Exploring the Relationship Between Stress-Physiology and Pain in the Daily Life of Patients With Chronic Widespread Pain

Emilie Pattyn, Nattapong Thammasan, Hannah Davidoff, Walter De Raedt, Gudrun Vera Eisele, Ruud van Stiphout, Maarten De Vos, Olivia J. Kirtley, Peter Van Wambeke, Bart Morlion, Elfi Vergaelen, and Chris Van Hoof

***Abstract*—Chronic widespread pain remains a complex and incompletely understood condition. To complement existing pain assessment strategies, this study explored the ecological validity of unobtrusively captured daily life physiological signals as indicators of pain. Therefore, we collected physiological data using a wearable wristband from 46 patients with chronic widespread pain for seven days. Linear mixed-effect models revealed several significant associations between physiological signals, such as mean heart rate and momentary pain intensity. However, making individual pain predictions with multivariate machine learning models did not add value. While this study underscores the potential of ambulatory physiology for pain assessment, future research should validate and expand upon these initial findings to further enhance pain management strategies.**

***Index Terms*—Chronic widespread pain, ecological momentary assessment, pain assessment, physiology, wearable sensors.**

I. Introduction

VALID and continuous pain assessment can enhance the evaluation of treatment efficacy in patients with chronic pain, defined as pain lasting or reoccurring for at least three months [1], thereby improving their quality of life. Wearable technology could offer a complementary strategy to existing pain monitoring methods by providing continuous, real-time, and personalized data [2]. Given the complex and individual nature of pain, long-term monitoring of physiological signals and pain-related symptoms could generate new insights into the underlying causes and mechanisms of chronic pain, leading to more effective and personalized treatments [3].

Traditionally, pain is assessed through verbal reporting, which is considered the gold standard but this method might be susceptible to recall bias [4]. Recall bias is more likely when patients have to recall their pain over extended periods (e.g., 14 days). To help reduce recall bias, alternative methods (with shorter reflection periods) such as questionnaire-based self-reports, pain diaries, or ecological momentary assessments (EMA) have been used for prolonged pain monitoring [5]. In recent years, research has also explored physiological signals captured by wearable sensors as another approach to objectively assess pain. For instance, previous research has demonstrated accurate discrimination between baseline and induced pain in laboratory settings among healthy individuals using physiological signals [6], [7]. These physiological signals could complement EMA, particularly when compliance is low or inconsistent [8]. However, limitations include possible discomfort from the monitoring devices [4].

Multiple physiological signals are indirectly linked to pain stimuli through the physiological stress response system. More specifically, acute pain triggers the physiological stress response, activating both the sympathetic-adrenal-medullary (SAM) system and the hypothalamic-pituitary-adrenal (HPA) axis. The SAM system activates the autonomic nervous system (ANS), resulting in increased sympathetic ANS activation and decreased parasympathetic ANS activation. In turn, these changes are reflected in several measurable physiological signals which can be captured using wearable sensors such as heart rate (HR), electrodermal activity (EDA), and distal skin temperature (ST) [9], [10], [11], [12]. For a more detailed overview per modality, we refer to Fernandez Rojas et al. (2023) [4]. These physiological signals have been already used to model different types of pain, as detailed in previous reviews: [3], [4].

This study explores the feasibility of using ambulatory-collected physiological signals for pain assessment in a clinical population of patients diagnosed with chronic widespread pain (CWP). Chronic widespread pain is a specific type of chronic pain, which is present in at least four of the five body regions,

Received 19 August 2024; revised 26 January 2025; accepted 21 February 2025. Date of publication 25 February 2025; date of current version 15 September 2025. The work of Gudrun Vera Eisele was supported by an FWO fellowship under Grant 1223725N. Recommended for acceptance by J. Hernandez. *(Corresponding author: Chris Van Hoof.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the ethical committee of the University Hospital of Leuven under Application No. S67061, and performed in line with the Declaration of Helsinki.

Emilie Pattyn and Hannah Davidoff are with KU Leuven, 3000 Leuven, Belgium, and also with Imec, B-300 Leuven, Belgium.

Nattapong Thammasan and Ruud van Stiphout are with OnePlanet Research Center, 6708 WE Wageningen, Netherlands.

Walter De Raedt is with Imec, B-300 Leuven, Belgium.

Gudrun Vera Eisele, Maarten De Vos, Olivia J. Kirtley, Bart Morlion, and Elfi Vergaelen are with KU Leuven, 3000 Leuven, Belgium.

Peter Van Wambeke is with the Leuven Centre for Algology & Pain Management, University Hospitals of Leuven, B-3000 Leuven, Belgium.

Chris Van Hoof is with KU Leuven, 3000 Leuven, Belgium, and with OnePlanet Research Center, 6708 WE Wageningen, Netherlands, and also with Imec, B-300 Leuven, Belgium (e-mail: chris.vanhoof@kuleuven.be).

This article has supplementary downloadable material available at https://doi.org/10.1109/TAFFC.2025.3545477, provided by the authors.

Digital Object Identifier 10.1109/TAFFC.2025.3545477

1949-3045 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

is associated with emotional distress or functional disability, and has a prevalence of about 9% in Europe [13], [14]. Social, psychological, and biological contributors can be defined [14]. Furthermore, the underlying nociceptive processes remain unclear. Given the incomplete understanding of its etiology and mechanisms, the detection and treatment of CWP remains inconsistent [15].

Given the complex nature of CWP, this study aims to explore whether ambulatory-collected physiological signals can be valuable for ambulatory pain assessment in individuals with chronic widespread pain. Therefore, we used ambulatory-collected data from 46 patients with CWP over seven days. Specifically, participants wore a wearable wristband to record their physiological signals and completed up to ten EMA questionnaires per day to assess their momentary pain and several psychological variables. The aims of the study are to:

- Exploring the associations between univariate physiological biomarkers and pain at group level using mixed-effects models,
- Evaluating the performance of different multivariate regression models for predicting pain at individual level,
- Investigating whether including physiological features alongside EMA-related features enhances the performance of individual-level pain predictions.

II. Related Work

Previous research has investigated the associations between pain and physiological signals captured by wearable sensors. To the best of our knowledge, no previous studies have examined the relationship between physiological features and pain in patients with chronic widespread pain in daily life. Hence in this section, we first elaborate on findings from previous laboratory studies followed by those from ambulatory studies in this section.

A. Laboratory Pain Physiology Associations

Multiple studies have investigated the associations between pain and physiological features in laboratory settings, both in healthy individuals and individuals with chronic pain. Studies have shown that changes in HR and skin conductance resulting from pain can be successfully captured by wearable sensors through photoplethysmography (PPG) and EDA, respectively [16]. For example, Loggia et al. (2011) [10] demonstrated that the intensity of the heat pain stimulus is related to the strength of the physiological response in healthy participants (within correlations: 0.1–1.0 for EDA with 31 of the 39 individuals having a p-value < 0.05 and -0.6–1.0 for HR with 12 of the 39 individuals having a p-value < 0.05, between correlations: 0.22–0.24 for EDA and 0.40-0.46 for HR). However, they also reported a high inter-individual variability modulated by confounding factors, such as gender [2], [17].

In patients with chronic widespread pain, research suggests dysfunction of the physiological stress system, characterized by heightened sympathetic and diminished parasympathetic activity [18], along with a blunted physiological stress response after stress or pain induction [19], [20] although findings have been inconsistent [21]. However, the causal relationship between stress response dysfunction and the onset of chronic pain remains unknown [22].

Previous studies explored the prediction of pain based on physiological signals during experimentally induced pain stimuli, including thermal, electrical, or mechanical stimuli [2]. These studies typically employed multiple physiological modalities, such as electrocardiography (ECG), EDA, or electromyography (EMG), mostly in healthy participants. While binary pain discrimination of pain stimuli, i.e., pain versus no pain, achieved rather high accuracies, e.g., 74-94% [7], [23], [24], [25], [26], [27], [28], [29], multi-level pain discrimination and regression remained challenging, with accuracies between 37-61% for 4- or 5-class pain classification [7], [23], [27], [28], [29]. These regression models obtained a coefficient of determination (R^2) between 0.24-0.48 [25]. Table I summarizes some pain studies that trained regression models in laboratory settings using several stress-physiology-related features as input.

B. Ambulatory Pain Physiology Associations

While laboratory studies provide valuable insights, they do not capture the complexity of real-life physiological variations and pain experiences. Table II summarizes several studies that modelled pain intensity based on stress-physiology-related signals captured by wearable sensors in ambulatory conditions using machine learning models. Currently, these studies contain only a limited number of data points for which the pain intensity is known, either caused by small sample sizes or short monitoring duration. Furthermore, the relatively low observed strength of the associations between ambulatory collected physiology and negative-arousal-related indicators, such as pain or stress, may be attributed to the complexity of monitoring stress physiology in daily life using wearables [37], [38]. Signals collected by wearables are influenced by various contextual factors, such as physical activity, environmental conditions, or substance intake, complicating accurate assessment [38], [39]. Additionally, comparisons between studies are hindered by differences in datasets with small sample sizes, as well as the lack of standardized normalization procedures and preprocessing choices, complicating the field's progress [38], [40].

Our study aims to evaluate the ecological validity of physiological signals for ambulatory pain measurement of 46 patients with CWP, who reported their pain up to 10 times per day over seven days. We seek to explore whether the relationships between physiological signals and chronic pain can be detected at group level despite the contextual influences present in ambulatory conditions. Additionally, we aim to demonstrate the potential of physiological data for pain assessment at individual level in patients with CWP in real-world settings.

III. Methods

The methods of this work have been postregistered [41], a type of preregistration for existing data, following data collection to the Open Science Framework prior to data analysis (osf.io/fevx5) and are described in this section. Deviations from the postregristration are described in the supplementary materials.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SUMMARY OF SEVERAL LABORATORY PAIN PREDICTION STUDIES THAT TRAINED REGRESSION MODELS

Study	Population	Number of participants	Measured signals	Pain modality	Pain levels	Models	Performance
[28]	Healthy participants	87 (BioVid dataset)	EMG, EDA, ECG	Thermal pain	4 levels + no pain	*Regression*: random forest *Classification*: random forest	*Regression*: mean absolute error: 0.892 *Classification*: accuracy: 40.1%
[29]	Healthy participants	87 (BioVid dataset)	EMG, EDA, ECG, facial video	Thermal pain	4 levels + no pain	*Regression*: random forest Classification: random forest	*Regression*: root mean squared error: 1.16, mean absolute error: 0.99 *Classification*: accuracy: 83.1% for binary classification (no pain vs. high pain)
[23]	Healthy participants	87 (BioVid dataset)	EDA, ECG	Thermal pain	4 levels + no pain	*Regression*: long short term memory, neural network, supported vector machine, and linear regression *Classification*: logistic regression, and supported vector machine	*Regression*: mean absolute error: 1.05, root mean squared error: 1.29, coefficient of determination: 0.24 *Classification*: accuracy: 74.2% for binary classification (no pain vs. high pain)
[25]	Healthy participants	16	EDA (on fingers)	Electrical pain	4 levels + no pain	*Regression*: multilayer perceptron, and supported vector machine *Classification*: multilayer perceptron, and supported vector machine	*Regression*: coefficient of determination: 0.38-0.48, root mean squared error: 1.11-1.21, mean absolute error: 0.87-0.93 *Classification*: balanced accuracy: 44.1%, root mean squared error: 1.279, mean absolute error: 0.883
[30]	Healthy participants	87 (BioVid dataset)	EMG, EDA, ECG, facial video	Thermal pain	4 levels + no pain	*Regression*: linear regression, supported vector machine, neural network, random forest, k-nearest neighbors, and extreme gradient boosting	*Regression*: mean absolute error: 0.93, root mean squared error: 1.13

TABLE II
SUMMARY OF SEVERAL AMBULATORY PAIN PREDICTION STUDIES

Study	Population	Number of participants	Study duration	Number of points with known pain	Measured signals and wearable	Pain levels	Models	Highest reported model performance
[31]	Patients with migraine	2	4-6 weeks	32 migraines	ECG, ST, EDA, peripheral oxygen saturation by PLUX-Wireless Biosignals and Nonin Onyx II	Relative changes in pain intensity or punctual pain levels	State space model	F1: 0.62
[32]	Patients with cancer	21	4 days	44 points self-reported at the beginning of a VR session	HR and HRV from PPG, EDA, ACC, ST	11-point NRS, rescaled to no pain and pain	*Classification*: SVM, RF, MP, log, AdaBoost	Accuracy (SVM): 73%
[33]	Patients with sickle cell disease	20	average 4 hours (std 2 hours)	40 points self-reported on an iPad mobile app and 67 nursing scores	HR from PPG, EDA, ST, accelerometer, gyroscope Microsoft Band 2	Reported 11-point VAS- scores rescaled into 3 pain classes + no pain (for classification)	*Regression*: Ridge, lasso, GP, SVR, and SVM (11 levels)	Root mean squared error (SVR): 1.430, Pearson correlation: 0.706
[34]	Patients with sickle cell disease	20	average 2 hours	median per patient: 980 points, but assumed pain to be constant for 15 min, thus 5 labelled points*	HR and HRV from PPG, calorie burn, basal calories burned by Apple watch	Pain reported between 0-10 by a nurse and extracted from electronic medical record	*Classification*: Logistic regression classification, gradient boosting, and RF	Root mean squared error (RF): 0.84, area under ROC curve: 0.98, F1: 0.85, and accuracy: 84%

* calculated from the numbers in the paper: median of 980 points per patient and one HR per 5 seconds: (980/ (15*60/5) =5.4)

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

TABLE III
EXTRACTED FEATURES FOR MIXED EFFECT MODELS PER SIGNAL

Signal	Features
HR	Mean, minimum, maximum, standard deviation
EDA	Mean, minimum, maximum, standard deviation, number of peaks per minute, peak amplitude
Tonic	Mean, minimum, maximum, standard deviation
Phasic	Mean, minimum, maximum, standard deviation, number of peaks per minute, peak amplitude
ST	Standard deviation, max slope per minute, mean slope per minute
ACC magnitude	Standard deviation

A. Data Collection

This work used both EMA data and ANS-related physiological data collected under ambulatory conditions for seven days from 48 patients with CWP. However, due to technical issues with two wristbands, resulting in severe data loss, the final dataset size was 46 patients. Data collection occurred from March 2023 until March 2024.

The inclusion criteria required patients to have a diagnosis of CWP, in which pain had to be present on both sides of the body, both above and below the waist and in the spinal region. Thereby, underlying explanatory pathologies were ruled out. Patients with a sub diagnosis of fibromyalgia (69%) met the 1990 American College of Rheumatology (ACR) criteria [42]. Moreover, they had to be able to speak and understand Dutch and be between 18 and 65 years old. Exclusion criteria included pregnancy or breastfeeding, an unstable medication intake schedule within one month before the start of the study and during the study, the use of beta-blockers, corticoids, or inflammatory drugs except for non-steroidal anti-inflammatory drugs, a history of drugs or alcohol abuse within six months before the start of the study, the use of antibiotics within three months before the start of the study, an active COVID-19 infection within two weeks before the start of the study, having symptoms of COVID-19 (e.g., fever, cough, diminished smell or taste capability) within two weeks before the start of the study, or having other serious disease(s) such as malignancy, severe heart disease, kidney disease, or neurological disease. Patients were allowed to participate if they took benzodiazepines only if their dosage had not changed for at least three months [43]. Patients were recruited at the University Hospital of Leuven's Psychiatry Department of and the Leuven Algological Center. Clinicians introduced the study to patients with CWP and connected interested individuals with researchers. Flyers were also distributed in patient communities. The study received approval from the ethical committee of the University Hospital of Leuven (S67061) and was conducted in accordance with the Declaration of Helsinki. All participants signed written informed consent and were compensated for their time.

Participants were instructed to wear a Chill+ wristband, developed by imec for research purposes, for seven days. The Chill+ wristband captured acceleration (ACC) in the x, y, and z-axis at 32 Hz, angular acceleration at 32 Hz, EDA at 64 Hz, PPG at 128 Hz, and ST at 1 Hz. Additionally, participants received 10 notifications per day on their own phone, distributed between 8 a.m. and 10 p.m. using the m-Path application [44] (Table S9). Participants could request to adjust the questionnaire time window to better suit their daily routine, which was done by two patients. The EMA notifications were randomly scheduled within 70-minute blocks throughout the day, ensuring a minimum gap of 15 minutes between consecutive notifications, e.g., the first notification between 8:00 a.m. and 9:10 a.m. was followed by the second notification between 9:25 a.m. and 10:35 a.m. Participants had a 10-minute time limit to respond, with one reminder sent after 5 minutes if no response was received. The EMA questionnaire consisted of a minimum of 18 items and a maximum of 29 items as some questions were asked conditionally. The prompts questioned pain (intensity, unpleasantness, course), affective state, pain catastrophizing, physical state, and context. The first questionnaire, filled in on a new day, contained six additional questions regarding sleep. Additionally, one control question was repeated three times during the EMA trial in the 6th, 25th, and 63rd scheduled prompts.

For more detailed study-related information, including the instructions given to the participants regarding the wristband and EMA and the full EMA questionnaire, we refer to osf.io/yxmdz.

B. Data Preprocessing

Both the physiological and the EMA data were preprocessed before analysis. First, we preprocessed the ambulatory physiological signals to remove and filter out segments contaminated by artefacts resulting from excessive movement, local pressure, or environmental conditions. These preprocessing steps were performed exactly as postregistered, and a more detailed description can be found in the supplementary materials. The magnitude of acceleration was computed by taking the square root of the sum of the squared signals from each axis, i.e., x, y, and z, with the standard deviation (std) of the magnitude serving as an approximation for activity [45]. Additionally, to further characterize activity levels, we used the activity classifier of Stojchevska et al., (2022) [46]. More specifically, we categorized the average estimated probabilities per minute into limited (lying, sitting, standing), moderate (walking), and vigorous (cycling, running) activity. On average, patients wore the wristband for 132 hours (std: 26) between 6 a.m. and midnight. The average fraction of high-quality windows for HR was 76% (std:13%), for EDA was 75% (std: 18%), and for both HR and EDA was 61% (std: 18%) as computed from the physiology extracted in 5-minute rolling windows (with a step of one minute). Furthermore, patients had on average an EDA of 2.05 μS (std:1.96 μS), HR of 80 bpm (std: 8 bpm), ST of 31°C (std: 2°C), ST variation of 0.20°C (std: 0.16°C), magnitude of acceleration of 1.02 g (std: 0.009 g), standard deviation of magnitude of acceleration 0.08 g (std: 0.02 g).

The collected EMA data, originally consisting of 2613 observations (compliance: 78%) were preprocessed. Using the method described by Jaso et al., (2022) [47] we identified and removed careless observations, characterized by unusually low response times, random answer selections, or a lack of attention

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

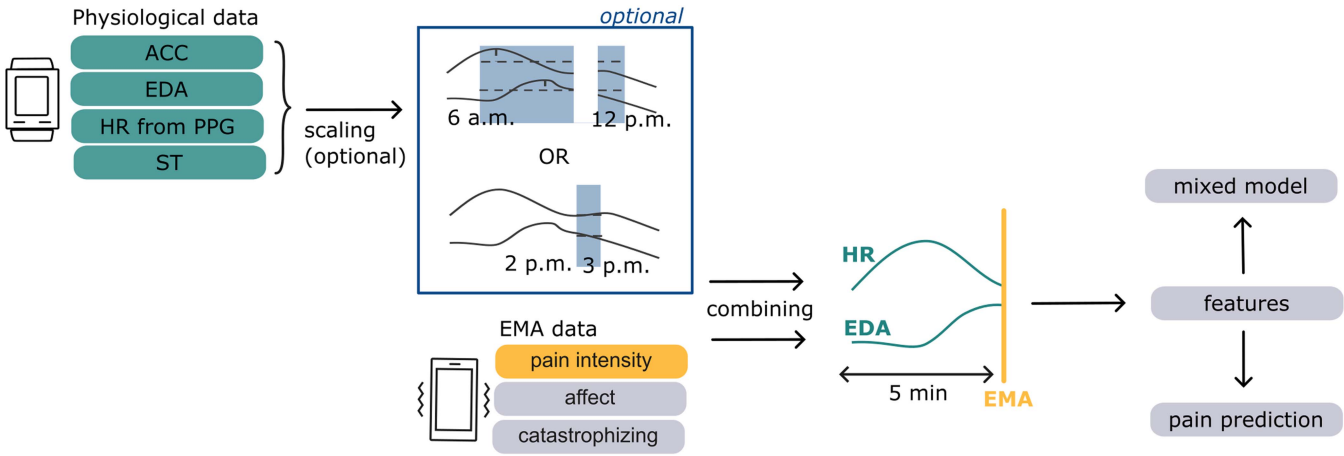

Fig. 1. Integration process of physiology signals and momentary reported EMA. Physiological data included accelerometer (ACC), electrodermal activity (EDA), heart rate (HR), photoplethysmography (PPG), and skin temperature (ST). For the mixed models, pain intensity was included as an EMA variable, whereas the pain prediction models also incorporated additional EMA variables such as affect and catastrophizing.

to question content. More specifically, we removed observations in which the response time of the last question was zero (N = 1), observations with a total response time larger than the preset fill-in window of 15 minutes (N = 12), and observations in which the control question was answered incorrectly (N = 10). We computed the average response time per item, in which we excluded questions with open response options to avoid bias from typing speed variations. Using the quartile rule threshold, we obtained a threshold of 153 seconds and removed 171 items accordingly. Additionally, we removed observations with a mean response time per item lower than one second, a standard deviation of response options lower than 0.5, and a fraction of the mode answer option higher than 45% (N = 23). Detailed results regarding the distributions, thresholds, and slopes are provided in the supplementary material (Fig. S1-S2 and Table S1).

Finally, the physiological data were coupled to the EMA data by extracting the physiological signals in 5-minute windows before the EMAs were initiated (Fig. 1). We opted for a 5-minute initial window to balance capturing sufficient information, given the uncertainty of the exact occurrence of the physiological stress response, while limiting the potential contribution of noise [48]. Of the 2396 valid observations, 1893 items also contained simultaneously recorded physiological data. To account for potential artefacts within the physiological window, low-quality segments, defined as having a mean high-quality fraction lower than 50% for the PPG-derived HR and lower than 80% for EDA, were removed in 5-second windows. Windows with a fraction of high-quality sub-windows lower than 50%, for either the HR or EDA signal, were dropped (N = 826). Moreover, windows with vigorous activity were excluded before the modelling stage, as activity can cause physiological changes not necessarily related to stress [45]. More specifically, we removed windows containing any vigorous activity and more than 50% moderate activity (i.e., walking) (N = 92) resulting in a total of 975 data points (Table S4). If the window was not dropped, pre-defined statistical features were extracted from the high-quality data (Table 3). Finally, we confirmed that the physiological features lay in physiologically possible ranges, e.g., EDA-related features minimally 0 μS.

C. Data Analysis

To explore potential physiological biomarkers for pain in daily life, we first explored the variability of three pain-related EMA variables: pain intensity (rated on a numeric rating scale (NRS) from 0 to 10, N = 2396), pain unpleasantness (rated on a 7-point Likert scale, N = 2389), and pain impact on daily life (rated on a 7-point Likert scale, N = 2340). Repeated measure correlations, calculated by *rmcorr* function in R, revealed moderate associations between these variables: 0.68 (95% confidence interval [CI]: [0.66, 0.70]) between pain intensity and unpleasantness, 0.49 (95% CI: [0.46, 0.53]) between pain intensity and impact, and 0.56 (95% CI: [0.53, 0.59]) between pain unpleasantness and impact. Given the greater variability observed in the distribution of pain intensity compared to the other two variables (Fig. S4), we proceeded with this variable for further analysis. Fig. 2 shows the distributions of the reported momentary pain intensity on an 11-point NRS scale, as well as the corresponding levels of no (NRS: 0), mild (NRS: 1-3), moderate (NRS: 4-6), and high (NRS: 7-10) pain intensity.

1) Exploring Pain-Physiology Associations At Group Level: Next, we fitted several mixed-effect models to explore univariate ambulatory physiological biomarkers for pain in patients with CWP. Before modelling, we logarithmically transformed the EDA, tonic, phasic, and ST features to better fit a normal distribution. Next, we created within-participant-mean-centered variables for each biomarker, which we will further refer to as within-biomarkers, along with variables representing each participant's mean value, which we will further refer to as between-biomarkers. Finally, we standardized all variables to align their scales.

The independent variables in the models included the within-biomarker (participant-centered biomarker), the between-biomarker (average participant's biomarker), substance use in the last hour (yes/no), gender (woman, man, or transwoman), and a random intercept per participant. Since one of the women was transitioning and actively taking hormones, we classified her as a separate gender category due to potential physiological impact.

We also investigated the impact of two alternative scaling methods and the size of the extracted physiological window

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

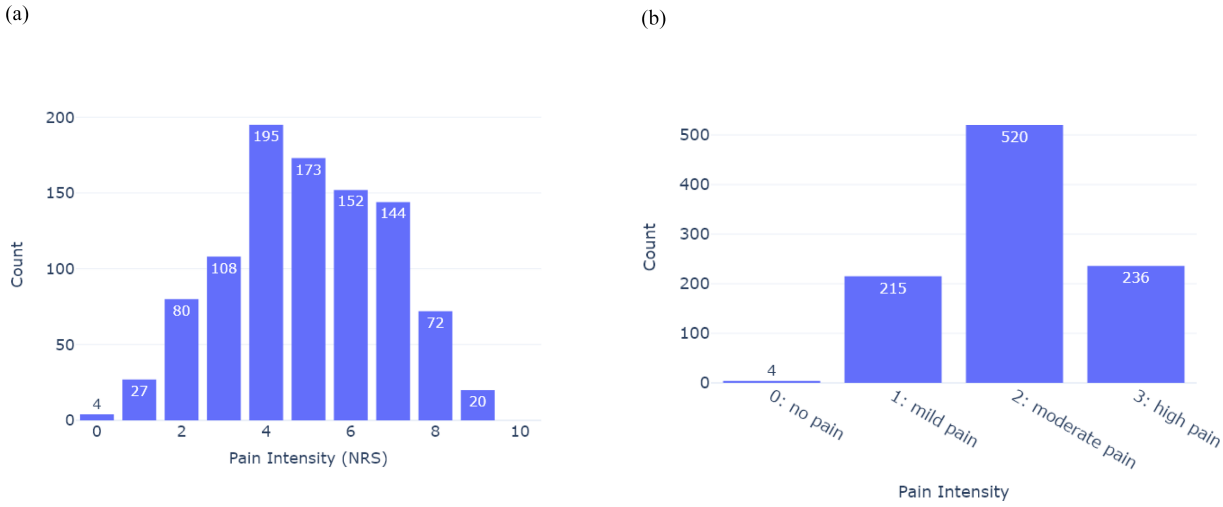

Fig. 2. Pain intensity distribution for patients, (a) 11-point numeric rating scale (NRS) scale, and (b) pain intensity classes.

on the linear mixed model outcomes using the same quality implementation pipeline as previous. First, we examined extracting features from standardized physiological signals relative to their weekly average, computed between 6 a.m. and midnight, which we assumed to be a simplified representation of baseline physiology. In these models, no within- and between-biomarkers were created as the signals were already standardized within participants. Second, we analyzed the impact of extracting features from standardized physiological signals relative to the hourly average, which we assumed to be a simplified correction for the diurnal rhythm present in some physiological measures, such as HR. Additionally, we evaluated the impact of different physiological window sizes preceding the EMA prompt, comparing 5, 10, and 30-minute intervals (Tables S4-S6). Estimates with corresponding p-values lower than 0.05 were considered significant.

2) Predicting Pain At Individual Level: To evaluate the predictive performance of physiological and EMA features for pain at individual level on unseen data, several regression models were trained. These models incorporated both EMA features (negative affect, positive affect, catastrophizing, physical exhaustion, momentary activity, substance intake, sleep quality, and how well rested they felt upon awakening), and pre-defined physiological features (Table S7). The dataset included the physiological measurements taken in 5-minute windows before the EMA prompts, standardized relative to each participant's weekly average. The constructs for catastrophizing, positive, and negative affect were developed and validated as detailed in the supplementary materials (Tables S2-S3 and Fig. S3). Furthermore, missing values in reported sleep (N = 6), catastrophizing (N = 24), phasic (N = 38), and tonic (N = 38) were imputed using the participant's average values. For pain intensity prediction, we considered the following pain levels: no pain (NRS: 0) was assigned to 0, mild pain (NRS: 1-3) to 1, moderate pain (NRS: 4-7) to 2, and high pain (NRS: 7-10) to 3.

First, the data was split iteratively into training and test sets using leave-one-participant-out cross validation, ensuring that each participant served once as the test set. Data preprocessing involved identifying and removing highly correlated features in the training data using a Spearman correlation coefficient threshold of 0.8 by retaining the feature which was most correlated with pain. The resulting features were standardized. Given the relatively small dataset, a Recursive Feature Elimination with Cross-Validation (RFECV) step using a Random Forest (RF) classifier with 30 estimators was applied to the training data to select the most relevant features for the model [48]. Next, we trained several regression models optimizing the hyperparameters through 3-fold-cross validation using grid search, optimizing the negative mean square error (Table IV). To assess the model's performance, we computed several metrics, including mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and the coefficient of determination (R^2).

TABLE IV
CONSIDERED HYPERPARAMETERS PER REGRESSION MODEL

Model type	Hyperparameters
Random Forest regressor (RF)	criterion: squared error or Poisson minimum samples per split: 2, 4, or 8 number of estimators: 20, 50, 100, or 200
XGBoost regressor (XGB)	Learning rate: 0.01, 0.05, or 0.1 Maximum depth: 2, 4, or 8 Number of estimators: 20, 50, 100, 200
Supported Vector regressor (SVR)	Kernel: linear, rbf, or sigmoid C: 0.1, 1, 10, 100, 1000 Gamma: 0.0001, 0.001, 0.01, 0.1, or 1
k- Nearest Neighbor regressor (kNN)	Number of neighbors: 3, 5, 7, 9, or 11
Elastic Net regressor (EN)	L1 ratio: 0, 0.25, 0.5, 0.75, or 1
Gaussian Process regressor (GP)	Kernel: WhiteKernel or RBF kernel

Next, we retrained the best-performing regression model, based on the MAE, using the dataset containing only EMA features to assess the contribution of physiological variables besides EMA-related variables for the prediction of pain. Changes in the model performance, measured by MAE per participant, against

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

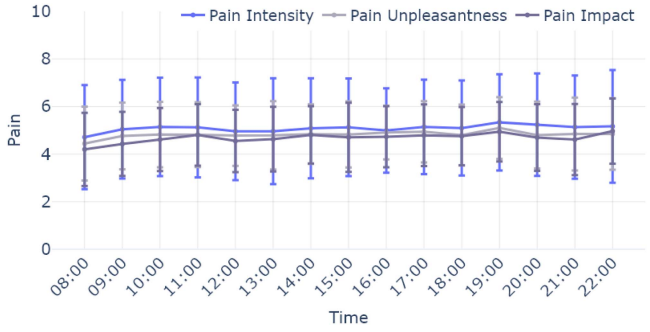

Fig. 3. Average reported pain with standard deviation. The plot includes all valid EMAs rounded to the nearest hour.

the model trained on physiological and EMA-features, were statistically tested using Wilcoxon signed rank tests. Subsequently, we evaluated if incorporating demographic variables would further enhance pain prediction. The added demographic variables included: age, gender, ethnicity, body mass index (BMI), blood pressure, smoking status, duration of pain complaints (months), highest educational degree, all questionnaire total scores, and psychiatric comorbidities as obtained by the Mini-International Neuropsychiatric Interview - simplified. Finally, we determined the 10 best-performing features using permutation importance. The permutation importance was evaluated using 10 repeats on the test data in every fold, after which the results were averaged per feature. For any feature not selected during model training, the permutation importance was set to 0. The ten features with the highest average impact were examined.

IV. Results

A. Data Collection

The sample consisted of 46 patients with a median age of 45 years (IQR: 17, range: 20-60). The gender distribution included 38 women and 8 men. The median BMI was 27.7 (IQR: 6.6, range: 19.4-47.9). Additionally, 43.8% of the patients were actively working, and 15% were active smokers. The median reported duration of pain complaints was 10 years (IQR: 10, range: 1-38). Patients took several types of medication, including benzodiazepines (N = 8), opioids (N = 19), and other pain medication (N = 34). A more detailed overview of all demographic variables and questionnaire scores can be found in Table S8.

B. EMA Descriptives

The average reported pain on an 11-point NRS scale per participant was 5.0 (std: 1.7) and the average standard deviation within a participant was 1.1 (std: 0.4). For pain unpleasantness and impact on daily life, the averages were 4.8 (std: 0.8) and 4.7 (std: 0.9), respectively, and average standard deviations were 1.0 (std: 0.4) and 0.8 (std: 0.4), respectively. The reported pain intensity did not significantly change with increasing beep number ($p = 0.185$), study day ($p = 0.129$), or their interaction ($p = 0.711$) (Fig. 3, model estimates: Table S10). The descriptives of the other EMA variables can be found in Table S11.

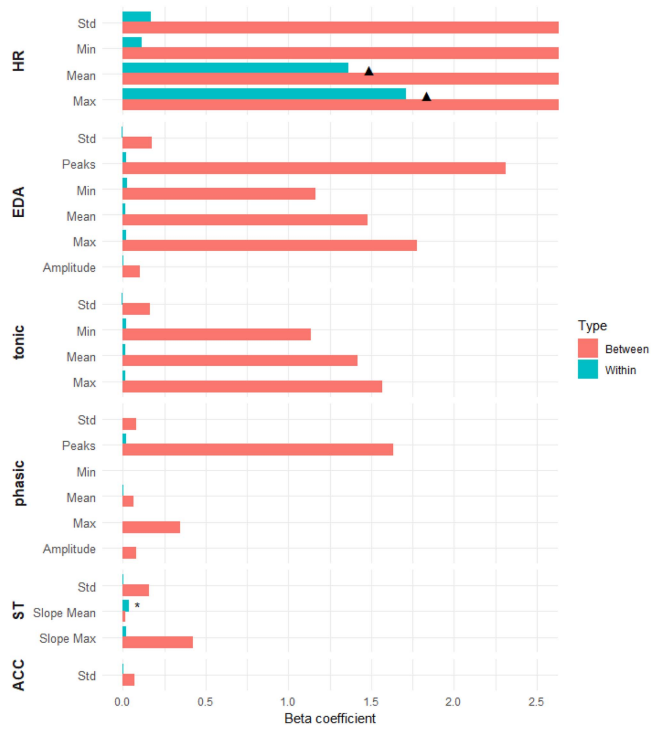

Fig. 4. Beta-coefficients for the unscaled data in 5 minutes of the within and between estimates for each of the biomarkers, if no substances were taken, with an indication of the significance, i.e., $p<0.001$ ■, $p<0.01$ ▲, $p<0.05$ ∗. The range of the x-axis is set to 2.5 and the between estimates for HR are 82, 61, 106, and 7, respectively.

C. Exploring Pain-Physiology Associations At Group Level

We started with exploring potential biomarkers for daily-life pain using linear mixed effect models based on the momentary physiological biomarkers five minutes before the pain intensity rating during moments of limited activity with gender and substance as covariates. Since the main effect of substance use was significant in all the models and had an impact on multiple physiological biomarkers, we also included substance use as an interaction effect with the within-biomarker (model without substances: Table S12 and Fig. S5). Fig. 4 shows the results of the within and the between estimates, which were backtransformed (b_t) to their original units for interpretation purposes, as well as the associated p-values. The within coefficients indicate the estimated effect on pain intensity if the within biomarker, extracted in a 5-minute window, increases with one unit compared to the patient's average while the confounding factors are held constant. The between coefficients indicate the estimated effect on average reported pain intensity for a participant for which the between biomarker increases with one unit compared to the population average.

In all the models the main effect of substances was significant ($b = 0.2$, $se = 0.09$, $t = 2.2$-2.6, $p = 0.01$-0.04) while there was no significant difference between genders (Table S13). Moreover, within mean HR was a positive predictor for pain intensity ($b_t = 1.4$, $b = 0.15$, $se = 0.05$, $t = 3.2$, $p = 0.001$), as well as maximum HR ($b_t = 1.7$, $b = 0.13$, $se = 0.05$, $t = 2.7$, $p = 0.007$), and mean ST slope per minute ($b_t = 0.03$, $b = 0.18$, $se = 0.07$, $t = 2.4$, $p = 0.02$) although the estimated effect of the

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

latter on pain is small. None of the between-biomarker estimates were found significant.

1) Alternative Scaling Methods: Subsequently, we standardized the signals with respect to the participant's weekly average (Fig. S6). The mean HR ($b_t = 0.2$, $b = 0.14$, $se = 0.05$, $t = 2.7$, $p = 0.006$), maximum HR ($b_t = 2.1$, $b = 0.12$, $se = 0.05$, $t = 2.3$, $p = 0.02$), and mean ST slope per minute ($b_t = 0.03$, $b = 0.11$, $se = 0.05$, $t = 2.3$, $p = 0.02$) remained significant predictors of pain intensity. Moreover, the maximum of EDA ($b_t = 3.1$, $b = -0.10$, $se = 0.05$, $t = -2.0$, $p = 0.04$) was now significant. Additionally, several within phasic features were also significant: mean ($b_t = 0.5$, $b = 0.10$, $se = 0.05$, $t = 2.1$, $p = 0.04$) and standard deviation ($b_t = 0.2$, $b = 0.11$, $se = 0.05$, $t = 2.1$, $p = 0.04$).

When standardizing with respect to the hour of the day, mean HR ($b_t = 0.02$, $b = 0.11$, $se = 0.05$, $t = 2.2$, $p = 0.03$) was again a significant predictor of pain intensity though the estimate was much smaller than when not standardized (Fig. S6). Moreover, maximum HR ($p = 0.28$) was not a significant predictor anymore. Finally, mean ST slope per minute ($b_t = 0.01$, $b = 0.11$, $se = 0.05$, $t = 2.1$, $p = 0.03$) also remained a significant predictor.

2) Alternative Window: Next, we compared estimates for the unscaled models if features were extracted in 10 and 30 minutes before the EMA instead of 5 minutes (Fig. S7). For a feature extraction window size of 10 minutes, the within mean HR ($b_t = 1.6$, $b = 0.19$, $se = 0.05$, $t = 4.0$, $p = <0.001$), maximum HR ($b_t = 2.4$, $b = 0.18$, $se = 0.05$, $t = 3.9$, $p = <0.001$), and mean ST slope per minute ($b_t = 0.03$, $b = 0.21$, $se = 0.07$, $t = 3.2$, $p = 0.001$) remained significant predictors of pain intensity. Moreover, the within standard deviation of HR ($b_t = 0.3$, $b = 0.12$, $se = 0.05$, $t = 2.5$, $p = 0.01$) and standard deviation of ACC magnitude ($b_t = 0.006$, $b = 0.13$, $se = 0.05$, $t = 2.7$, $p = 0.006$) were found significant predictors of pain intensity. Again, no between estimates were found significant.

For a feature extraction window size of 30 minutes, within mean HR ($b_t = 1.7$, $b = 0.22$, $se = 0.06$, $t = 3.7$, $p = <0.001$), maximum HR ($b_t = 2.2$, $b = 0.17$, $se = 0.06$, $t = 3.0$, $p = 0.003$), standard deviation of HR ($b_t = 0.3$, $b = 0.14$, $se = 0.06$, $t = 2.3$, $p = 0.02$), and standard deviation of ACC magnitude ($b_t = 0.006$, $b = 0.13$, $se = 0.05$, $t = 2.7$, $p = 0.006$) remained significant (Fig. S7). Within mean ST slope per minute was no longer significant ($p = 0.07$). In contrast to the previous models, the minimum of the tonic component ($b_t = 0.08$, $b = 0.16$, $se = 0.06$, $t = 2.6$, $p = 0.009$) as well as the maximum of the tonic component ($b_t = 0.08$, $b = 0.13$, $se = 0.06$, $t = 2.0$, $p = 0.046$) were significant predictors. Finally, the between ST maximum slope per minute was a significant predictor of pain intensity ($b_t = 0.23$, $b = -0.49$, $se = 0.23$, $t = -2.1$, $p = 0.04$).

D. Predicting Pain At Individual Level

Table V presents the median and interquartile ranges of the leave-one-patient-out test performance indicators for all models trained on both the physiological and EMA features. The EN regression model achieved the lowest MAE of 0.49 with the average pain intensity being 2 (moderate pain). The lowest MAPE of 0.29 was achieved by both the EN and support vector regression (SVR). Meanwhile, the gradient boosted decision tree (XGB) regressor attained the lowest RMSE of 0.64. The highest R^2 was -0.4 for the XGB and SVR. The residuals versus predicted test-scores illustrate the tendency of the model to predict the most frequent pain level, which is moderate pain (Fig. S10).

TABLE V
MEDIAN AND IQR OF THE TEST REGRESSION SCORES ACROSS PARTICIPANTS

Model	MAE	MAPE	RMSE	R^2
Random forest (RF)	0.56 (0.48)	0.32 (0.18)	0.70 (0.42)	-0.78 (3.1)
Gradient boosted decision tree (XBG)	0.54 (0.50)	0.31 (0.30)	**0.64 (0.44)**	**-0.40 (2.8)**
Supported vector machine regression (SVR)	0.50 (0.50)	**0.29 (0.37)**	0.67 (0.42)	**-0.40 (3.1)**
k-nearest neighbor (kNN)	0.59 (0.39)	0.31 (0.29)	0.73 (0.34)	-0.86 (2.7)
Elastic net (EN)	**0.49 (0.53)**	**0.29 (0.32)**	0.65 (0.44)	-0.38 (3.0)
Gaussian process (GP)	0.53 (0.49)	0.31 (0.26)	0.67 (0.45)	-0.82 (2.7)

Fig. 5 shows the test outcomes of the EN regression models, which were retrained using only the EMA-related features (light grey) against the models trained on both physiological and EMA-features (blue). Dropping the physiological features from the model did not significantly impact the performance indicators. Moreover, adding demographic-related features into the model trained on both EMA and physiological features resulted in a slightly lower median MAE and MAPE, though not significantly. The top ten predictive features of the EN model, as indicated by the permutation importance on the test set, mainly consisted of EMA-related variables, such as catastrophizing and positive affect. Although some physiological features were also included, they contributed only a limited median decrease in negative mean square error.

V. DISCUSSION

This study explored the associations between physiological signals, measured in daily life, and reported momentary pain within 46 patients with chronic widespread pain. To limit the effect of artefacts on the results, we implemented quality indicators and focused only on data from periods of limited activity, despite this substantially reducing dataset size. We found that regular self-reporting of pain did not affect the reported pain intensity over the day. Linear mixed model results indicated several significant associations between within-patient physiological features and pain intensity on average. Additionally, our results stressed the impact of preprocessing choices such as window size

A. Exploring Pain-Physiology Associations At Group Level

We found significant positive associations between within-participant mean and maximum HR over 5-minutes and momentary reported pain intensity. These associations were also

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

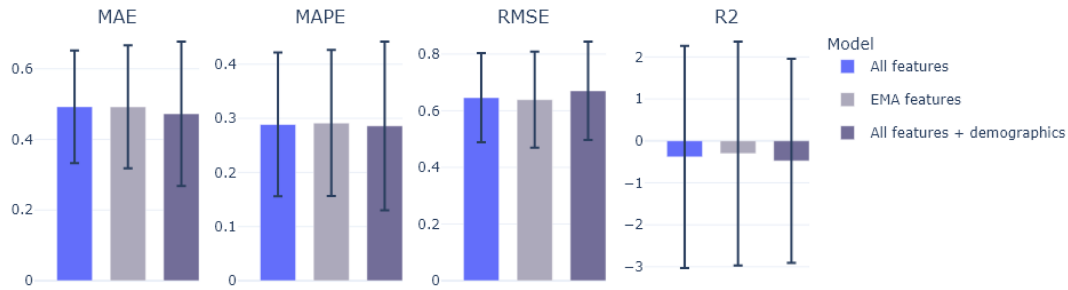

Fig. 5. Median and interquartile ranges of the performance indicators for models trained on all features (physiological and EMA), only EMA features, and only all features and demographic features, respectively. Stars indicate significant outcomes of Wilcoxon signed rank tests against the models containing EMA and physiological features.

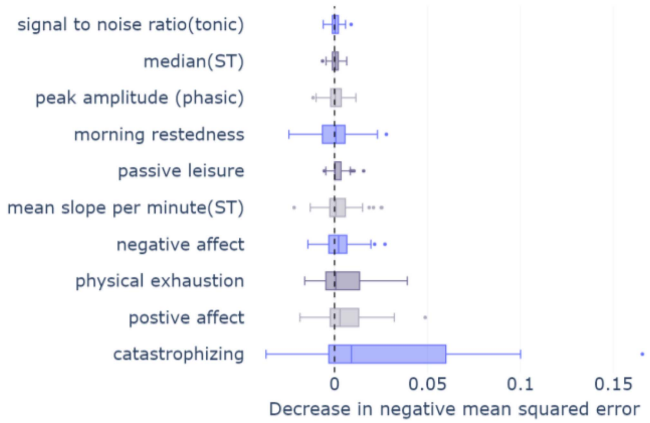

Fig. 6. Permutation importance of the top 10 features with the highest median predictive power on the test set.

present in 10- and 30-minute windows, in which the within standard deviation was also found significant. These associations disappeared when HR was standardized with respect to the hour of the day. Since no diurnal variations were observed in reported pain intensity (Fig. 3), standardizing by the hour likely reduced HR variance (Fig. S8). These observations suggest that HR and pain intensity are prolongedly associated in CWP patients, consistent with the idea that the pain variability in these patients is likely gradually changing [49]. Possibly, increases in clinical chronic pain may result in a prolonged hyporeactive sympathetic nervous system response. Previous studies reported significant associations between pain intensity and HR in both patients with CWP [49], [50], [51] and healthy individuals [10], [52], [53] in laboratory settings, although not consistently [52], [54]. For example, Treister et al. (2021) [54] reported that only the linear combination of different modalities, i.e., HR variability high-frequency component, HR, number of EDA fluctuations, EDA level, and PPG amplitude, could discriminate pain intensity levels induced by three different phasic heat pain intensities in 45 healthy individuals. However, it is difficult to directly compare the results to this study as none of these studies examined ambulatory-assessed pain intensity.

We did not find significant estimators of momentary pain intensity level in the unscaled 5 and 10-minute models for both EDA, tonic, and phasic. However, when using a 30-minute window the within-participant minimum and maximum tonic were significant predictors. Moreover, when standardizing with respect to the participant's weekly average, several within-participant phasic-related biomarkers and the maximum EDA became significant. The significance of the tonic components related to range suggests that extreme values might be insightful in longer windows. These findings align with the prolonged hyporeactive response indicated by the HR-related findings. In contrast to the prolonged responses seen with tonic features, several significant phasic biomarkers emerged when data was standardized with respect to the participant's weekly average, suggesting immediate pain-physiological interactions. It is possible that by standardizing by the weekly average, the phasic signal for participants with small variations was stretched, making detecting responses more likely (Fig. S9). Previous studies have reported mixed results including significant associations between pain intensity and EDA level [55], [56], EDA responses [10], and no significant associations [55], [57]. For example, Posada-Quintero et al. (2021) [57], found no significant EDA-related features discriminating between three induced pain stimuli levels and no pain in 23 healthy individuals exposed to a thermal grill for maximally 10 seconds per stimulus. However, the phasic EDA was found to be significantly different between no pain and high pain and the tonic EDA differed between mild pain and no pain. These observed differences might be explained by the more transient stimuli and shorter observation windows. In line with our results, the study by Scheuren et al. (2020) [58] demonstrated that inducing secondary hyperalgesia significantly affected EDA levels. Secondary hyperalgesia is characterized by enhanced sensitivity to pain in the areas surrounding the site tissue injury, and hyperalgesia is a symptom of CWP. Again, direct comparisons with other studies are challenging since most studies were executed in laboratory settings or with healthy individuals. Interestingly, Sugimine et al. (2020) [56] suggested that the physiological stress response intensity, as indicated by the EDA level, might relate to pain stimulus intensity rather than pain intensity. However, Mischkowski et al. (2018) [59] indicated that pain intensity could modulate this relationship. Unfortunately, our study cannot discriminate between the strength of nociception and pain intensity, and it is therefore impossible in this study to validate these findings in patients with CWP.

We found a consistently significant association between the average ST slope per minute and the reported pain intensity

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

aligning with the prolonged hyporeactive response. However, the estimated association was too small to be clinically relevant. Consistent with our results, Sano et al. (2018) [60] found ambulatory wrist ST to be a predictor for perceived stress in healthy individuals. Moreover, Vinkers et al. (2013) [12], found indications for reduced ST at the wrist, after mental stress induction, despite no statistical significance. This reversed relationship, compared to ours, could potentially be attributed to the controlled versus ambulatory conditions, in which variations in ambient temperature might also influence ST changes. For instance, increased pain could be associated with higher activity levels or pain behaviors, such as napping, which may be linked with ambient conditions [61].

Even after correction for the activity intensity, we still found certain significant associations between pain intensity and the standard deviation of the accelerometer's magnitude in 10- and 30-minute windows. However, the estimates were small (almost 0). These small effects in longer windows might still indicate the impact of minor activity on pain. Previous research has reported heterogeneity in the activity-pain association but has not yet investigated such small windows [62]. For example, Alschuler et al. (2011) [61] monitored 20 patients with low-back pain for 5 days and did not find significant pain-activity associations within 4-hour windows.

Finally, none of the between-participant estimates were found to be significant, except the maximum slope of ST per minute in 30 minutes-windows, although the estimate was small and thus not clinically relevant. These findings contrast with the study of Loggia et al. (2011) [10], which reported HR to be a significant between participant indicator of pain. It can be assumed that our sample size was too limited to detect HR-related between-participant differences as our study was designed to maximize within-participant power.

The observed associations between momentary pain intensity and physiological variables at group level highlight the potential of physiological monitoring as a complementary pain monitoring methodology, which could potentially unveil underlying pathophysiological mechanisms and allow evaluation of treatment efficacy within patients with CWP.

B. *Predicting Pain At Individual Level*

Next, we trained six different types of regression models to assess the predictive performance of physiological and EMA-related features for pain intensity using leave-one-patient-out cross-validation for generalizability. The EN regression model yielded the lowest median MAE of 0.49, while the XGBoost model achieved the lowest median RMSE of 0.69 with the average reported pain-level being two (moderate pain). These MAE and RMSE performance scores align with previously reported numbers for multilevel regression in laboratory studies (Table I). For instance, Walter et al. (2015) [24] reported an MAE of 1.05 and an MSE of 1.29, for their best-performing model predicting thermal pain intensity using EDA and ECG-related features in the BIOVID dataset containing 87 healthy individuals. Their model was based on a balanced set of 80 pain stimuli given at pain threshold, pain tolerance, and two intermediate levels. However, their coefficient of determination of 0.24 is higher than ours, suggesting a better fit. Nevertheless, direct comparisons were difficult because of the different number of pain levels. Johnson et al. (2019) [35] conducted an ambulatory study and trained regression models to predict pain intensity in 20 patients with sickle cell anemia during vaso-occlusive crises, using HR, EDA, and motion-related features. They obtained an RMSE of 1.4 and a Pearson correlation of 0.71 when predicting pain intensity across 11 levels with the majority of reported pain-level being between 5 and 8. This study obtained lower predictive performances (Fig. S9) compared to similar research in other pain conditions, possibly because of the distinct dysfunctions of the physiological stress system or diagnostic criteria unique to CWP [20]. However, the obtained results are unique to CWP, characterized by widespread pain and indications for nociplastic pain, and direct comparisons with other chronic pain conditions are challenging. The residual versus predicted plot (Fig. S9) shows that while our model captures some meaningful information regarding pain, it only accounts for a modest portion of the observed variance. One possible explanation is the limited variability in pain intensity levels within some participant's test sets, which may not be adequately reflected in the physiological features, resulting in a potential mismatch between the training and test conditions. Additionally, our model never predicted 'no pain', likely due to its occurrence in less than 1% of the data, which suggests that the trained model might encounter challenges in differentiating between features associated with extreme pain levels compared to those with average pain values [38].

Moreover, we found that removing the physiological features or adding the demographic features did not significantly impact the performance indicators. This suggests that physiological features trained on data from other patients may not generalize well to predict pain intensity in unseen patients. However, a few physiological features, such as mean temperature slope per minute, mean phasic response amplitude, median temperature, and signal-to-noise ratio in the tonic component, were among the top 10 predictive features, albeit with low median impact on the mean squared error. Interestingly, these features do not correspond to the significant biomarkers, except mean ST slope, found in the statistical models. This discrepancy might be due to an imbalance in the amount of variance in the data of some participants. Additionally, some features may have strong predictive power for certain individuals but not for others, leading the RFECV algorithm to discard them. These results align with Krause et al. (2023) [63], who observed only a slight performance increase when adding physiological features to a resilience model only trained on EMA features. In contrast with our findings, the study of Olugbade et al. (2015) [64], showed that the performance of the RF model improved with the addition of depression-related features, whereas the performance of the Support Vector Machine model did not.

C. *Limitations and Future Work*

Though this work strived to explore the associations and predictive ability between physiological features and pain in the

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

daily life of patients with CWP, there were several limitations. First, our sample predominantly consisted of white women, limiting the generalizability to other patient populations but represents the gender imbalance in CWP [65]. For instance, racial differences in pain sensitivity have been documented [66], so future studies should include a more diverse population. Moreover, this work focused on momentary associations and did not account for potential temporal dynamics in the data. For example, research has demonstrated that increased pain can influence negative affect (but not positive affect) on the following day [67]. Accounting for these relationships could improve the performance of pain prediction models. Third, the linear mixed model results should be interpreted carefully, given the complexity of both ambulatory pain and physiological measurements. For instance, the physiological stress response is not specific to pain, but can also be triggered by mental stress, physical activity, or thermal stimuli [68], [69]. While this study provides preliminary support for the associations between physiological features and pain intensity in ecologically valid contexts, we executed exploratory analyses without controlling for multiple testing and adjusted the model structure after data inspection. Therefore, the results should be validated in larger samples. Although we adjusted for substance use within the last hour, other contextual variables influencing physiology, pain, or both in ambulatory conditions were not fully accounted for. Examples of contextual variables include ambient temperature affecting EDA and ST, specific substances or medications affecting physiological signals, or psychological factors like catastrophizing or attentional focus on the pain [16], [70]. The observed associations could be attributed to one of these unincluded variables. Furthermore, this is difficult to further investigate as we do not know the exact moment when the physiological stress responses occurred, or the pain intensity changed. Additionally, by dropping the random slope, we assumed that the relationship between physiology and pain was consistent across individuals, which should be further explored. Specifically, different relationships may emerge among patients with underlying pathophysiology that could be better explained by a central driven mechanism, i.e., central sensitization, compared to those whose pathophysiology might be better explained by peripheral mechanisms [71].

The predictive performance of the pain prediction models was insufficient to support practical applicability. Future work should further explore the capability of physiological signals to contribute to clinical pain assessment in patients with chronic widespread pain. Specifically, individual or personalized models should be explored based on clinically relevant parameters (such as psychiatric comorbidities or other demographic variables), subgroups of patients (such as patients who exhibit avoidance or persistence behavior towards activity as defined by Huijnen et al. (2011) [72]), or the physiological stress response in laboratory settings [73]. The limited complementary added value of physiological signals beyond EMA, should be further investigated using larger windows as they may capture greater variability. Additionally, extending the overall monitoring duration could similarly reveal more variation and could be informative about the treatment effect. For example, Patterson et al. (2023) [73] showed that physiological biomarkers can be used to follow treatment outcomes in patients with chronic pain who were followed for six months. To increase knowledge regarding the temporal relationship between stress physiology and pain the use of more advanced modelling techniques, such as Long Short-Term Memory (LSTM) networks should be explored, or potentially more questionnaires could be prompted on days with more stress or pain. Moreover, the findings on the added value of demographic features in the pain prediction models should be interpreted with caution, as the sample size and balance of some demographic variables (e.g., gender) were limited. Next, this study focused on exploring the added value of physiological features in pain prediction models using ambulatory data, opting for simple features given the complexity of the data, potentially containing remaining artifacts or low-quality segments. Future work could incorporate more complex features (e.g., frequency based), which may offer greater sensitivity for pain [27]. Furthermore, integrating other features, such as related to environmental conditions or location, could improve pain prediction. Finally, this work used RFECV to preselect features, a method that can be still prone to overfitting. Future research could benefit from implementing other feature selection methods (e.g., factor or principal component analysis), regularization techniques (e.g., lasso regression), or more advanced models (e.g., deep learning with data augmentation).

VI. Conclusion

This study aimed to investigate the potential of ambulatory-collected physiological signals for assessing momentary pain in individuals with chronic widespread pain. This study provided complementary findings at both the group- and individual-level, obtained from data collected in ambulatory settings, unique to a clinical population of patients with chronic widespread pain. Our findings revealed several significant physiological biomarkers at group level, such as mean heart rate, which were associated with the reported momentary pain intensity, highlighting the potential of physiological monitoring to unveil underlying pathophysiological mechanisms and evaluate treatment efficacy.

The inherent complexity of both pain and ambulatory measurements required focusing on valid, high-quality physiological data during limited activities, despite substantially reducing the dataset size. Our study found limited added value of physiological features in predicting pain intensity at individual level suggesting that future work should focus on personalized models. Additionally, the limited performance of the pain prediction models underscores the ongoing challenge in the field, revealing only weak associations.

As this study presents the first exploration of pain physiology associations and predictive capabilities in ecologically valid conditions for patients with chronic widespread pain, further validation of our findings is needed before clinical implementation.

Acknowledgment

The authors would like to thank Marlies Houben for her contribution to the development of the EMA questionnaire and

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

Kristof Vansteenlandt for his statistical expertise. Furthermore, we would like to thank all the study participants for their time and effort as well as the clinical staff of the hospital who supported the study. OnePlanet Research Center acknowledges financial support from the Province of Gelderland.

References

[1] International Association for the Study of Pain, "Terminology | international association for the study of pain," 2020. [Online] Available: https://www.iasp-pain.org/resources/terminology/

[2] P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, and R. W. Picard, "Automatic recognition methods supporting pain assessment: A survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 530–552, First Quarter 2022, doi: 10.1109/TAFFC.2019.2946774.

[3] M. Matsangidou, A. Liampas, M. Pittara, C. S. Pattichi, and P. Zis, "Machine learning in pain medicine: An up-to-date systematic review," *Pain Ther.*, vol. 10, no. 2, Dec. 2021, Art. no. 1067, doi: 10.1007/s40122-021-00324-2.

[4] R. F. Rojas, N. Brown, G. Waddington, and R. Goecke, "A systematic review of neurophysiological sensing for the assessment of acute pain," *npj Digit. Med.*, vol. 6, no. 1, pp. 1–25, Apr. 2023, doi: 10.1038/s41746-023-00810-1.

[5] M. Gendreau, M. R. Hufford, and A. A. Stone, "Measuring clinical pain in chronic widespread pain: Selected methodological issues," *Best Pract. Res.: Clin. Rheumatol.*, vol. 17, no. 4, pp. 575–592, Aug. 2003, doi: 10.1016/s1521-6942(03)00031-7.

[6] P. Thiam et al., "Multi-modal pain intensity recognition based on the SenseEmotion database," *IEEE Trans. Affect. Comput.*, vol. 12, no. 3, pp. 743–760, Jul. 2019, doi: 10.1109/TAFFC.2019.2892090.

[7] P. Werner, A. Al-Hamadi, S. Gruss, and S. Walter, "Twofold-multimodal pain recognition with the X-ITE pain database," in *Proc. 8th Int. Conf. Affect. Comput. Intell. Interaction Workshops Demos*, Cambridge, U. K., 2019, pp. 290–296, doi: 10.1109/ACIIW.2019.8925061.

[8] I. Myin-Germeys and P. Kuppens, "The open handbook of experience sampling methodology a step-by-step guide to designing, conducting, and analyzing ESM studies," Leuven, The center for Research on Experience sampling and Ambulatory methods Leuven REAL, 2021. [Online]. Available: https://www.kuleuven.be/samenwerking/real/real-book

[9] B. N. Kyle and D. W. McNeil, "Autonomic arousal and experimentally induced pain: A critical review of the literature," *Pain Res. Manage.*, vol. 19, no. 3, pp. 159–167, 2014, doi: 10.1155/2014/536859.

[10] M. L. Loggia, M. Juneau, and M. C. Bushnell, "Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity," *Pain*, vol. 152, no. 3, pp. 592–598, Mar. 2011, doi: 10.1016/j.pain.2010.11.032.

[11] H. Storm, "Changes in skin conductance as a tool to monitor nociceptive stimulation and pain," *Curr. Opin. Anesthesiol.*, vol. 21, no. 6, pp. 796–804, Dec. 2008, doi: 10.1097/ACO.0b013e3283183fe4.

[12] C. H. Vinkers et al., "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, pp. 520–530, Sep. 2013, doi: 10.3109/10253890.2013.807243.

[13] P. Andrews, M. Steultjens, and J. Riskowski, "Chronic widespread pain prevalence in the general population: A systematic review," *Eur. J. Pain*, vol. 22, no. 1, pp. 5–18, Jan. 2018, doi: 10.1002/ejp.1090.

[14] ICD-11, "MG30.0 chronic primary pain," Jan. 2025. [Online]. Available: https://icd.who.int/browse11/l-m/en#/http://id.who.int/icd/entity/1326332835

[15] J. Lee, B. Ellis, C. Price, and A. P. Baranowski, "Chronic widespread pain, including fibromyalgia: A pathway for care developed by the British Pain Society," *Brit. J. Anaesth.*, vol. 112, pp. 16–24, 2014, doi: 10.1093/bja/aet351.

[16] T. A. Ajayi, L. Salongo, Y. Zang, N. Wineinger, and S. Steinhubl, "Mobile health-collected biophysical markers in children with serious illness-related pain," *J. Palliat. Med.*, vol. 24, no. 4, pp. 580–588, Apr. 2021, doi: 10.1089/jpm.2020.0234.

[17] Y. Tousignant-Laflamme, P. Rainville, and S. Marchand, "Establishing a link between heart rate and pain in healthy subjects: A gender effect," *J. Pain*, vol. 6, no. 6, pp. 341–347, Jun. 2005, doi: 10.1016/j.jpain.2005.01.351.

[18] J. Koenig, A. Loerbroks, M. N. Jarczok, J. E. Fischer, and J. F. Thayer, "Chronic pain and heart rate variability in a cross-sectional occupational sample evidence for impaired vagal control," *Clin. J. Pain*, vol. 32, no. 3, pp. 218–225, 2016, doi: 10.1097/AJP.0000000000000242.

[19] P. Chalaye et al., "Comparing pain modulation and autonomic responses in fibromyalgia and irritable bowel syndrome patients," *Clin. J. Pain*, vol. 28, no. 6, pp. 519–526, Jul. 2012, doi: 10.1097/AJP.0b013e31823ae69e.

[20] E. Coppens et al., "Cortisol and subjective stress responses to acute psychosocial stress in fibromyalgia patients and control participants," *Psychosomatic Med.*, vol. 80, no. 3, pp. 317–326, Apr. 2018, doi: 10.1097/PSY.0000000000000551.

[21] H. Van Middendorp, M. A. Lumley, J. H. Houtveen, J. W. G. Jacobs, J. W. J. Bijlsma, and R. Geenen, "The impact of emotion-related autonomic nervous system responsiveness on pain sensitivity in female patients with fibromyalgia," *Psychosomatic Med.*, vol. 75, no. 8, pp. 765–773, 2013, doi: 10.1097/PSY.0B013E3182A03973.

[22] A. Wyns et al., "The biology of stress intolerance in patients with chronic pain-state of the art and future directions," *J. Clin. Med.*, vol. 12, no. 6, Mar. 2023, Art. no. 2245, doi: 10.3390/JCM12062245.

[23] S. Gruss et al., "Pain intensity recognition rates via biopotential feature patterns with support vector machines," *PLoS One*, vol. 10, no. 10, Oct. 2015, Art. no. e0140330, doi: 10.1371/journal.pone.0140330.

[24] S. Walter et al., "Data fusion for automated pain recognition," in *Proc. 9th Int. Conf. Pervasive Comput. Technol. Healthcare PervasiveHealth*, 2015, pp. 261–264. doi: 10.4108/icst.pervasivehealth.2015.259166.

[25] D. Lopez-Martinez and R. Picard, "Continuous pain intensity estimation from autonomic signals with recurrent neural networks," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2018, pp. 5624–5627, doi: 10.1109/EMBC.2018.8513575.

[26] M. Jiang et al., "Acute pain intensity monitoring with the classification of multiple physiological parameters," *J. Clin. Monit. Comput.*, vol. 33, no. 3, pp. 493–507, Jun. 2019, doi: 10.1007/S10877-018-0174-8.

[27] Y. Kong, H. F. Posada-Quintero, and K. H. Chon, "Sensitive physiological indices of pain based on differential characteristics of electrodermal activity," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 10, pp. 3122–3130, Oct. 2021, doi: 10.1109/TBME.2021.3065218.

[28] P. Gouverneur et al., "Explainable artificial intelligence (XAI) in pain research: Understanding the role of electrodermal activity for automated pain recognition," *Sensors*, vol. 23, no. 4, Feb. 2023, Art. no. 1959, doi: 10.3390/S23041959.

[29] P. Thiam, P. Bellmann, H. A. Kestler, and F. Schwenker, "Exploring deep physiological models for nociceptive pain recognition," *Sensors*, vol. 19, no. 20, Oct. 2019, Art. no. 4503, doi: 10.3390/S19204503.

[30] M. Kächele, P. Thiam, M. Amirian, F. Schwenker, and G. Palm, "Methods for person-centered continuous pain intensity assessment from bio-physiological channels," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 5, pp. 854–864, Aug. 2016, doi: 10.1109/JSTSP.2016.2535962.

[31] M. Kächele et al., "Adaptive confidence learning for the personalization of pain intensity estimation systems," *Evolving Syst.*, vol. 8, no. 1, pp. 71–83, Mar. 2017, doi: 10.1007/s12530-016-9158-4.

[32] F. Pouromran, S. Radhakrishnan, and S. Kamarthi, "Exploration of physiological sensors, features, and machine learning models for pain intensity estimation," *PLoS One*, vol. 16, no. 7, Jul. 2021, Art. no. e0254108, doi: 10.1371/JOURNAL.PONE.0254108.

[33] J. Pagán et al., "Robust and accurate modeling approaches for migraine per-patient prediction from ambulatory data," *Sensors*, vol. 15, no. 7, pp. 15419–15442, Jun. 2015, doi: 10.3390/S150715419.

[34] S. Moscato, S. Orlandi, A. Giannelli, R. Ostan, and L. Chiari, "Automatic pain assessment on cancer patients using physiological signals recorded in real-world contexts," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2022, pp. 1931–1934, doi: 10.1109/EMBC48229.2022.9871990.

[35] A. Johnson et al., "Use of mobile health apps and wearable technology to assess changes and predict pain during treatment of acute pain in sickle cell disease: Feasibility study," *JMIR Mhealth Uhealth*, vol. 7, no. 12, 2019, Art. no. e13671, doi: 10.2196/13671.

[36] R. S. Stojancic et al., "Predicting pain in people with sickle cell disease in the day hospital using the commercial wearable apple watch: Feasibility study," *JMIR Formative Res.*, vol. 7, 2023, Art. no. e45355, doi: 10.2196/45355.

[37] S. Taylor, N. Jaques, E. Nosakhare, A. Sano, and R. Picard, "Personalized multitask learning for predicting tomorrow's mood, stress, and health," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 200, Second Quarter 2020, doi: 10.1109/TAFFC.2017.2784832.

[38] B. M. Booth, H. Vrzakova, S. M. Mattingly, G. J. Martinez, L. Faust, and S. K. D'Mello, "Toward robust stress prediction in the age of wearables: Modeling perceived stress in a longitudinal study with information workers," *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 2201–2217, Second Quarter 2022, doi: 10.1109/TAFFC.2022.3188006.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.

[39] W. Boucsein, *Electrodermal Activity: Second Edition*. Boston, MA, USA: Springer, 2012. doi: 10.1007/978-1-4614-1126-0.
[40] M. Naegelin et al., "An interpretable machine learning approach to multimodal stress detection in a simulated office environment," *J. Biomed. Inform.*, vol. 139, Mar. 2023, Art. no. 104299, doi: 10.1016/J.JBI.2023.104299.
[41] S. D. Benning, R. L. Bachrach, E. A. Smith, A. J. Freeman, and A. G. C. Wright, "The registration continuum in clinical science: A guide toward transparent practices.," *J. Abnorm. Psychol.*, vol. 128, no. 6, pp. 528–540, Aug. 2019, doi: 10.1037/abn0000451.
[42] F. Wolfe et al., "The american college of rheumatology 1990 criteria for the classification of fibromyalgia," *Arthritis Rheumatol.*, vol. 33, no. 2, pp. 160–172, Feb. 1990, doi: 10.1002/ART.1780330203.
[43] J. A. Chalmers, D. S. Quintana, M. J. A. Abbott, and A. H. Kemp, "Anxiety disorders are associated with reduced heart rate variability: A meta-analysis," *Front. Psychiatry*, vol. 5, Jul. 2014, Art. no. 80, doi: 10.3389/FPSYT.2014.00080.
[44] M. Mestdagh et al., "m-path: An easy-to-use and flexible platform for ecological momentary assessment and intervention in behavioral research and clinical practice," *PsyArxiv*, 2022. [Online]. Available: https://www.researchgate.net/publication/358108356_m-Path_An_easy-to-use_and_flexible_platform_for_ecological_momentary_assessment_and_intervention_in_behavioral_research_and_clinical_practice
[45] E. Smets et al., "Large-scale wearable data reveal digital phenotypes for daily-life stress detection," *npj Digit. Med.*, vol. 1, no. 1, pp. 1–10, Dec. 2018, doi: 10.1038/s41746-018-0074-9.
[46] M. Stojchevska et al., "Assessing the added value of context during stress detection from wearable data," *BMC Med Inform Decis Mak*, vol. 22, no. 1, pp. 1–13, Dec. 2022, doi: 10.1186/S12911-022-02010-5/TABLES/5.
[47] B. A. Jaso, N. I. Kraus, and A. S. Heller, "Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring," *Psychol. Methods*, vol. 27, no. 6, pp. 958–981, Sep. 2022, doi: 10.1037/MET0000312.
[48] L. Luebke, P. Gouverneur, T. M. Szikszay, W. M. Adamczyk, K. Luedtke, and M. Grzegorzek, "Objective measurement of subjective pain perception with autonomic body reactions in healthy subjects and chronic back pain patients: An experimental heat pain study," *Sensors*, vol. 23, no. 19, Oct. 2023, Art. no. 8231, doi: 10.3390/s23198231.
[49] M. M. Nickel, E. S. May, L. Tiemann, M. Postorino, S. T. Dinh, and M. Ploner, "Autonomic responses to tonic pain are more closely related to stimulus intensity than to pain intensity," *Pain*, vol. 158, no. 11, pp. 2129–2136, Nov. 2017, doi: 10.1097/J.PAIN.0000000000001010.
[50] P. Chalaye et al., "Comparing pain modulation and autonomic responses in fibromyalgia and irritable bowel syndrome patients," *Clin. J. Pain*, vol. 28, no. 6, pp. 519–526, Jul. 2012, doi: 10.1097/AJP.0b013e31823ae69e.
[51] G. A. Reyes del Paso, S. Garrido, Á. Pulgar, and S. Duschek, "Autonomic cardiovascular control and responses to experimental pain stimulation in fibromyalgia syndrome," *J. Psychosomatic Res.*, vol. 70, no. 2, pp. 125–134, Feb. 2011, doi: 10.1016/J.JPSYCHORES.2010.09.012.
[52] K. Hamunen, V. Kontinen, E. Hakala, P. Talke, M. Paloheimo, and E. Kalso, "Effect of pain on autonomic nervous system indices derived from photoplethysmography in healthy volunteers," *Brit. J. Anaesth.*, vol. 108, no. 5, pp. 838–844, May 2012, doi: 10.1093/BJA/AES001.
[53] J.-J. Ye, K.-T. Lee, J.-S. Lin, and C.-C. Chuang, "Observing continuous change in heart rate variability and photoplethysmography-derived parameters during the process of pain production/relief with thermal stimuli," *J. Pain Res.*, vol. 10, pp. 527–533, Mar. 2017, doi: 10.2147/JPR.S129287.
[54] R. Treister, M. Kliger, G. Zuckerman, I. G. Aryeh, and E. Eisenberg, "Differentiating between heat pain intensities: The combined effect of multiple autonomic parameters," *Pain*, vol. 153, no. 9, pp. 1807–1814, Sep. 2012, doi: 10.1016/j.pain.2012.04.008.
[55] S. Geuter, M. Gamer, S. Onat, and C. Büchel, "Parametric trial-by-trial prediction of pain by easily available physiological measures," *Pain*, vol. 155, no. 5, pp. 994–1001, May 2014, doi: 10.1016/j.pain.2014.02.005.
[56] S. Sugimine, S. Saito, and T. Takazawa, "Normalized skin conductance level could differentiate physical pain stimuli from other sympathetic stimuli," *Sci. Rep.s*, vol. 10, no. 1, Jul. 2020, Art. no. 10950, doi: 10.1038/s41598-020-67936-0.
[57] H. F. Posada-Quintero, Y. Kong, and K. H. Chon, "Objective pain stimulation intensity and pain sensation assessment using machine learning classification and regression based on electrodermal activity," *Amer. J. Physiol.-Regulatory, Integrative Comp. Physiol.*, vol. 321, no. 2, pp. R186–R196, Aug. 2021, doi: 10.1152/AJPREGU.00094.2021/ASSET/IMAGES/LARGE/AJPREGU.00094.2021_F006.JPEG.
[58] P. S. Scheuren, J. Rosner, A. Curt, and M. Hubli, "Pain-autonomic interaction: A surrogate marker of central sensitization," *Eur. J. Pain*, vol. 24, no. 10, pp. 2015–2026, Nov. 2020, doi: 10.1002/EJP.1645.
[59] D. Mischkowski, E. E. Palacios-Barrios, L. Banker, T. C. Dildine, and L. Y. Atlas, "Pain or nociception? Subjective experience mediates the effects of acute noxious heat on autonomic responses," *Pain*, vol. 159, no. 4, pp. 699–711, Apr. 2018, doi: 10.1097/J.PAIN.0000000000001132.
[60] A. Sano et al., "Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study," *J. Med. Internet Res.*, vol. 20, no. 6, Jun. 2018, Art. no. e210, doi: 10.2196/JMIR.9410.
[61] K. N. Alschuler, F. Hoodin, S. L. Murphy, J. Rice, and M. E. Geisser, "Factors contributing to physical activity in a chronic low back pain clinical sample: A comprehensive analysis using continuous ambulatory monitoring," *Pain*, vol. 152, no. 11, pp. 2521–2527, Nov. 2011, doi: 10.1016/j.pain.2011.07.017.
[62] M. Tynan, N. Virzi, J. S. Wooldridge, J. L. Morse, and M. S. Herbert, "Examining the association between objective physical activity and momentary pain: A systematic review of Studies using Ambulatory Assessment," *J. Pain*, vol. 25, no. 4, pp. 862–874, Apr. 2024, doi: 10.1016/J.JPAIN.2023.10.021.
[63] F. Krause et al., "Predicting resilience from psychological and physiological daily-life measures," *OSF Preprints*, vol. 2023, Jul. 2023, doi: 10.31219/OSF.IO/F28UV.
[64] T. A. Olugbade, N. Bianchi-Berthouze, N. Marquardt, and A. C. Williams, "Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain," in *Proc. Int. Conf. Affect. Comput. Intell. Interaction*, 2015, pp. 243–249, doi: 10.1109/ACII.2015.7344578.
[65] K. Bimpong et al., "The gender pain gap: Gender inequalities in pain across 19 European countries," *Scand. J. Public Health*, vol. 50, no. 2, pp. 287–294, Mar. 2022, doi: 10.1177/1403494820987466.
[66] H. J. Kim et al., "Racial/ethnic differences in experimental pain sensitivity and associated factors – Cardiovascular responsiveness and psychological status," *PLoS One*, vol. 14, no. 4, Apr. 2019, Art. no. e0215534, doi: 10.1371/JOURNAL.PONE.0215534.
[67] H. W. Mak and S. Schneider, "Individual differences in momentary pain-affect coupling and their associations with mental health in patients with chronic pain," *J. Psychosomatic Res.*, vol. 138, Nov. 2020, Art. no. 110227, doi: 10.1016/j.jpsychores.2020.110227.
[68] N. R. Bossenger, G. N. Lewis, D. A. Rice, and D. Shepherd, "The autonomic and nociceptive response to acute exercise is impaired in people with knee osteoarthritis," *Neurobiol. Pain*, vol. 13, Jan. 2023, Art. no. 100118, doi: 10.1016/J.YNPAI.2023.100118.
[69] K. Thieme, U. Rose, T. Pinkpank, C. Spies, D. C. Turk, and H. Flor, "Psychophysiological responses in patients with fibromyalgia syndrome," *J. Psychosomatic Res.*, vol. 61, no. 5, pp. 671–679, Nov. 2006, doi: 10.1016/J.JPSYCHORES.2006.07.004.
[70] C. Keppler, T. Rosburg, P. Lemoine, M. Pflüger, N. Gyr, and R. Mager, "Functional somatic syndromes: Skin temperatures and activity measurements under ambulatory conditions," *Appl. Psychophysiol. Biofeedback*, vol. 41, no. 4, pp. 363–373, Dec. 2016, doi: 10.1007/S10484-016-9337-X/TABLES/4.
[71] B. Ghavidel-Parsa and A. Bidari, "The crosstalk of the pathophysiologic models in fibromyalgia," *Clin. Rheumatol.*, vol. 42, no. 12, pp. 3177–3187, Dec. 2023, doi: 10.1007/S10067-023-06778-3.
[72] I. P. J. Huijnen et al., "Differences in activity-related behaviour among patients with chronic low back pain," *Eur. J. Pain*, vol. 15, no. 7, pp. 748–755, Aug. 2011, doi: 10.1016/j.ejpain.2010.11.015.
[73] D. G. Patterson et al., "Objective wearable measures correlate with self-reported chronic pain levels in people with spinal cord stimulation systems," *npj Digit. Med.*, vol. 6, no. 1, pp. 1–9, Aug. 2023, doi: 10.1038/s41746-023-00892-x.

Authorized licensed use limited to: KU Leuven University Library. Downloaded on January 27,2026 at 07:26:04 UTC from IEEE Xplore. Restrictions apply.