

Smartwatch-derived digital biomarkers distinguish episodic pain phenotypes in chronic low back pain

Karim Khattab

University of California, San Francisco

Aniket Pratapneni

University of California, San Francisco

Gregorij Kurillo

University of California, San Francisco

Trisha Hue

University of California, San Francisco

Patricia Zheng

University of California, San Francisco

Jeffrey Lotz

University of California, San Francisco

Abel Torres-Espin

University of California, San Francisco

Jeannie F Bailey

`jeannie.bailey@ucsf.edu`

University of California, San Francisco

Short Report

Keywords:

Posted Date: March 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8745290/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This study evaluated smartwatch-derived digital biomarkers for chronic low back pain phenotyping. Temporal features extracted from six months of smartwatch data (resting heart rate, heart rate variability, step counts) in 261 patients discriminated episodic from non-episodic pain phenotypes with 82% accuracy. Frequency-domain and local variability measures outperformed summary statistics, with combined heart and activity features providing superior classification. These findings demonstrate feasibility of passive wearable monitoring for objective pain phenotyping.

INTRODUCTION

Chronic low back pain (cLBP) is the most prevalent and costly cause of disability worldwide¹. However, clinical phenotyping of cLBP patients remains challenging due to its multifactorial etiology and heterogeneous presentation². While cLBP is often defined as “the persistence of pain beyond 3 months of symptoms”³, the temporal dynamics of pain vary substantially between patients^{4–6}. Some experience relatively stable pain levels over time, while others exhibit fluctuating, episodic patterns with distinct flare-ups^{5,7}. Patients often follow a specific pain trajectory over the time course of their low back pain, with little change in the nature of their trajectory⁸. Due to this stability and association with patient-specific measures, pain trajectories may represent more clinically meaningful phenotypes that could guide personalized treatment approaches⁴.

Current methods for tracking pain trajectories rely heavily on patient recall and self-report, introducing subjectivity and recall bias⁹. The emergence of wearable devices such as smartwatches presents an opportunity to develop objective, continuously monitored digital biomarkers for pain trajectory phenotyping.

Digital health technologies are increasingly recognized for their potential to transform chronic pain management^{10,11}. Wearables enable passive collection of physiological and behavioral data in real-world settings, offering insights unattainable through traditional episodic clinical assessments. Measures of resting heart rate (RHR) and heart rate variability (HRV) are associated with pain and have been shown to predict pain intensity^{12–14}. Physical activity has been shown to exhibit complex, non-linear relationships with pain, particularly in cLBP¹⁵. However, these studies have focused on daily summary measures (e.g., daily averages of step count, sedentary time, or resting HRV) that do not capture temporal trends. Importantly, a recent study has found that day-to-day variability in activity is associated with self-reported flare-ups in cLBP¹⁶, suggesting that objective longitudinal measures may be valuable biomarkers of patient pain trajectories.

We hypothesized that objective temporal measures extracted from smartwatch-derived activity and heart rate trajectories are associated with pain trajectories characterized by episodic fluctuations. We tested our hypothesis by evaluating whether we could discriminate between episodic and non-episodic pain phenotypes in cLBP patients using six months of smartwatch data from the BACKHOME UCSF cLBP cohort¹⁷ (Fig. 1), and we show associations and predictive utility of physiological and activity-based dynamic features with episodic pain phenotypes. We further investigate dynamic properties of the wearable data to gain insights on their use as dynamic digital biomarkers in cLBP.

RESULTS

Cohort characteristics

261 cLBP patients (108M, 153F) from the BACKHOME¹⁷ study cohort participated in the smartwatch study. Of those, 195 (74.7%) reported 'episodic' pain trajectories and 66 (25.3%) reported 'non-episodic' pain trajectories, with no significant group differences in age, sex distribution, or BMI. Across all subjects, the average resting heart rate was 64 ± 8.3 bpm and the average heart rate variability was 29.1 ± 10.4 ms. Of the 261 subjects, only 199 had daily step data, with an average daily step count of 5472 ± 2653 steps.

Discrimination of episodic vs. non-episodic pain phenotypes

We evaluated four elastic-net logistic regression models using nested 3-fold cross validation (25 repeats) to predict episodic vs non-episodic pain classification. The RHR model (using only resting heart rate trajectory features) resulted in 18 selected features and achieved a nested AUC of 0.718 ± 0.012 and final AUC of 0.718 ± 0.012 (Nagelkerke $R^2 = 0.045$). The HRV model (using only heart rate variability trajectory features) resulted in 14 selected features and performed more modestly, achieving a nested AUC of 0.668 ± 0.024 and final AUC of 0.671 ± 0.020 (Nagelkerke $R^2 = -0.070$). Combining RHR and HRV features (Heart model) resulted in 30 features and substantially improved performance, with a nested AUC of 0.781 ± 0.019 and final AUC of 0.780 ± 0.019 (Nagelkerke $R^2 = 0.28$). The Heart + Steps model, incorporating features from all three variables, resulted in 31 features and achieved the best discrimination: nested AUC 0.816 ± 0.029 , final AUC 0.820 ± 0.024 (Nagelkerke $R^2 = 0.48$), demonstrating that activity trajectory features provide complementary information to heart rate metrics.

Model feature importance

SHAP value analysis revealed that frequency-domain features (FFT coefficients) had the highest relative importance across all models. In the RHR model, the mean day-to-day change in RHR was also important, with lower day-to-day change associated with non-episodic pain (Fig. 3B). In the HRV model, the weekly autocorrelation and the weekly partial autocorrelation were important, with higher values in both associated with non-episodic pain (Fig. 3D). For both the RHR model and the HRV model, BMI was a relatively important feature with higher BMI values associated with non-episodic pain. In the Heart model, features with high relative importance included both RHR and HRV FFT features (Fig. 3F). The Heart + Steps model had RHR, HRV, and Steps features amongst the features with the highest relative importance (Fig. 3H).

DISCUSSION

We demonstrate that temporal features extracted from consumer smartwatch data can discriminate between episodic and non-episodic pain phenotypes in cLBP patients. Notably, the combination of RHR, HRV, and activity trajectory features improves classification. While overall model performance for the RHR- and HRV-only models was modest, combining these with step count trajectory features improved model performance to an AUC of 0.820. Given the sparse nature of the trajectory data, this suggests that variations in wearable-derived step count and heart rate data may be associated with fluctuations in pain experience. This is corroborated by the SHAP values showing that measures of frequency and daily or weekly trends may have more relative importance than summary measures. Further, this matches previous findings showing associations between changes in activity and heart rate to fluctuations in pain^{16,18}.

Our results align with emerging evidence that wearable-derived digital biomarkers can capture pain-relevant physiological and behavioral changes. Summary measures of physical activity, heart rate, and HRV over long periods of time have been found to correlate with chronic pain intensity¹². We extend these findings by demonstrating that temporal trajectory features – particularly those capturing short-term dynamics – offer additional discriminative power for phenotyping. The prominence of frequency-domain and ‘local’ variability features (e.g., day-to-day changes, weekly autocorrelations) over global summary measures suggests that fluctuations at shorter timescales may better reflect the episodic nature of pain experience. This corroborates findings that day-to-day variability in step count associates with flare-up incidence in cLBP patients¹⁶.

The superior performance of the combined Heart + Steps model (AUC = 0.820) suggests that heart rate and activity trajectory features appear to explain different variances in pain classification. Heart rate metrics may reflect autonomic responses and stress-related processes^{13,14}, while activity patterns may reflect functional capacity and pain-related

movement avoidance¹⁹. While studies have shown that heart rate and activity may be changed in chronic pain patients¹⁹⁻²¹, these signals are also sensitive to other stressors. Studies have shown that self-reported flare-ups in cLBP do not always coincide with increased pain intensity^{22,23}, suggesting that other factors may influence this reporting. The multimodal approach may therefore provide robustness against such confounding factors, offering a more comprehensive picture of a patient's chronic pain trajectory.

These findings have implications for digital health approaches to chronic pain management. Real-time monitoring of objective, passively collected trajectory features could enable earlier detection of pain flare-ups, facilitating timely intervention. However, this study is not without limitations. First, self-reported pain trajectories using the VTQ-Pain, while validated²⁴, remain susceptible to recall bias and do not allow for identification of specific flare-up incidences, instead relying on the assumption that patients reporting an episodic trajectory type are more likely to experience a higher frequency of flare-ups. Second, while subjects with large portions of missing data were excluded, missing data persisted in the final dataset, as is inherent to passive wearable data collection, and this may have attenuated our ability to identify associations between features. In addition, decreases in model performance may be related to the subjectivity and potential for bias in the VTQ-Pain rather than a lack of prediction power in the activity features. This further emphasizes the importance of objective measures for patient pain tracking and classification. Future work will incorporate daily pain surveying, which may better reflect patient pain experience at than the VTQ-Pain at the cost of exhaustive daily surveying.

In conclusion, we demonstrate that temporal features extracted from RHR, HRV, and step count trajectories are associated with episodic vs non-episodic pain classification. This suggests that heart rate and activity tracking, especially in tandem, may provide an opportunity for objective monitoring of patient pain, fluctuations, and response to interventions. Further, this study shows the importance of 'local' measures of variability in heart rate and activity trajectories, rather than summary measures alone, as a possible objective surrogate for subjective self-reported measures of pain experience. Lastly, we find that the combination of features from resting heart rate, HRV, and step count trajectories resulted in the best association with pain classification, suggesting that, in combination, these trajectory features provide more information about the patient pain experience. Future study will seek to utilize more comprehensive wearable data like sleep quality, walking speed and stride length, and other metrics to further explore associations between health trajectories and contemporary pain. This work establishes the feasibility of digital biomarkers for pain trajectory classification and suggests a path toward the development of reliable, continuous monitoring of patient pain experience using objective measures rather than subjective and discrete self-reported metrics.

METHODS

Ethics approval and informed consent

This study was approved by the University of California, San Francisco Institutional Review Board (institutional IRB #21-34420) and WCG IRB (IRB #20212455) as part of the multicenter BACPAC BACKHOME study. All participants provided written informed consent prior to enrollment. All study procedures were conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. Participants consented to the retrospective collection and analysis of their Apple Watch health data from iPhone Healthkit.

Study population

The study population was a subset of the BACKHOME¹⁷ study cohort (, in which the key eligibility requirement was current chronic low back pain defined as pain between the lower posterior margin of the rib cage and the horizontal gluteal fold, which has persisted for at least the past 3 months and has resulted in pain on at least 50% of days in the past 6 months (per the NIH Task Force on Research Standards for cLBP^{25,26}), and that their low back pain was worse than any other bodily pain they were experiencing. BACKHOME participants who allowed access to their Apple Healthkit data and had enough data for analysis (data available for at least 120 days of the 180-day period) were included in this study.

Data collection

With IRB approval and informed consent, cLBP patients in the BACPAC BACKHOME cohort completed the Visual Trajectories Questionnaire-Pain (VTQ-Pain)²⁴, a single question prompting patients to select the visual pain trajectory which best describes the course of their pain over the past 6 months (the only exclusion criterion was the inability to read or write English¹⁷). Patients were binned into ‘episodic’ or ‘non-episodic’ pain groups based on their response to the VTQ-Pain (Fig. 2). Then, for 261 participants who consented to iPhone Healthkit data access, we retrospectively collected resting heart rate, heart rate variability, and step count data, from iPhone HealthKit, for the six-month period described in the VTQ-Pain survey. We cleaned and aggregated the raw HealthKit data into daily average resting heart rate, daily median heart rate variability, and daily average step counts. From these daily values, we created six-month trajectories of RHR, HRV, and step count for each patient. Patients were excluded if they had more than 60 days of missing data in the 180-day time frame, resulting in a final cohort of 258 patients.

Feature engineering

We applied the tsfresh package²⁷ to extract 600 time-series features per modality. Features included: summary statistics (mean, standard deviation, min/max, percentiles), distribution properties (skewness, kurtosis), linear trends, autocorrelation functions, spectral properties (FFT coefficients, entropy), complexity measures, and change quantifications. Features with missing values or zero variance across participants were removed. To address multicollinearity, we identified highly correlated feature pairs ($> 90\%$) and randomly removed one feature from each pair. This yielded a total of 300 features for model development.

We employed stability selection²⁸ to identify robust features for each model. Using 100 bootstrap resamples (50% sample size, stratified by outcome to maintain class imbalance), we ranked features by univariate ANOVA F-statistic and selected the statistically significant features ($p < 0.05$) in each resample. Features selected in $\geq 70\%$ of resamples were retained for final modeling. This yielded 18 RHR features, 14 HRV features, 30 Heart (RHR + HRV) features, and 31 Heart + Steps features.

Statistical modeling

We trained elastic-net logistic regression models (scikit-learn) to predict episodic vs non-episodic classification using four feature sets: (1) RHR only, (2) HRV only, (3) Heart (RHR + HRV), and (4) Heart + Steps.

Models were trained on a 70 – 30 training-testing split with a nested cross-validation. Both inner and outer cross-validations were 3-fold with 25 repeats. Hyperparameters were tuned in the inner fold using grid search with the final model hyperparameters taken from the best model across folds and repeats. For each model, we calculated the mean area under the receiver-operating curve (AUC) of the nested cross-validation (nested AUC), as well as the AUC of the final model refit following nested cross-validation. The final model coefficients and AUC were calculated from a single pass through the inner cross-validation using the full dataset. We report mean AUC across all outer folds (nested AUC) and final model AUC from refitting the full dataset with optimal hyperparameters.

Model performance was further assessed using Nagelkerke pseudo- R^2 , a scaled version of Cox-Snell R^2 normalized to a maximum of 1, calculated from the final refitted model.

Model interpretation

We computed SHAP (SHapley Additive exPlanations) values²⁹ using the shap package to quantify feature importance and direction of effect. SHAP values represent each feature's contribution to individual predictions. We report mean absolute SHAP values (global

feature importance) and SHAP value distributions stratified by feature values to visualize feature-outcome relationships.

Statistical analysis

Group comparisons used independent t-tests for continuous variables and chi-square tests for categorical variables. All analyses were conducted in Python using numpy, pandas, scipy, scikit-learn, shap, and tsfresh libraries.

Declarations

CONFLICT OF INTEREST

I declare that the authors have no competing interests as defined by Nature Portfolio, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

FUNDING

This research did not receive funding.

Author Contribution

KK performed the analysis for this study. KK and AP wrote the main manuscript text and prepared figures. KK, GK, TH, PZ, JL, AT-E, and JFB were involved in study design and data acquisition. All authors reviewed the manuscript.

Data Availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at the University of California, San Francisco.

References

1. The Lancet Rheumatology. The global epidemic of low back pain. *Lancet Rheumatol.* 5, e305 (2023).
2. Koes, B. W., Van Tulder, M. W. & Thomas, S. Diagnosis and treatment of low back pain. *BMJ* 332, 1430–1434 (2006).

3. Rozenberg, S. [Chronic low back pain: definition and treatment]. *Rev. Prat.* 58, 265–272 (2008).
4. Kongsted, A., Kent, P., Axen, I., Downie, A. S. & Dunn, K. M. What have we learned from ten years of trajectory research in low back pain? *BMC Musculoskelet. Disord.* 17, 220, s12891-016-1071-2 (2016).
5. Tamcan, O. *et al.* The course of chronic and recurrent low back pain in the general population. *Pain* 150, 451–457 (2010).
6. Macedo, L. G. *et al.* Nature and Determinants of the Course of Chronic Low Back Pain Over a 12-Month Period: A Cluster Analysis. *Phys. Ther.* 94, 210–221 (2014).
7. Axén, I. & Leboeuf-Yde, C. Trajectories of low back pain. *Best Pract. Res. Clin. Rheumatol.* 27, 601–612 (2013).
8. Dunn, K. M., Campbell, P. & Jordan, K. P. Long-term trajectories of back pain: cohort study with 7-year follow-up. *BMJ Open* 3, e003838 (2013).
9. Stanton, T. R., Latimer, J., Maher, C. G. & Hancock, M. J. How do we define the condition ‘recurrent low back pain’? A systematic review. *Eur. Spine J.* 19, 533–539 (2010).
10. Patel, M. S., Asch, D. A. & Volpp, K. G. Wearable Devices as Facilitators, Not Drivers, of Health Behavior Change. *JAMA* 313, 459 (2015).
11. Dunn, J., Runge, R. & Snyder, M. Wearables and the Medical Revolution. *Pers. Med.* 15, 429–448 (2018).
12. Patterson, D. G. *et al.* Objective wearable measures correlate with self-reported chronic pain levels in people with spinal cord stimulation systems. *Npj Digit. Med.* 6, 146 (2023).
13. Forte, G., Troisi, G., Pazzaglia, M., Pascalis, V. D. & Casagrande, M. Heart Rate Variability and Pain: A Systematic Review. *Brain Sci.* 12, 153 (2022).
14. Rampazo, É. P., Rehder-Santos, P., Catai, A. M. & Liebano, R. E. Heart rate variability in adults with chronic musculoskeletal pain: A systematic review. *Pain Pract.* 24, 211–230 (2024).
15. Heneweer, H., Vanhees, L. & Picavet, S. J. H. Physical activity and low back pain: A U-shaped relation? *Pain* 143, 21–25 (2009).
16. Costa, N. *et al.* Are objective measures of sleep and sedentary behaviours related to low back pain flares? *Pain* 163, 1829–1837 (2022).
17. Hue, T. F. *et al.* Design of the COMEBACK and BACKHOME Studies, Longitudinal Cohorts for Comprehensive Deep Phenotyping of Adults with Chronic Low-Back Pain (cLBP): a part of the BACPAC Research Program. Preprint at <https://doi.org/10.1101/2024.04.09.24305574> (2024).

18. Dudarev, V., Barral, O., Radaeva, M., Davis, G. & Enns, J. T. Night time heart rate predicts next-day pain in fibromyalgia and primary back pain. *PAIN Rep.* 9, e1119 (2024).
19. Ogawa, T., Castelo-Branco, L., Hatta, K. & Usui, C. Association Between Step Count Measured With a Smartphone App (Pain-Note) and Pain Level in Patients With Chronic Pain: Observational Study. *JMIR Form. Res.* 6, e23657 (2022).
20. Koenig, J., Loerbroks, A., Jarczok, M. N., Fischer, J. E. & Thayer, J. F. Chronic Pain and Heart Rate Variability in a Cross-Sectional Occupational Sample: Evidence for Impaired Vagal Control. *Clin. J. Pain* 32, 218–225 (2016).
21. Danilin, L. K., Spindler, M., Sörös, P. & Bantel, C. Heart rate and heart rate variability in patients with chronic inflammatory joint disease: the role of pain duration and the insular cortex. *BMC Musculoskelet. Disord.* 23, 75 (2022).
22. Costa, N. *et al.* ISSLS PRIZE IN CLINICAL SCIENCE 2021: What are the risk factors for low back pain flares and does this depend on how flare is defined? *Eur. Spine J.* 30, 1089–1097 (2021).
23. Costa, N. *et al.* Low Back Pain Flares: How do They Differ From an Increase in Pain? *Clin. J. Pain* 37, 313–320 (2021).
24. Dunn, K. M., Campbell, P. & Jordan, K. P. Validity of the Visual Trajectories Questionnaire for Pain. *J. Pain* 18, 1451–1458 (2017).
25. Deyo, R. A. *et al.* Focus article: report of the NIH task force on research standards for chronic low back pain. *Eur. Spine J.* 23, 2028–2045 (2014).
26. Mauck, M. C. *et al.* The Back Pain Consortium (BACPAC) Research Program: Structure, Research Priorities, and Methods. *Pain Med.* 24, S3–S12 (2023).
27. Christ, M., Braun, N., Neuffer, J. & Kempa-Liehr, A. W. Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing* 307, 72–77 (2018).
28. Meinshausen, N. & Bühlmann, P. Stability Selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72, 417–473 (2010).
29. Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. Preprint at <https://doi.org/10.48550/ARXIV.1705.07874> (2017).

Figures

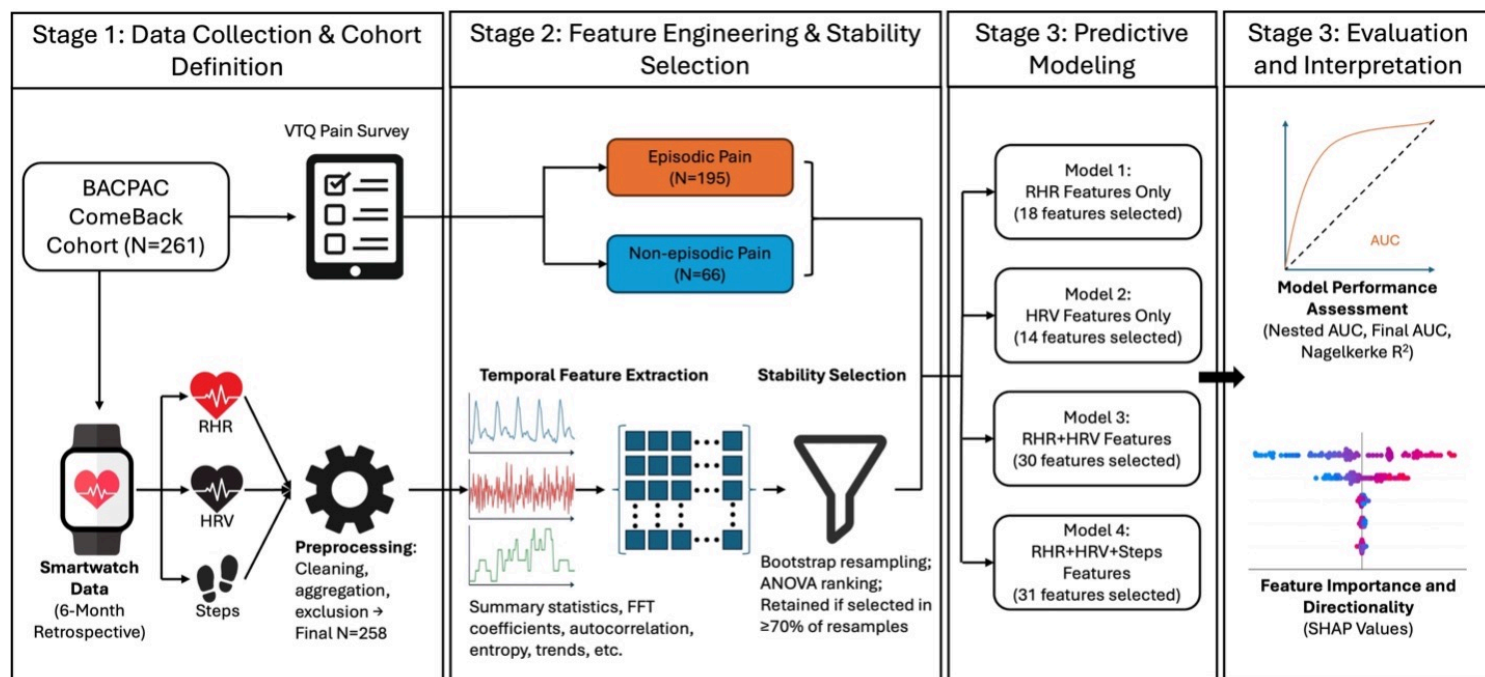


Figure 1

Study design and analytical pipeline.

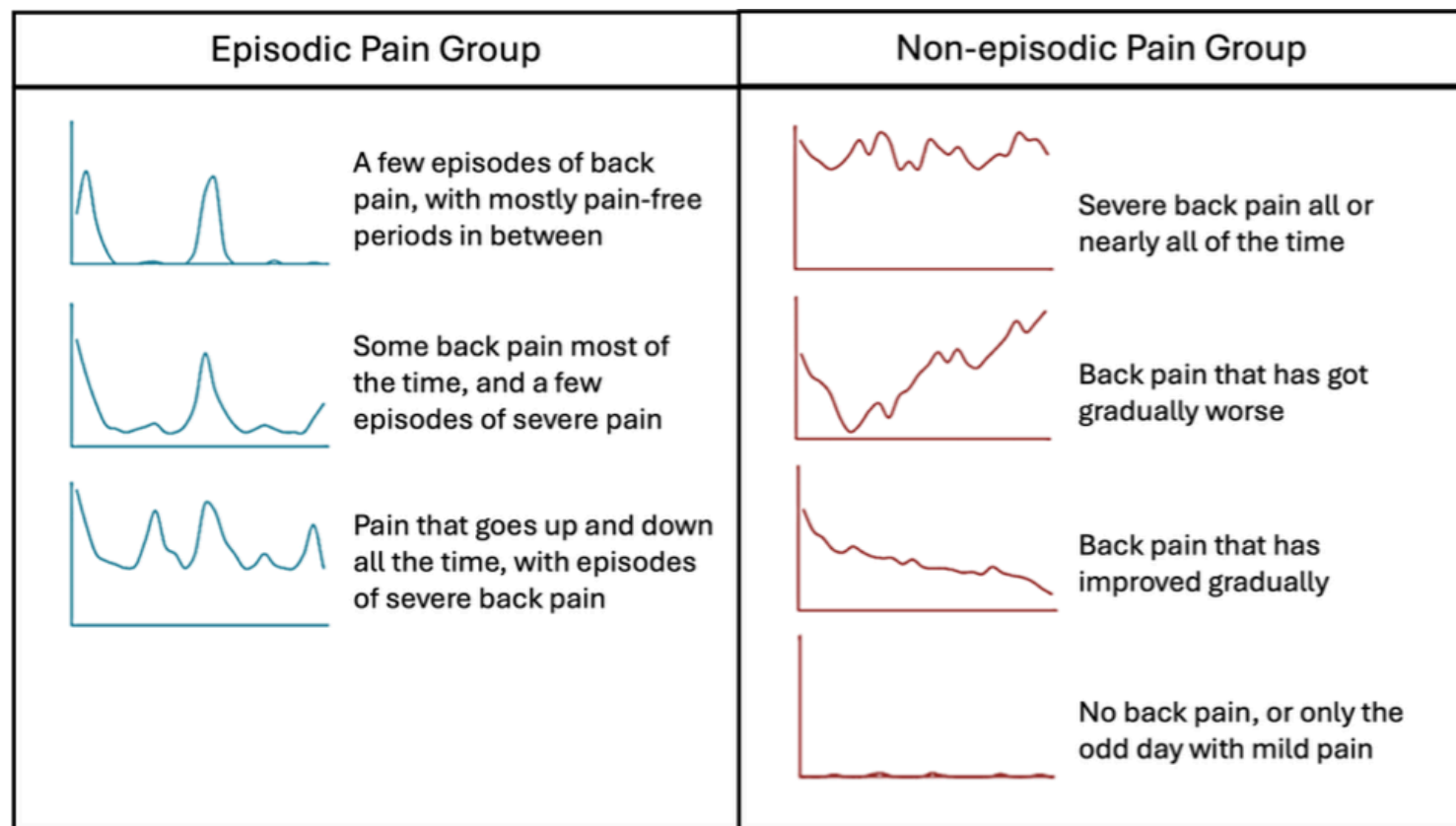


Figure 2

Episodic and non-episodic pain group classification from the Visual Trajectory Questionnaire for Pain. Patients were binned into two groups, episodic or non-episodic, based on their response to the VTQ-Pain questionnaire asking them to select the visual trajectory representing their pain in the past six months.

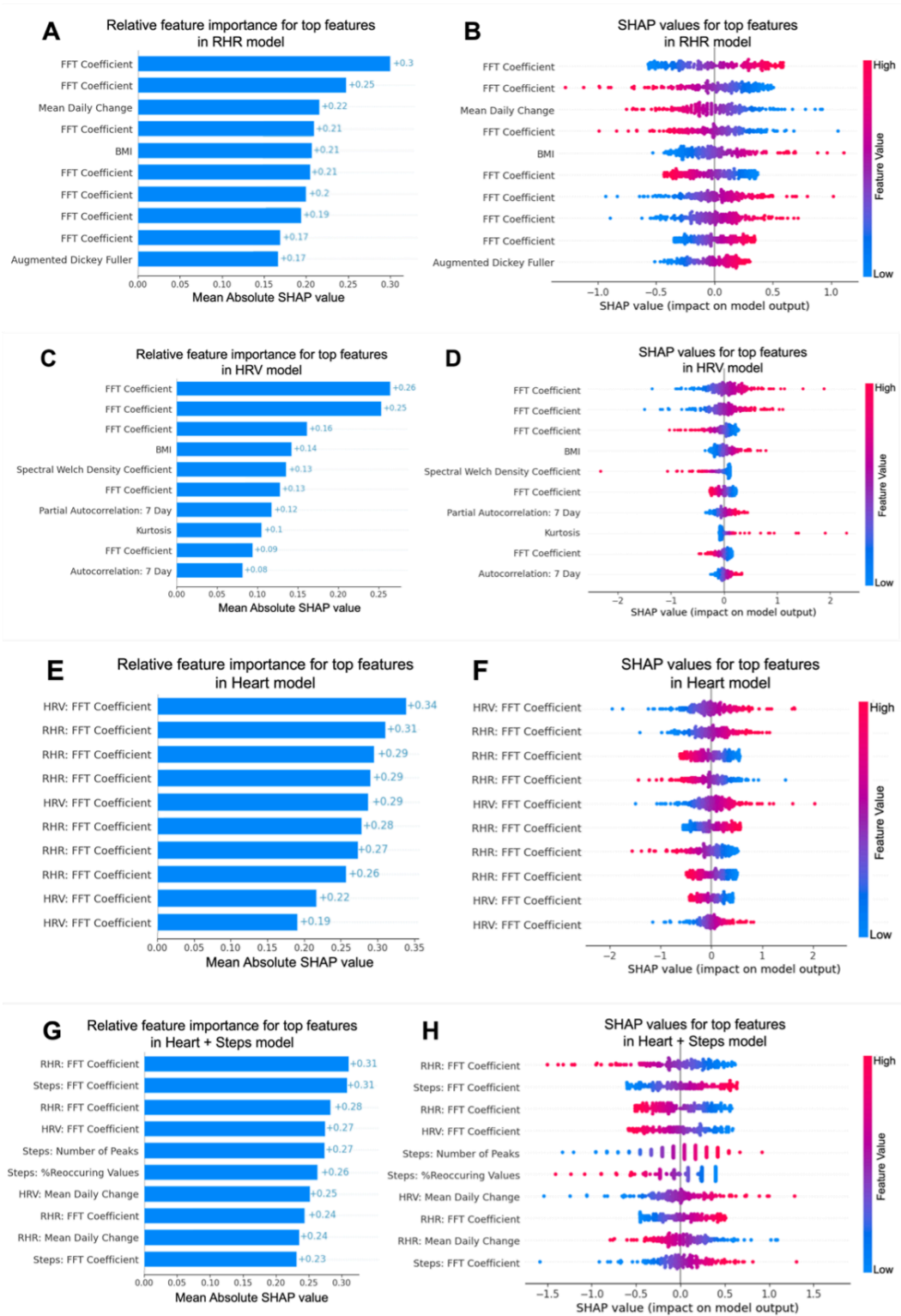


Figure 3

Relative feature importances and directions for the top ten features in the RHR (A, B), HRV (C, D), Heart (E, F), and Heart+Steps (G, H) models. A, C, E, G: Mean absolute SHAP values for the top 10 features in the RHR, HRV, Heart, and Heart+Steps models, respectively. B, D, F, H: SHAP values of the top ten features for each patient, color coded by feature value. Positive SHAP values represent relative importance for non-episodic pain group classification and negative SHAP values represent relative importance for episodic pain group classification.