, S., Stabile, V., Han, X., Shotwell, M., & McQueen, K. A. (2015). Prevalence of chronic pain in low-income and middle-income countries: a systematic review and meta-analysis. *The Lancet*, *385*(Suppl 2), S10. https://doi.org/10.1016/S0140-6736(15)60805-4.

[Kachele et al. 2016] Kachele, M., Thiam, P., Amirian, M., Schwenker, F., & Palm, G. (2016). Methods for person-centered continuous pain intensity assessment from bio-physiological channels. *IEEE Journal of Selected Topics in Signal Processing*, *10*(5), 854–864. https://doi.org/10.1109/JSTSP.2016.2535962.

[Kenefati et al. 2023] Kenefati, G., Rockholt, M. M., Ok, D., McCartin, M., Zhang, Q., Sun, G., Maslinski, J., Wang, A., Chen, B., Voigt, E. P., Chen, Z. S., Wang, J., & Doan, L. V. (2023). Changes in alpha, theta, and gamma oscillations in distinct cortical areas are associated with altered acute pain responses in chronic low back pain patients. *Frontiers in Neuroscience*, *17*, 1278183. https://doi.org/10.3389/fnins.2023.1278183.

[Khan et al. 2024] Khan, L., Shahreen, M., Qazi, A., Jamil Ahmed Shah, S., Hussain, S., & Chang, H.-T. (2024). Migraine headache (MH) classification using machine learning methods with data augmentation. *Scientific Reports*, *14*(1), 5180. https://doi.org/10.1038/s41598-024-55874-0.

[Kong et al. 2021] Kong, W., Zhou, X., et al. (2021). Real-time high-level acute pain detection using a smartphone and electrodermal activity. *Sensors*, *21*(12), 3956. https://doi.org/10.3390/s21123956.

[Lee et al. 2017] Lee, P.-S., Low, I., Chen, Y.-S., Tu, C.-H., Chao, H.-T., Hsieh, J.-C., & Chen, L.-F. (2017). Encoding of menstrual pain experience with theta oscillations in women with primary dysmenorrhea. *Scientific Reports*, *7*(1), 15977. https://doi.org/10.1038/s41598-017-16039-4.

[Levitt et al. 2020] Levitt, J., Edhi, M. M., Thorpe, R. V., Leung, J. W., Michishita, M., Koyama, S., Yoshikawa, S., Scarfo, K. A., Carayannopoulos, A. G., Gu, W., Srivastava, K. H., Clark, B. A., Esteller, R., Borton, D. A., Jones, S. R., & Saab, C. Y. (2020). Pain phenotypes classified by machine learning using electroencephalography features. *NeuroImage*, *223*, 117256. https://doi.org/10.1016/j.neuroimage.2020.117256.

[Lopez-Martinez and Picard 2017] Lopez-Martinez, D., & Picard, R. W. (2017). Multi-task neural networks for personalized pain recognition from physiological signals. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 181–184). https://doi.org/10.1109/ACIIW.2017.8272611.

[Lopez-Martinez and Picard 2018] Lopez-Martinez, D., & Picard, R. W. (2018). Continuous pain intensity estimation from autonomic signals with recurrent neural networks. In *Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5624–5627). IEEE. https://doi.org/10.1109/EMBC.2018.8513575.

[Macchia et al. 2024] Macchia, L., Delaney, L., & Daly, M. (2024). Global pain levels before and during the COVID-19 pandemic. *Economics & Human Biology*, *52*, 101337. https://doi.org/10.1016/j.ehb.2023.101337.

[Mari et al. 2022] Mari, T., Henderson, J., Maden, M., Nevitt, S., Duarte, R., & Fallon, N. (2022). Systematic review of the effectiveness of machine learning algorithms for classifying pain intensity, phenotype or treatment outcomes using electroencephalogram data. *The Journal of Pain*, *23*(3), 349–369. https://doi.org/10.1016/j.jpain.2021.07.011.

[Mari et al. 2023] Mari, T., *et al.* (2023). External validation of binary machine learning models for pain intensity perception classification from EEG in healthy individuals. *Scientific Reports*, *13*(1), 143. https://doi.org/10.1038/s41598-022-27298-1.

[Mussigmann et al. 2022] Mussigmann, T., Bardel, B., & Lefaucheur, J.-P. (2022). Resting-state electroencephalography (EEG) biomarkers of chronic neuropathic pain. A systematic review. *NeuroImage*, *258*, 119351.

[Nahavandi et al. 2022] Nahavandi, D., Alizadehsani, R., Khosravi, A., & Acharya, U. R. (2022). Application of artificial intelligence in wearable devices: Opportunities and challenges. *Computer Methods and Programs in Biomedicine*, *213*, 106541.

[Othman et al. 2023] Othman, E., Mutlu, B., & Tönnies, K. (2023). Classification networks for continuous automatic pain intensity monitoring in video using facial expression on the X-ITE Pain Database. *Journal of Visual Communication and Image Representation*, *91*, 103743. https://doi.org/10.1016/j.jvcir.2022.103743.

[Patlar Akbulut et al. 2025] Patlar Akbulut, F., Toktay, B., Işık Orhan, İ., & Yıldırım, E. (2025). PhysioPain Dataset. *Mendeley Data*, V4. https://doi.org/10.17632/mf2cgph9cy.4.

[Petrušić et al. 2024] Petrušić, I., Savić, A., Mitrović, K., Bačanin, N., Sebastianelli, G., Secci, D., & Coppola, G. (2024). Machine learning classification meets migraine: recommendations for study evaluation. *The Journal of Headache and Pain*, *25*(1), 215. https://doi.org/10.1186/s10194-024-01924-x.

[Prasad et al. 2022] Prasad, V. K., Bhattacharya, P., Maru, D., Tanwar, S., Verma, A., Singh, A., Tiwari, A. K., Sharma, R., Alkhayyat, A., Ţurcanu, F.-E., et al. (2022). Federated learning for the internet-of-medical-things: A survey. *Mathematics*, *11*(1), 151.

[Raja et al. 2020] Raja, S. N., Carr, D. B., Cohen, M., Finnerup, N. B., Flor, H., Gibson, S., Keefe, F. J., Mogil, J. S., Ringkamp, M., Sluka, K. A., Song, X.-J., Stevens, B., Sullivan, M. D., Tutelman, P. R., Ushida, T., & Vader, K. (2020). The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *Pain*, *161*(9), 1976–1982. https://doi.org/10.1097/j.pain.0000000000001939.

[Rathi et al. 2025] Rathi, G., Anuradha, R., & Pramila, M. (2025). Impact of yoga asanas on primary dysmenorrhea: EEG band power analysis and deep learning based classification. *Egyptian Informatics Journal*, *31*, 100741.

[Stovner et al. 2007] Stovner, L. J., Hagen, K., Jensen, R., Katsarava, Z., Lipton, R. B., Scher, A. I., Steiner, T. J., & Zwart, J.-A. (2007). The global burden of headache: a documentation of headache prevalence and disability worldwide. *Cephalalgia*, *27*(3), 193–210.

[Stovner et al. 2022] Stovner, L. J., Hagen, K., Linde, M., & Steiner, T. J. (2022). The global prevalence of headache: an update, with analysis of the influences of methodological factors on prevalence estimates. *The Journal of Headache and Pain*, *23*(1), 34. https://doi.org/10.1186/s10194-022-01402-2.

[Thiam et al. 2019] Thiam, P., Bellmann, P., Amirian, M., Layher, G., Zhang, H., Kessler, V., Gruss, S., Walter, S., Eskofier, B., André, E., & Al-Hamadi, A. (2019). Exploring deep physiological models for nociceptive pain estimation. *Sensors*, *19*(20), 4503. https://doi.org/10.3390/s19204503.

[Thiam et al. 2021] Thiam, P., Hihn, H., Braun, D. A., Kestler, H. A., & Schwenker, F. (2021). Multi-modal pain intensity assessment based on physiological signals: A deep learning perspective.

Frontiers in Physiology, *12*, 720464. https://doi.org/10.3389/fphys.2021.720464.

[Treede et al. 2015] Treede, R.-D., Rief, W., Barke, A., Aziz, Q., Bennett, M. I., Benoliel, R., Cohen, M., Evers, S., Finnerup, N. B., First, M. B., Giamberardino, M. A., Kaasa, S., Kosek, E., Lavand'homme, P., Nicholas, M., Perrot, S., Scholz, J., Schug, S., Smith, B. H., Svensson, P., Vlaeyen, J. W. S., & Wang, S.-J. (2015). A classification of chronic pain for ICD-11. *Pain*, *156*(6), 1003–1007. https://doi.org/10.1097/j.pain.0000000000000160.

[Treede et al. 2019] Treede, R.-D., Rief, W., Barke, A., Aziz, Q., Bennett, M. I., Benoliel, R., Cohen, M., Evers, S., Finnerup, N. B., First, M. B., Giamberardino, M. A., Kaasa, S., Kosek, E., Lavand'homme, P., Nicholas, M., Perrot, S., Scholz, J., Schug, S., Smith, B. H., Svensson, P., Vlaeyen, J. W. S., & Wang, S.-J. (2019). Chronic pain as a symptom or a disease: the IASP classification for ICD-11. *Pain*, *160*(1), 19–27.

[Vijayakumar et al. 2017] Vijayakumar, V., Case, M., Shirinpour, S., & He, B. (2017). Quantifying and characterizing tonic thermal pain across subjects from EEG data using random forest models. *IEEE Transactions on Biomedical Engineering*, *64*(12), 2988–2996.

[Walter et al. 2013] Walter, S., Gruss, S., Ehleiter, H., Tan, J., Traue, H. C., Werner, P., Al-Hamadi, A., Crawcour, S., Andrade, A. O., & Moreira da Silva, G. (2013). The BioVid Heat Pain Database: Data for the advancement and systematic validation of an automated pain recognition system. In *2013 IEEE International Conference on Cybernetics (CYBCONF)* (pp. 128–131). https://doi.org/10.1109/CYBCONF.2013.6617456.

[Wang et al. 2024] Wang, C., He, J., Feng, X., Qi, X., Hong, Z., Dun, W., Zhang, M., & Liu, J. (2024). Characteristics of pain empathic networks in healthy and primary dysmenorrhea women: an fMRI study. *Brain Imaging and Behavior*, *18*(5), 1086–1099.

[Wickstrøm et al. 2022] Wickstrøm, K., Kampffmeyer, M., Mikalsen, K. Ø., & Jenssen, R. (2022). Mixing up contrastive learning: Self-supervised representation learning for time series. *Pattern Recognition Letters*, *155*, 54–61.

[Williamson and Hoggart 2005] Williamson, A., & Hoggart, B. (2005). Pain: a review of three commonly used pain rating scales. *Journal of Clinical Nursing*, *14*(7), 798–804. https://doi.org/10.1111/j.1365-2702.2005.01121.x.

[World Health Organization 2018] World Health Organization. (2018). *WHO Guidelines for the Pharmacological and Radiotherapeutic Management of Cancer Pain in Adults and Adolescents*. Geneva: World Health Organization. ISBN 978-92-4-155039-0. https://iris.who.int/handle/10665/279700.

[Wu et al. 2024] Wu, R., Wang, H., Chen, H.-T., & Carneiro, G. (2024). Deep multimodal learning with missing modality: A survey. *arXiv preprint* arXiv:2409.07825. https://arxiv.org/abs/2409.07825.

[Zebhauser et al. 2023] Zebhauser, P. T., Hohn, V. D., & Ploner, M. (2023). Resting-state electroencephalography and magnetoencephalography as biomarkers of chronic pain: A systematic review. *Pain*, *164*(6), 1200–1221. https://doi.org/10.1097/j.pain.0000000000002825.