Title

Personalized entropy-informed deep learning for identifying opioid misuse

Journal

Nature Mental Health, 4(1)

ISSN

2731-6076

Authors

Luo, Yunfei

Deznabi, Iman

Gullapalli, Bhanu Teja

et al.

Publication Date

2026

DOI

10.1038/s44220-025-00555-8

Personalized entropy-informed deep learning for identifying opioid misuse

Yunfei Luo[1] , Iman Deznabi[2] , Bhanu Teja Gullapalli[1] ,
Mark Tuomenoksa[3] , Madalina Brostean Fiterau[2] ,
Eric L. Garland[4,5*], Tauhidur Rahman[1*]

[1]Halıcıoğlu Data Science Institute, University of California San Diego
[2]Manning College of Information and Computer Sciences,
University of Massachusetts Amherst
[3]Ilumivu, Inc.
[4]Sanford Institute for Empathy and Compassion,
University of California San Diego
[5]Department of Psychiatry, School of Medicine,
University of California San Diego

Accepted manuscript version.
This is the accepted manuscript of the article:
Luo, Yunfei, Iman Deznabi, Bhanu Teja Gullapalli, Mark Tuomenoksa, Madalina Brostean Fiterau, Eric L. Garland, and Tauhidur Rahman. "Personalized entropy-informed deep learning for identifying opioid misuse." Nature Mental Health (2026): 1-13..
DOI: 10.1038/s44220-025-00555-8
The final authenticated version is available online at:
https://www.nature.com/articles/s44220-025-00555-8

Abstract

Fluctuations in pain, stress, and craving are thought to opioid misuse. Developing accurate prediction models is vital for intervention and prevention efforts. In this work, we leverage physiological data and semantic analysis of electronic health records to tackle the challenge of detecting opioid misuse. Utilizing personalized hierarchical deep learning models, we analyze trajectories of predicted pain, stress, and craving states with 10,140 hours of heart rate data collected by wearables from patients on long-term opioid therapy. From these trajectories, we extract nonlinear dynamical features and develop a novel relevance-based temporal fusion model of opioid misuse risk. We incorporate clinical data into a natural language processing (NLP) model to enhance opioid misuse risk detection. We then fuse these results to achieve a robustly accurate opioid misuse risk assessment (PRC AUC = 0.94 ± 0.05). This approach, combining physiological data, machine learning, and semantic inference, marks a significant advancement in personalized prediction of addictive behavior by elucidating the entropic nature of underlying affective state dynamics.

Keywords: Digital Healthcare, Affective State, Opioid Misuse, Machine Learning, Artificial Intelligence

1 Introduction

The misuse of opioids continues to pose a significant public health challenge, with far-reaching consequences for individuals and communities. The United States particularly has observed a concerning escalation in national overdose deaths caused by opioid misuse and disorder, rising from 21,089 in 2010 to 47,600 in 2017. This alarming trend intensified with the onset of the COVID-19 pandemic [1], with reported deaths reaching 68,630 in 2020 and 80,411 in 2021 [2]. Regrettably, reported opioid misuse deaths continued to escalate to 112,000 in 2023 [3]. Chronic, high dose opioid use is thought to induce neuroplastic changes in corticolimbic and corticostriatal circuits that govern stress and reward-related

*These authors jointly supervised this work.

processes, resulting in heightened stress, increased pain sensitivity, and intense craving states that dynamically interact and contribute to opioid misuse risk [4]. Despite growing awareness of these affective drivers of opioid misuse, clinicians lack tools to monitor them in real time, leaving critical windows for intervention unaddressed. Traditional assessment approaches, including self-report and periodic clinical visits, are too infrequent to capture the rapid, nonlinear fluctuations in stress, pain, and craving that often precede misuse. In this study, we present a novel, wearable-based system that passively detects such high-risk affective states using heart rate variability (HRV) and machine learning. By modeling the chaotic and dynamic nature of these states, our approach offers a data-driven and clinically relevant solution to predicting opioid misuse risk-enabling early detection, personalized care, and just-in-time intervention in everyday settings.

Affectively-valenced states integral to opioid misuse, including stress [5], pain [6], and craving [7], manifest and kindle one another over time in temporally dynamic and complex networks that are stimulated and perturbed by ongoing life events (e.g., the onset of a new stressor or pain episode) and constrained by underlying stable dispositional tendencies [8]. Moreover, high frequency changes in these states from moment-to-moment may be superimposed over low frequency changes in mood, addictive tendencies, and chronic pain over the course of hours and days. As such, traditional linear models may fail to capture the non-linear dynamics and entropy inherent in such complex systems, requiring more complex computational modeling approaches [9]. Experience sampling methods and ecological momentary assessments (EMA; typically recorded a few times a day) may not be of sufficient density to reflect momentary fluctuations in stress, pain, and craving that increase the risk of opioid misuse behavior. Because affective processes unfold on the order of several hundred milliseconds, sparse self-reports may fail to capture critical windows where affect shifts in response to or remains stable in spite of changing environmental contingencies [10]. Yet, the relative inertia or lability of momentary affective states may contribute to addictive behaviors [11]. For instance, patients may report feeling "stuck" in negative affective states that are hard to break, and consequently escalate opioid use as a means of obtaining relief from the unremitting suffering.

Such complex system dynamics can be modeled via chaos theory, which provides a more intuitive and comprehensive understanding of the emergent behaviors seen in conditions like chronic pain, stress, and craving among people who use opioids [12, 13]. In chaos theory, nonlinear dynamics and chaotic features account for the system's sensitivity to initial conditions and the presence of multiple stable states, or attractors, that the system gravitates towards. For instance, chronic pain can be seen as a maladaptive attractor state in the brain's neural networks, resulting from non-linear feedback loops and plasticity [14]. Similarly, the regulation of stress and craving through allostatic processes demonstrates how the system can achieve stability through adaptive changes to hedonic setpoints in the brain, reflecting the principles of nonlinear dynamics [15]. Dynamic responses in the face of homeostatic perturbations (e.g., the onset of a stressor, a pain flare, or an episode of craving) activate a central-autonomic network that governs appetitive behavior (e.g., opioid misuse) via downstream effects on neurovisceral pathways, including vagal modulation of heart rate known as HRV [16, 17]. For example, phasic decreases in high frequency HRV reliably index stress [18]. The advent of affordable, commercially-available wearables (e.g., smartwatches) has made continuous ambulatory monitoring of HRV a feasible means of assessing affective dynamics. In that regard, ambulatory measures of HRV have been shown to be associated with momentary ratings of stress, pain, and craving [19–24]. By leveraging the nonlinear and chaotic features of the stress, pain and craving scores automatically estimated by HRV timeseries data, modern machine learning approaches can better capture the unpredictable and multifaceted nature of affective states associated with opioid misuse risk.

Developing a passive, continuous and unobtrusive opioid misuse risk monitoring system [25–27] with mobile and wearable technology can enable early misuse screening and just-in-time interventions. Here, we test an opioid misuse risk model that uses passively and continuously captured HRV data from a commercially-available smartwatch to estimate states predictive of opioid misuse. While mobile and wearable sensor data have demonstrated efficacy in monitoring opioid administration moments [26–28], tracking affective states [29] associated with opioid misuse presents a significant challenge due to the highly individual-specific nature of perceived stress, pain, and craving levels. To meet this challenge, we leveraged chaotic system theory to analyze and model opioid misuse behavior based on dynamical affective states indexed by HRV signals obtained by a wearable. We developed a machine learning system that detects stress, pain, and craving patterns from the HRV data. We captured the nonlinear and chaotic dynamics of these affective states, and then used these features to estimate opioid misuse risk. To enhance clinical relevance, we then integrated these physiological dynamics with a large language model (LLM) to semantically represent patient demographic and health data. This integrated approach

was then used to predict opioid misuse risk with a high degree of accuracy.

Through this work, we present a passive, continuous, and unobtrusive opioid misuse risk monitoring system using commercially available wearable devices [25–27], providing a practical and scalable advance over traditional screening methods. Unlike scheduled assessments, our method enables early identification of misuse risk and just-in-time interventions while minimizing patient burden. Continuous passive monitoring of physiological markers reflecting affective states allows risk detection between clinical visits, supports automatic integration with electronic health records, and delivers personalized alerts to healthcare providers. By shifting monitoring from patients and clinicians to a low-cost, automated system, this work advances the field of intelligent opioid use monitoring, bringing continuous, patient-friendly, and clinically actionable surveillance closer to real-world deployment.

2 Results

To evaluate our approach, we adopt a two-stage process designed to capture personalized affective dynamics and assess opioid misuse risk. In Stage 1 (Personalized Affective State Prediction), we perform within-subject 5-fold cross-validation using self-reported stress, pain, and craving levels as ground truth. This personalized modeling approach is motivated by the idiographic nature of affective processes; notably, the affect prediction model developed in this stage is never exposed to any misuse risk labels. In Stage 2 (Misuse Risk Assessment), we use the affective representations learned in Stage 1 to infer trajectories across the entire dataset, which includes a substantial portion of data not used in Stage 1 due to the absence of self-reported affective scores. We then perform 5-fold cross-validation across participants, ensuring subject-level independence between folds. All reported metrics reflect performance on unseen validation data, with reasonable partitioning to prevent information leakage and maintain sufficient data coverage.

2.1 Affective State Prediction

2.1.1 Evaluation Schema

To evaluate our methods compared to the baseline and for the ablation study, we used 5-fold cross-validation. The split is stratified by subjects, and samples from each subject are split into 5 folds. The baselines include models without personalization, as well as models with vanilla fusion of personalized characteristics. Because we examined the incremental validity of taking a personalized prediction approach, some amount of data needs to be included in the training set for each subject. 8 out of 51 subjects are removed due to an insufficient number of samples with either lack of inter-beat interval (IBI) or ground truth. The main modeling task is regression. Thus, at the same time we report the testing mean absolute error (MAE), we also report the Pearson correlation coefficient between the predicted values from our model and the ground truth in the test sets. Using this evaluation framework, we can not only assess the disparity between the predicted values and the ground truth, but also gauge the extent of their linear correlation.

2.1.2 Personalization and Dynamic User Segmentation

Here we present the results of our ablation study on demonstrating the added predictive power of using personalized and dynamic grouping modules. We observe that direct concatenation of personalized demographic and clinical characteristics as suggested by the literature provides limited gain in performance. Adding our proposed personalization module equipped with a branching algorithm following guidance from [30] significantly improves performance above baseline model, the predicted stress, pain, and craving value achieves the Pearson correlation coefficient of respectively 0.70, 0.72, and 0.81. Similarly, the mean absolute error are 1.44, 1.21, and 1.13 respectively. Detailed performance comparison and ablation studies are presented in Appendix G in Table 17.

The shared temporal and spatial convolutional layers in the backbone, branching layers, and personalized layers of our model act as global, group-level, and individual-level representation learners respectively to produce optimal predictive performance. Figure 2 presents a more intuitive visualization of the model prediction, indicating a significant gain in performance from our personalized approach with branching for dynamic user segmentation and novel temporal fusion. The box plots in figure 2 illustrates that while direct fusion of demographic information to predict affective states with wearable data produces a flat response to increasing ratings, our learning to branch approach shows a diagonal trend.

2.2 Opioid Misuse Detection

The physiological signal (i.e., interbeat interval [IBI]) is systematically processed in multiple stages before it is converted to the opioid misuse risk score. First, as described above, HRV features are extracted from the raw IBI data which are then used to train a personalized stress, pain, and craving score prediction network. By dynamically identifying optimal user group segmentation, the model generates personalized predictions for levels of stress, pain, and craving longitudinally based on HRV data. The personalized affect score prediction network generates a set of stress, craving and pain trajectories from the wearable data from each of the participant's data. The trajectories are divided by missing wearable data due to moments of non-use, primarily due to sleep and battery charge. On average, each participant contributed 8.28 (SD=6.72) days worth of wearable data which resulted in 42 (SD=29) continuous trajectories on average where each trajectory is 4.75 (SD=4.02) hours long.

2.2.1 Evaluation Schema

For conducting the task of opioid risk assessment, we construct the modeling problem as a binary classification task. Considering that we have time series of predicted affective pain, stress, and craving states and text-based clinical data as 2 main modalities in the available data, we carefully process each modality separately, aiming to feed informative features to train the classifier.

For evaluating the methods, we conduct 5 fold cross validation where the users are divided into 5 groups. The train and test process is conducted such that each time one of the 5 groups is left as a test set, and we can evaluate the performance of our proposed model on data from unseen participants.

2.2.2 Experimental Performance

For conducting the task of opioid risk assessment, we construct the modeling problem as a binary classification task. For the time series of affective states, we leverage nonlinear dynamics features to represent each series. The nonlinear dynamics features are a fixed set of metric numbers that are used to describe the chaotic and predictability of the trajectories of affective states inferred from raw physiological signals. For the text based clinical data including opioid prescription, symptoms, psychiatric disorders, and basic demographics, we compute semantic representation from the raw text by extracting embedding vectors from pretrained transformer based language models. These semantic representations will carry the information about the users' health status, and are able to differentiate between users with different prescribed types of opioids and different demographics. We then train a lightweight neural network model with the nonlinear dynamics features (e.g., Persistence Entropy) from the user's states estimated from the wearable signals. Furthermore, we fuse the language models' (ClinicalBERT and ClinicalTinyLlama) embeddings. With the novel relevance based fusion mechanism (NNRF), we are able to assess the extent of relevance of predicted state trajectories from various days with various recorded lengths. With a linear probing layer, we are able to directly fuse wearable based nonlinear dynamics features with health records based semantic features and output final opioid misuse risk score.

From Table 1, we can see that LLM ClinicalTinyLlama with all EHR features and inferential results from NNRF achieve the highest ROC and PRC AUC scores of 0.863 and 0.940 respectively. Opioid misuse detection performance improves when the nonlinear dynamics feature-based model is fused with the semantic representation model for clinical data.

Figure 3 presents the T-SNE plot of the test time embedding representation from the best version of the model: ClinicalTinyLlama fusing inferential score from nonlinear dynamics model. These data indicate that the latent space comprising information from nonlinear dynamics features (NLD) and semantic embedding clearly differentiates opioid misusers from individuals who take opioids as prescribed by their physician.

2.2.3 Reduced Entropy in Opioid Misuse

In Figure 4, we visualize the stress, pain and craving trajectories in a 3-dimensional space where the 3 coordinates indicate respectively the state at time t (i.e., current time), t-1 and t-2 which is also widely known as Takens embedding [31]. Takens embedding is one of the widely used time series representation techniques to transfer the time series into points cloud. Since the personalized state prediction model estimates stress, pain, and craving ratings every 30 minutes (with a window shift of 30 minutes), the Takens embedding in this plot captures the current and past one hour's states. Figure 4 illustrates that the state trajectories of prescription users are significantly more irregular, unpredictable and chaotic in

nature compared to that of the miusers' states. For misusers, the three states (of the 3 consecutive hours') are more correlated, with less variability, and follow a smoothed, spiral pattern.

In order to understand the statistical significance of our observation, we employed persistent homology analysis of trajectories of stress, pain and craving separately. The term 'homology' here refers to structures that enclose an area. For instance, homology of 0 dimension refers to connected components which represent the presence of isolated clusters in the data, and homology of 1 dimension refers to loops or tunnels within the data. We first extract these homology on the point cloud generated by Takens embedding from the predicted trajectories of stress, pain, and craving, then we calculate the entropy among 0 and 1 homology dimensions. Entropy here is a concept commonly used in nonlinear dynamical data analysis to measure the predictability or randomness of a system which quantifies the total information gained when the state of a system changes. The more the information gained, the more divergent the states are, hence implying less predictability.

In order to find the statistical relationship between the nonlinear dynamics patterns in pain, stress, and craving states of prescription and misusers, we conduct permutation tests on the distributions of persistence entropy between opioid misusers and those who take opioids as prescribed. We conduct this permutation test on each of the states with entropy calculated on homology of zero and one dimensions separately. Table 3 indicates that prescription users tend to have significantly more unpredictable stress, pain, and craving trajectories than those of misusers.

2.3 Ablation Study

We conducted an ablation study on different nonlinear dynamics features including persistence entropy [32], Lyapunov exponent [33], and DFA [34] which is similar to Hurst exponent [35]. We also experimented with widely used statistical features in time and frequency domains, as well as some cross-domain combinations between these features. The result is shown in Table 9. Persistence entropy and Lyapunov exponent are the two most representative features that describe the degree of chaos and predictability of state trajectories across the day. A key takeaway from this ablation study is that nonlinear dynamical features which comprise information related to the chaotic state of the system yield the best performance for predictive opioid misuse. We also test the performance of the model when given trajectories from various days. With constraints set on 2 days, 4 days, 1 week, and all of available data, we observe AUC PRC of 0.738, 0.760, 0.821, and 0.875 respectively. Thus, the more available data there are, the better the performance.

2.4 Generalizability and Cross-Cohort Evaluation

To further justify the generalizability of our proposed methodology and to emphasize the nonlinear dynamical features from the inferential trajectories of affective states across the day, we obtained new data from 21 additional opioid users with chronic pain, including 13 opioid misusers and 7 prescription users. The dataset comprises 2807 hours of raw IBI data (on average 5.29 hours per continuous raw signal series, 133.67 total discontinued hours per subject). We then re-ran the analysis and ablation studies on this second external dataset. From Table 16, we observe that integrating such nonlinear dynamical entropy features boosts model performance significantly by 7.14% with respect to AUC ROC comparing to approach that only leverage HRV features, from 0.700 ± 0.292 to 0.750 ± 0.316. Notably, this dataset does not have EHR data due to privacy consent, which on the other hand, further strengthens the feasibility of our methodology on inferring opioid misuse behavior from wearable (i.e., smart watch) physiological data as a primary data source.

Finally, because this data set has a similar nature to the main dataset we analyzed in the original paper, both in terms of raw data format and time scale,, we combined all the nonlinear dynamical features together then re-ran the statistical analysis. This integrated feature dataset comprises information from all 72 subjects, 12947 hours of raw IBI data, and 2666 inferential trajectories of affective states across the days when subjects wear the smart watch. From the table below, we observe that prescription users consistently have significantly larger persistence entropy than the misusers at 1 dimension homology across all affective states ($P < .001$).

The detailed description and results of additional external datasets are presented in the supplementary section F. The experimental results from these two independent datasets are consistent with our original observations, highlighting that: (a) people who misuse opioids have significantly low persistence entropy compared to individuals who take opioids as medically prescribed and (b) such nonlinear dynamical features always bring benefit to performance on opioid misuse risk prediction.

3 Discussion

Affective states of stress, pain, and craving are known to increase risk of opioid misuse among people with chronic pain. Here we proposed a integrated personalized temporal fusion and semantic model that combines physiological, demographic, and clinical data to estimate nonlinear trajectories among these states and provide a means of accurately detecting opioid misuse. We observed that patients who misuse opioids show markedly less entropy in estimated stress, pain, and craving states over time relative to patients who take opioids as prescribed. Moreover, fusing state estimation from cardiac-autonomic signals with semantic representation of clinical data resulted in a model with predictive performance that surpasses the accuracy of "gold standard" clinical assessments of opioid misuse.

The attenuated entropy in stress, pain, and craving trajectories among patients who misuse opioids is consistent with prior research demonstrating reduced heart rate variability in opioid misuse and opioid use disorder [36–39], and might be explicated allostatic models of addictive behavior [4, 40–42]. These models propose that chronic use of substances like opioids causes allostatic shifts in brain stress and reward thresholds, resulting in increased sensitivity to stress, pain, and craving, which drives substance misuse as a means of self-medicating the resulting dysphoria. Yet, this regulatory attempt comes at cost by further increasing sensitization to stress, pain, and craving, trapping the individual in a downward spiral of behavioral escalation [4] toward the loss of control of opioid use. Consistent with the downward spiral model linking chronic pain to opioid misuse and addiction, here we observed that opioid misuse was associated with spiraling pattern of affective state trajectories reflective of reduced entropy. In the downward spiral underpinning opioid misuse, negentropic processes (underpinned by neuroplasticity in corticolimbic and corticostriatal circuitry) make it difficult to for a given individual to shift out of a negative affective state (e.g., stress, pain, and craving) to a more positive affective state. Thus, negative affective experiences tend to predominate and be conserved in the form of dysphoric moods that are relatively intractable to changes in the environment. In contrast, individuals who take opioids as prescribed may exhibit greater flexibility in responding to changing life conditions, manifested in increased state variability and greater entropy than people who misuse opioids. Thus, opioid misuse represents a condition characterized by relative inflexibility and stereotypic or habitual responses. In that regard, reduced affective and autonomic flexibility has been proposed as a pathogenic risk factor for a wide range of psychiatric conditions, including substance use [43–46].

Combating opioid misuse is a critical and complex problem that requires a multifaceted approach to risk detection. Our proposed pipeline that models both physiological and clinical data can significantly advance risk detection efforts. Firstly, our proposed model analyzed patterns in physiological data to discriminate between patients who misuse opioids from those who take opioids as prescribed by their physicians on the basis of predicted stress, pain, and craving states. Trends in these states over time could indicate a worsening of pain or opioid dependence towards opioid misuse and addiction. By integrating clinical semantic data (such as that which is commonly available in electronic medical records), including opioid prescription and pain diagnostic information, our system can further refine its predictions and identify patients with a higher likelihood of developing opioid misuse problems.

Secondly, our final model enables personalized risk assessments based on individual patient characteristics such as age, gender, psychiatric diagnosis, and opioid prescriptions. Moreover, our state prediction module leverages advanced personalization techniques, namely dynamic grouping via learning-to-branch, which improve performance on estimating stress, pain, and craving. Such a platform enables passive monitoring of opioid misuse-related states, and therefore may be useful in informing future just-in-time adaptive interventions at moments of particularly high opioid misuse risk.

Finally, our proposed pipeline leverages a lightweight neural network to make predictions based on nonlinear dynamical features and semantic embedding from relatively small natural language models. We ensured that the entire pipeline can be deployed on local or portable devices, such as mobile phone or personal laptop, ensuring privacy and minimizing the cost of deployment. Moreover, our approach leverages passive wearable sensing technology already widely accepted in consumer markets Indeed, we used a widely commercially available and inexpensive (e.g., Garmin watch with cost of $100 USD) wearable smartwatch to collect the primary study dataset. Compared to traditional opioid misuse screening methods that require frequent in-person visits, our approach offers a promising alternative by enabling passive inference of misuse risk scores from wearable physiological signals, hence, introduces minimal patient burden and is highly feasible for daily clinical use. Clinically, the data collected through wearable devices can be seamlessly integrated into existing healthcare workflows, automatically syncing with electronic health records and analytic platforms. This setup allows healthcare providers to receive real-time alerts and personalized risk assessments, facilitating timely interventions tailored to individual patients.

As a result, this work could push the field of intelligent opioid use monitoring and intervention one step closer to the actual deployment in clinical settings.

The typical approach to detecting and managing opioid misuse often involves regular on-site screenings and assessments, necessitating significant time and effort in clinical settings. This approach not only places an extra burden on healthcare infrastructure but also lacks the continuous and unobtrusive monitoring essential for detecting opioid misuse in real time. If opioid misuse can only be detected at the intervals of scheduled clinical assessments, opioid misuse events that occur between those intervals may go unaddressed, increasing risk for negative health outcomes, including opioid overdose. Our proposed approach, employing personalized models based on wearable wristbands, analyzing the chaotic features of stress, pain, and craving states, and incorporating semantic clinical data (such as that included in most electronic medical records), offers a less obtrusive alternative for real-time monitoring. By enabling passive monitoring of opioid misuse-related states via physiological data, we mitigate the need for frequent intrusive on-site assessments, respecting individuals' autonomy. Furthermore, continuous passive monitoring allows for detection of opioid misuse risk in moments between scheduled clinical assessment appointments, thereby preventing adverse consequences. The advancements of our approach strategy have also shown in other digital healthcare applications, such as remote patient monitoring in chronic illnesses [47, 48], where a non-invasive, continuous approach has proven more ethical and patient-friendly than frequent hospital visits.

Monitoring of opioid misuse behaviors associated with heightened pain, stress, and craving may help to prevent the development of opioid use disorder. On-body sensor-based monitoring reduces the burden of repeated self-reported measures. In our dataset, we observed that we need, on average, 3 weeks of wearable and EMA data from a user to achieve a strong predictive association between model inferential and self-reported states. This finding has practical implications for clinical implementation. For instance, if a participant is prescribed opioids as part of a surgical procedure (e.g., spinal fusion), medical staff could help ensure high quality data are obtained in the postoperative inpatient stay and for the first follow up appointments post-discharge. Following this period, our system would require minimal supervision and could continue to predict opioid misuse with a high degree of accuracy using passively collected physiological data, making it more practical for long-term monitoring in people with chronic pain.

In addition, our proposed approach provides a more interpretable metric for monitoring of opioid users. Instead of raw HRV features that contain more complex information for interpretation, a granular trajectory of stress, pain, and craving levels throughout the day could better help healthcare professionals develop more effective intervention strategies and mitigate foreseeable health harms.

In a real-world scenario, a privacy consent is mandatory for obtaining the demographic information. As a suggestion for future works, in addition to obtaining a larger dataset, it is also worth exploring disentanglement learning [49] to obtain a generative model. The best practice such as variational autoencoder (VAE) [50] can be leveraged with latent variables being instructed by the demographic data [51]. With such an approach, the pre-trained VAE can be further leveraged for data augmentation under the data-hungry scenario.

3.1 Limitations

The main challenge faced by our current approach is that the model conducting prediction of stress, pain, and craving states makes inference under personalized manner. This approach is mainly proposed to address the inter-subject variability. As a result, when a new user joins the system, the model is facing a cold-start problem, i.e. having difficulty to make accurate predictions due to a lack of initial data. The lack of training data from new users will not only reduce the predictive accuracy for detecting risk in these individuals, but could also potentially distort the model because the model could easily get overfitted on the limited initial data, hence the model could lose generalization. According to our ablation study, we found that approximately 3 weeks of wearable data along with self-reports of stress, pain, and craving would be sufficient to help our model to learn person-specific trends and achieve stable performance.

Finally, to mitigate privacy concerns related to collecting and processing physiological sensor data and electronic health records, we will only use inferred stress, pain, and craving states for risk assessment, avoid storing raw data, and anonymize data to exclude personal identifiable information from the training prediction process.

4 Methods

In this work, we leverage novel deep learning techniques to achieve promising inference of affective states in opioid users including stress, pain, and craving by taking 1 hour inter-beat interval (IBI) and quantitative demographic data as model input. With such a model, we are able to generate dense affective states trajectories across the day(s) which minimize the opioid user's effort on self-reporting.

With such an inference model as prerequisite, we then aim to conduct an opioid misuse risk assessment. Towards this aim, we first leverage chaotic theory [52] to analyze the non-linear dynamic features of the affective state trajectories. Then we developed a lightweight machine learning model to conduct binary classification on whether or not participants show signs of opioid misuse. We also integrate a novel temporal fusion mechanism within the model which allows us, as researchers, to retrieve which affective states trajectory contains the most abnormality for indicating opioid misuse situation.

In addition, we further explore the feasibility of LLM [53] on making semantic inference [54] based on EHR data that are recorded by natural language, which haven't been used by the pipeline introduced so far. We finally fuse this to the pipeline, which enables the system to be able to output opioid misuse risk score, along with affective states trajectory(s) that have high probability of being abnormal, as well as semantic reasoning in natural language format. In summary, the general philosophy behind this step-by-step modeling strategy, from raw wearable sensing data to the final inferential opioid misuse risk, is aligned with empirically validated analytic schema used in several recent machine learning studies [55–57].

Each part of the system, as shown in Figure 1, introduced above will be described in detail in the following subsections.

4.1 Data Collection

4.1.1 Study Protocol

The study was approved by the Institutional Review Board of the University of Utah. We recruited a total of 51 patients with chronic pain on long-term opioid therapy from electronic health records, physician referrals, community advertisements, and online advertisements targeted in the Intermountain West. The inclusion criteria for participants were as follows: proficiency in English, age >= 18, daily prescribed opioid use for >= 3 months, and a reported chronic pain condition as confirmed by physician assessment. Each participant was asked to wear a Garmin Vivosmart 4 smartwatch throughout the day, during which heartbeat (IBI) data were recorded. Participants were then randomized to receive one of two, 8-week long behavioral interventions (Mindfulness-Oriented Recovery Enhancement or supportive group therapy). , to help them better manage their pain and opioid craving. Results from this clinical trial have been reported elsewhere [58]. During the 8 week study treatments, participants were asked to respond to one random EMA probe each day from 7 am to 9 pm. In addition, from 7 am to 9 pm, participants were sent EMA probes when the Garmin Vivosmart automatically detected a deviation in a HRV marker of physiological stress exceeding 1 SD of a participant's moving average value. The statistics of the dataset is shown in Table 2.

The Garmin Vivosmart collects photoplethysmogram (PPG), accelerometer, and pulse oximeter data to detect physiological stress via Firstbeat PRO's heartbeat analysis software [59]. The Firstbeat PRO software estimates the individual's maximal HR (HR_{max}), maximum respiration rate, and maximum oxygen consumption (VO_{2max}) based on participant gender, age, weight, height, and physical activity level. To correctly classify heartbeats, RR-interval data obtained by the Garmin Vivosmart are processed with an artifact detection filter before linear interpolation is used to re-sample artifact-corrected RR-intervals at 5 Hz. The software then removes low frequency trends and variances outside the frequency bands of interest from the re-sampled RR-interval data with a polynomial filter and digital FIR band-pass (.03-1.2 Hz) filter. The software computes time domain and frequency domain HRV variables (e.g., high frequency HRV at .15-.40 Hz). The software uses neural network modeling to segment second-by-second RR-interval data into stationary segments. Prior to detection of physiological stress, the software distinguishes physical activity, postural changes, and movement artifacts from other non-metabolic factors that influence cardiac activity (e.g., stress, pain, craving states). RR intervals and HRV spectral frequency data provide estimates of respiration rate [60] as the frequency of peak power (amplitude) in the respiratory spectrum. Physical activity is detected via accelerometry and VO_2 , which when $> 20\%$ of the participant's VO_{2max} the segment is labeled as activity and excluded from the physiological stress metric. Accelerometer data further refines the detection of postural changes and other body movements.

For non-physical activity, the segments when the body is in a stress state are determined. Stress is defined as a state of sympathetic predominance, indicated by decreases in HRV from the individual's typical resting HRV level and elevated HR. The Firstbeat PRO algorithm-derived physiological stress metric has been validated in multiple other studies [61–67].

4.1.2 Ground Truth Opioid Misuse Measure

After providing informed consent, participants were assessed with demographic and clinical measures, including the Current Opioid Misuse Measure (COMM) [68], a well-validated and broadly used measure of aberrant drug-related behaviors associated with opioid misuse. In a study of chronic pain patients on long-term opioid analgesic therapy, receiver–operator characteristic curve analyses revealed that a score of 9 or higher on the COMM had maximum sensitivity and specificity to identify prescription opioid misuse. Participants with COMM scores ≥ 9 were classified as opioid misusers, whereas those with COMM scores < 9 were classified as prescription users. Our sample was enriched for opioid misusers to enable us to comprehensively study opioid misuse.

4.2 Affective States Prediction

4.2.1 Motivation behind Personalized Prediction

With a preliminary data analysis we observed distinct groups or clusters formed based on various demographic factors exhibiting differing distributions of stress, pain, and craving levels. For each selected demographic and clinical marker, we first conducted k-means clustering where the optimal number of clusters is determined through the elbow method. Then we conducted permutation tests with test statistics of average values on pairwise clusters. The statistical tests revealed several significant associations. Younger participants (20s to 40s) reported higher levels of stress, pain, and craving ($P < .001$). College graduates tended to report lower levels of stress, pain, and craving ($P = .02$). Participants with longer pain durations also reported lower levels of stress, pain, and craving ($P < .01$). Similarly, those with longer durations of opioid use reported lower levels of craving ($P < .01$). Finally, participants with psychiatric disorder diagnoses reported higher levels of stress and craving compared with those without such diagnoses ($P < .001$). The detailed statistical results are presented in Appendix B. These observations are in line with literature in psychology and neuroscience that find individual-specific traits (e.g., genetic, physiological, psychological, and environmental factors) to be significant predictors of affect states [69].

However, demographic factors alone fail to fully explain the opioid misuse risk. We mapped raw sensor data obtained from a wearable wristband to stress, pain, and craving state ratings. The scatter plots in Figure 5 show how different levels of each state are uniformly distributed in different regions of the feature space formed by heart rate variability (HRV) features. Given this high degree of individual variability, we designed a personalization scheme based on dynamic user grouping during the training of computational neural networks for accurate state modeling.

We built a personalized model equipped with a dynamic grouping mechanism to continuously predict each individual's levels of stress, pain, and craving with a window size of 30 minutes. Our model is able to achieve mean absolute error less than 1.5, and Pearson correlation coefficient above 0.7 ($p < .001$) during test time. This predictive model of stress, pain, and craving states is one of the foundational modules in our pipeline of opioid misuse risk assessment.

4.2.2 Personalized Affective State Prediction

As a foundational module in the pipeline of opioid misuse risk assessment, monitoring affective states longitudinally is crucial because the misuse behaviors are highly related to the arouse and suppression of stress, pain, and craving.

We observed, from exploratory data analysis, that there are statistically significant differences in the distribution of affective states across different demographic groups. In addition, with support from the literature, we also know that segmenting users into different groups is an important strategy used to personalize the model. We leverage this observation and the information that lies in the user segmentation to design a personalized affective state prediction model in opioid users.

The most common approach for personalization is to directly fuse the personalized representation to the deep learning model before the final prediction output [70–73]. This approach will serve as one of the baseline methods.

Another intuitive but brute-force approach for personalization is to train or fine-tune a general model for each individual, but this approach underutilizes the information from the full general population and is susceptible to overfitting. To overcome this limitation, we adopt a more scalable approach based on learning to branch where we reserve a small set of personalized parameters in the neural networks for each individual while the remaining parameters are shared across all subjects.

To account for this, we propose a novel dynamic user segmentation mechanism implemented with learning-to-branch [30, 74]. Using the branching technique, the network selects the output from a specific branch for each user to optimize the objective function. This process ensures that the chosen branch's output aligns with the overarching goal of achieving the desired outcome in the network's performance. A visualization of the pipeline is shown in Figure 1, module A. More formally, our deep learning model with hierarchical structure includes shared time-series embedding layers that extract both spatial and temporal features. The branching (group) layers, indicated as $F_{branch\ i}$ in module A, are connected to the embedding layers, followed by individual layers for each user, which are connected to all branching layers with the learn-to-branch mechanism. By training on the task of predicting levels of stress, pain, and craving with heart rate variability data as input, the model learns branches in the network that assign each individual to the optimal group. We denote the output from branching layer i as B_i, the likelihood of the personalized layer j receiving the output from the branch i as $\theta_{i,j}$ as shown in the gray area in module A. Each personalized layer take the output from the branch with the highest likelihood. Then the output, denoted as d_j, which will be received by the personalized layer j is

$$d_j = B^T one\_hot\{argmax_i(\theta_{i,j})\} \tag{1}$$

The argmax function is not differentiable. In order to make $\frac{\partial d_j}{\partial \theta_j}$ differentiable and the results being close to a one-hot vector, we used *Gumbel-trick* [75]. As a result, the branching operation becomes:

$$d_j = B^T \frac{\exp((\log \theta_j + \epsilon)/\tau)}{\sum_k \exp((\log \theta_{k,j} + \epsilon_k)/\tau)} \tag{2}$$

The noise term ϵ is generated from the Gumbel distribution. This helps prevent the network from favoring branches with higher likelihood at initialization, which could lead to gradient saturation issues for branches with lower likelihood. The sharpness of the smoothed one-hot vector is regulated by the temperature parameter, denoted as τ. We begin by initializing τ with a relatively high value and then gradually decrease it in a linear fashion across epochs. This linear decay helps shift the likelihood distribution towards a one-hot vector. The heat map in Appendix G Figure 9 presents an example evolution of branching probabilities during the training. When training is done, users who choose the same branch are considered to be in the same group.

4.3 Misuse Risk Inference

4.3.1 Non-Linear Dynamic Analysis

Nonlinear dynamics features in time series analysis and signal processing, as shown in Figure 1 module B, refer to the phenomena where the behavior of a system or signal is not easily explained by linear relationships and exhibits complex, often chaotic patterns. These features encompass a range of behaviors such as periodicity, bifurcations, and attractors, which are characterized by sensitive dependence on initial conditions and the presence of feedback loops.

Persistence entropy, in this context, quantifies the degree of predictability or randomness within the time series by measuring the rate at which information is lost over time. Higher persistence entropy indicates greater unpredictability and complexity in the time series, while lower entropy suggests more regular and predictable behavior.

Lyapunov exponent is another metric for quantifying the chaotic behavior and sensitivity to initial conditions exhibited by nonlinear dynamical systems. It measures the average rate at which nearby trajectories in phase space diverge over time, providing insight into the system's long-term predictability. A positive Lyapunov exponent indicates exponential divergence of trajectories, suggesting chaotic dynamics where small perturbations grow rapidly, rendering long-term predictions unreliable. Conversely, a negative Lyapunov exponent implies convergence of nearby trajectories, indicative of stability or periodic behavior.

Detrended fluctuation analysis (DFA), as we also explored, is a powerful method used in time series analysis and signal processing to quantify the presence of long-range correlations and fractal-like

properties in data. DFA involves the following steps: first, the time series is integrated to remove its mean and detrended using a polynomial fitting procedure. Then, the fluctuation function is calculated at various time scales, measuring the root-mean-square fluctuation of the detrended data. By analyzing the relationship between these fluctuation values and the corresponding scale, DFA can reveal whether the time series exhibits short-term or long-term correlations.

4.3.2 Relevance Based Temporal Fusion

With inspiration from retrieving memory stream in a previous study [76], we adopt a similar idea and implement it in a novel way to fuse the same philosophy into the neural network, as visually demonstrated in Figure 10. More specifically, we define the recency score W_{rec} of vector at each timestamp t as an exponential decay function:

$$W_{rec}(t) = p^{T-t}, t \in (0, T), p < 1 \tag{3}$$

where T is the total sequence length and p is the decay factor. We then define the relevance score W_{rel} as:

$$W_{rel}(t) = cos\_sim(x_t, \frac{1}{C}\sum_{c=1}^{C} w_c) \tag{4}$$

where x_t is the vector at step t, C is the output number of class, w_c is the weight vector of the linear mapping in the output layer with respect to the output class c, and cos_sim is the cosine similarity calculation function. The integrated weight W is finally defined as:

$$W(t) = \frac{W_{rec}(t) + W_{rel}(t)}{\sum_{i=1}^{T} W_{rec}(i) + W_{rel}(i)} \tag{5}$$

which will result in final output $X = \sum_{t=1}^{T} x_t \cdot W(t)$. Intuitively, the latent representations from all time steps are fused before the final output layer with weighted scores computed according to how close they are to the most current time step and how relevant they are to the objective tasks. No trainable parameters are added with this temporal fusion method, and the output layer is instructed to select the most informative representations to optimize objective tasks.

4.3.3 Language Model Embedding

With the emerging technique of LLM [54], the potential benefits from having LLM driven assistant system are revealed [53]. However, LLM not only introduce much challenges in the local deployment, but also raise the privacy concern that the data will have to be sent to the API provided by lots of third-party service provider company. As a result, we leverage relatively smaller and typical natural language model including ClinicalBERT [77] and ClinicalTinyLlama [78] that have been finetuned on clinical text corpus and can be easily deployed even on edge devices. In our clinical application on opioid users specifically, we leveraging prompt engineering [79], as shown in Figure 1 module C, to integrate text based data including basic demographic information, opioid prescription details, and clinical symptoms associated with pain. We then takes the embedding from language model and leverage these semantic representations for further misuse classification with lightweight linear probing method.

Prompt Engineering: We design the input to the language model with prompt engineering, ensuring the information preservability and the quality of the semantic embedding. All the prompt is coded in markdown format. The example template of prompt compressing EHR data is shown in the following text box, where text enclosed in {} are the placeholders for filling in the corresponding data for each sample:

Basic information:
- This patient is {*age*} years old
- pron's education level is education, and current annual income is {*income*}.

Prescription record
{*nthprescription*}
- Opioid name is {*opioid*}, with each dose being {*dose*}, {*freq*}.
- Recorded usage duration is {duration}.

Current symptom
Current Pain Condition
{*pain_list*}
Self reported severity of {*severity*} out of 10. Duration as reported: "{*years*}"

Current Affective States
- Self reported craving level of {*craving*} out of 10.
- Self reported distress level of {*distress*} out of 10.
- Self reported depression level of {*depress*} out of 10.

5 Conclusion

In conclusion, this study advanced a novel, personalized approach to monitoring opioid misuse risk among individuals on long-term opioid therapy, combining multi-task deep learning with dynamic branching to account for inter-individual variability. By modeling stress, pain, and craving trajectories and integrating nonlinear dynamical features from wearable data with semantic embeddings from electronic health records, we achieved superior predictive accuracy and nuanced insight beyond traditional binary classification. Reduced entropy in opioid misuse-related affective state trajectories suggests diminished regulation of maladaptive attractor states that drives a downward spiral of symptom worsening and behavioral escalation over time. This fusion of physiological data and semantic inferences offers a clinically applicable, scalable tool for early detection and targeted intervention, laying the groundwork for more responsive, evidence-based strategies to mitigate opioid misuse in vulnerable populations.

Declarations

Data Availability Statement Due to the sensitive nature of the data collected in this study, access will be granted only upon request and contingent upon the execution of a formal Data Access Agreement (DAA). This approach ensures compliance with institutional review board (IRB) requirements, participant consent parameters, and applicable privacy regulations (e.g., HIPAA, GDPR). The dataset contains potentially identifiable information and/or clinical variables that, even when de-identified, may pose re-identification risks. Releasing the data through a controlled access mechanism allows the authors to uphold ethical standards of participant confidentiality, monitor appropriate use, and ensure that secondary analyses align with the original scope of consent. Researchers seeking access must demonstrate a legitimate scientific purpose, agree to data use limitations, and commit to maintaining data security protocols. This model balances transparency and reproducibility with the imperative to protect participant privacy and institutional integrity. For the detailed permission to access the data, please contact egarland@health.ucsd.edu.

Acknowledgments We thank the Supercomputing Center and Halıcıoğlu Data Science Institute at University of California San Diego for providing the working space and the computational resources necessary to conduct all experiments. EG was supported by R01DA058621 (PI: Garland), R01AT011772

(PI: Garland), R01DA056537 (PI: Garland), and R01DA057631 (PI: Garland) from the National Institutes of Health during the preparation of this manuscript. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the National Institutes of Health. TR, YL, and BG were supported by NSF Smart and Connected Health (2320678), Google Research Scholar Award (Gift), and Optum Labs Award (UCSD Proposal Number 30223298; KR 46245).
Author Contributions Statement EG and TR are co-principal investigators conceived the original study, signal processing pipeline, and acquired funding and ran the data collection and analysis of the study. MT are the data engineer contribute to the preliminary data processing and analysis. MF are the senior peers contribute to the guidance of notions and methodologies in the clinical and machine learning fields. YL, ID, and BG lead the research, data analysis, data visualization and machine learning model development.
Competing Interests Statement The authors have no conflicts of interests.
Ethical Statement The data was collected from an IRB-approved study at University of Utah with IRB#00078615. Informed consent was received prior to the data collection. The wearable data was fully anonymized and additional privacy safeguards were put in place before data analysis. The application of such predictive models must be accompanied by careful consideration of potential biases, ensuring equitable outcomes for all demographic groups.

Regarding the privacy safeguards, the data were collected with the ilumivu mEMA app, an encrypted application that is compliant with HIPAA, EU-GDPR (EU general data protection regulation), and FDA 21 CFR Part 11 regulations. The ilumivu mEMA app incorporates robust data privacy safeguards to protect research participants and ensure ethical compliance in EMA and wearable studies. Personal identifiers are not stored by ilumivu; instead, participant data are linked via numeric keys maintained solely by the research team, preserving anonymity. All data—including self-report measures, biometric inputs, and location information—are encrypted during transmission to prevent unauthorized access. The app does not sell personal information and only shares data with third parties under legally mandated conditions. Access to collected data is restricted to the approved research team, with ilumivu serving as a data processor rather than a controller. For users in the European Economic Area, the app complies with GDPR regulations. Any reporting or analysis conducted using mEMA data is anonymized and aggregated to prevent individual identification, ensuring that participant confidentiality remains protected throughout the research process.

Tables

		When wearable NLD features are not added		When wearable NLD features are added	
Model	**Features**	**ROC AUC**	**PRC AUC**	**ROC AUC**	**PRC AUC**
Logistic Regression	Demo	**0.606±0.146**	**0.817±0.088**	0.550±0.076	0.758±0.077
XGBoost	Demo	0.571±0.160	0.751±0.084	**0.652±0.133**	**0.790±0.107**
Neural Network	Demo	0.674±0.193	0.824±0.110	**0.694±0.193**	**0.875±0.082**
LSTM	Demo, Affect	0.537±0.185	0.757±0.103	**0.617±0.260**	**0.819±0.116**
ConvTran	Demo, Affect	0.573±0.092	0.782±0.064	**0.614±0.174**	**0.810±0.104**
ClinicalBERT	Demo, Pres, Symp, Affect	0.567±0.142	0.811±0.066	**0.605±0.219**	**0.812±0.112**
ClinicalTinyLlama	Demo, Pres, Symp, Affect	0.807±0.122	0.913±0.062	**0.863±0.108**	**0.940±0.053**

Table 1: Performance on the task of opioid misuse classification. 'Demo', 'Pres', 'Symp' and 'Affect' stands for features capturing respectively demographics, opioid prescription, pain symptom and affective state information. 'NLD' stands for nonlinear dynamics. The mean and standard deviation across the 5 fold cross validations are reported. The peak performance across the two experimental settings (with and without nonlinear dynamical features) are highlighted in bold.

Factors	**Domain**	**Statistics**
Age, mean (STD)		53.2 (12.0)
Gender, n (%)	Male	28 (55)
	Female	23 (45)
College, n (%)	Complete	32 (63)
	Incomplete	19 (37)
OUD, n (%)	Positive	16 (31)
	Negative	35 (69)
Misuse, n (%)	Positive	34 (67)
	Negative	17 (33)
Treatment, n (%)	Supportive	20 (39)
	Mindfulness	31 (61)
Pain duration in years mean (STD)		17.8 (9.5)
Opioid use duration in years mean (STD)		6.4 (6.7)
Wearable data in hours mean (STD)		198.8 (161.2)

Table 2: Demographic factors and statistics of sample population ($N = 51$). The statistics are presented in count (n) followed by percentage (%) or mean with standard deviation (STD).

Figures

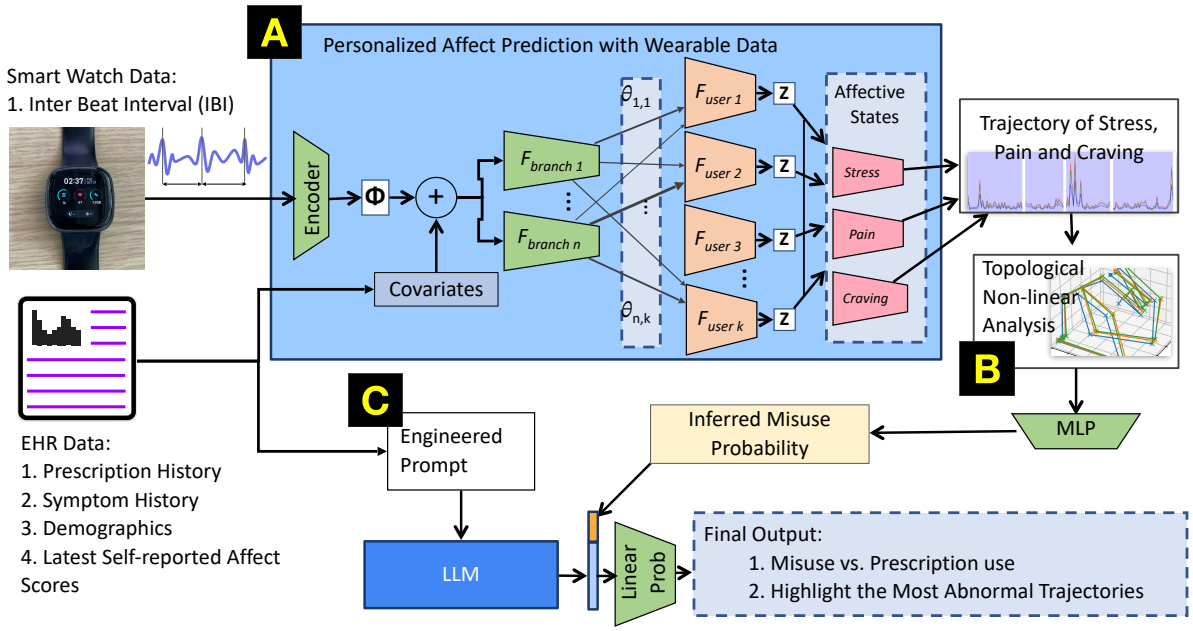

Figure 1: Overview of the entire pipeline. **(A).** Personalized modeling approach for predicting levels of affective states from raw wearable signals. **(B).** Extracting nonlinear dynamical features from the predicted trajectories of affective states across days. **(C).** Output module that fuse the EHR information for predicting risk of having opioid misuse.

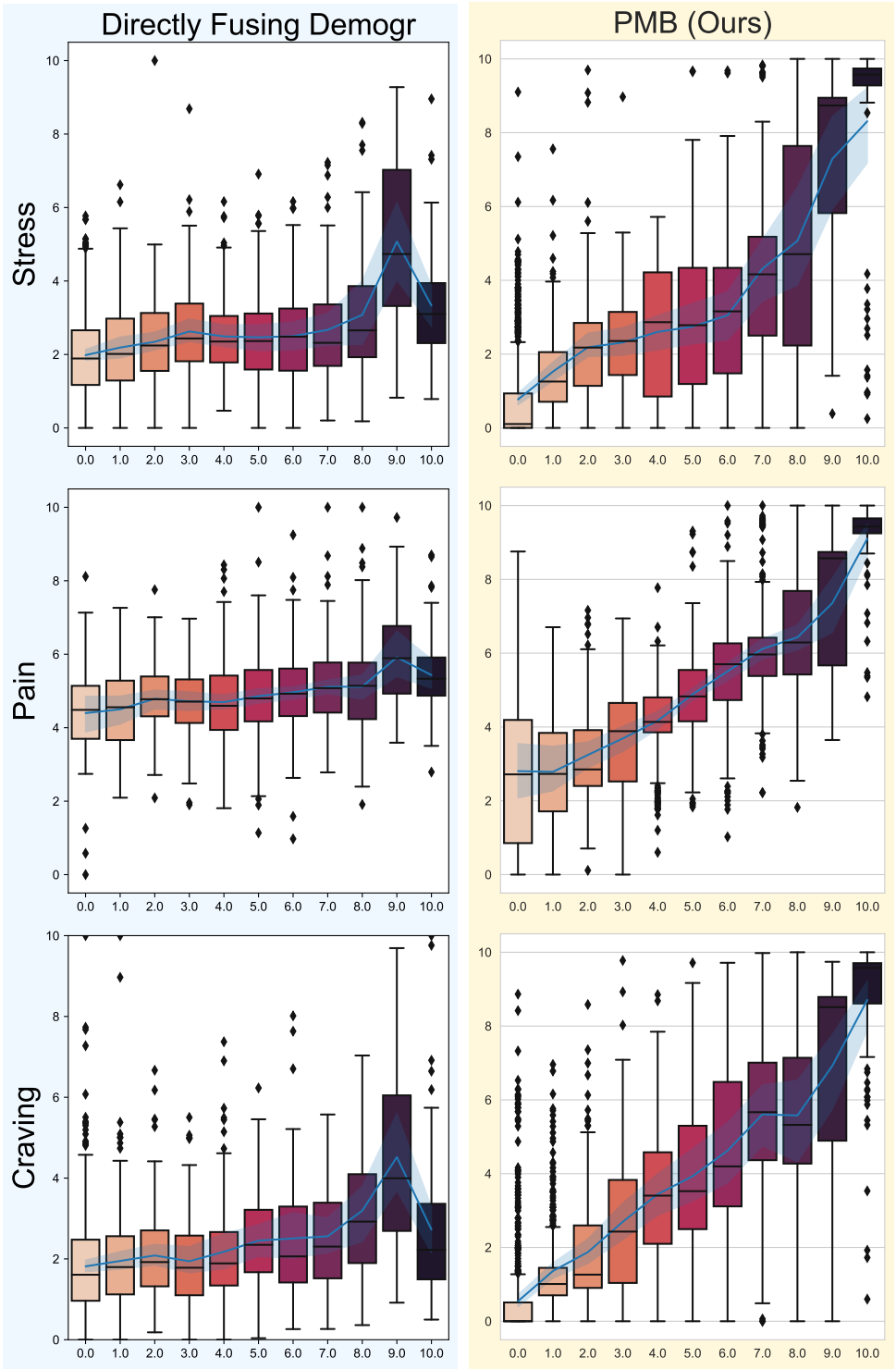

Figure 2: Visualization of test time quality. The x-axis and y-axis are the ground truth and model predicted levels of affective states, within range from 0 to 10, respectively. For each level of affective states (integers from 0 to 10), a box plot of the distribution of the predicted levels are presented. The top and bottom bar stand for the minima and maxima, and the box in the middle represents upper quartile, median, and lower quartile. The scatter points beyond the box plot whiskers represent statistical outliers, defined as values that fall more than 1.5 times the interquartile range above the third quartile or below the first quartile. The column on the left shows the performance of typical approach which directly fusing the personalized characteristics (demographics and clinical characteristics), and column on the right shows the performance from our proposed method.

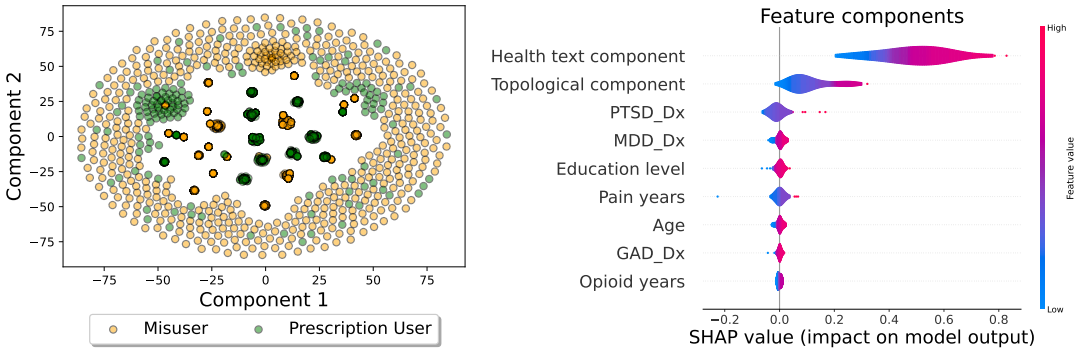

Figure 3: (left) T-SNE plot of test-time embedding from the best model (ClinicalTinyLlama+NNRF inference). (right) SHAP analysis result on the feature contribution. The health text component contains unstructured health related information including text based description of symptoms and opioid prescriptions; Topological component contains the representation from nonlinear dynamic analysis; And the rest are items in the structured tabular data. From this result, we observe that health information and topological representations contribute the most to the inference.

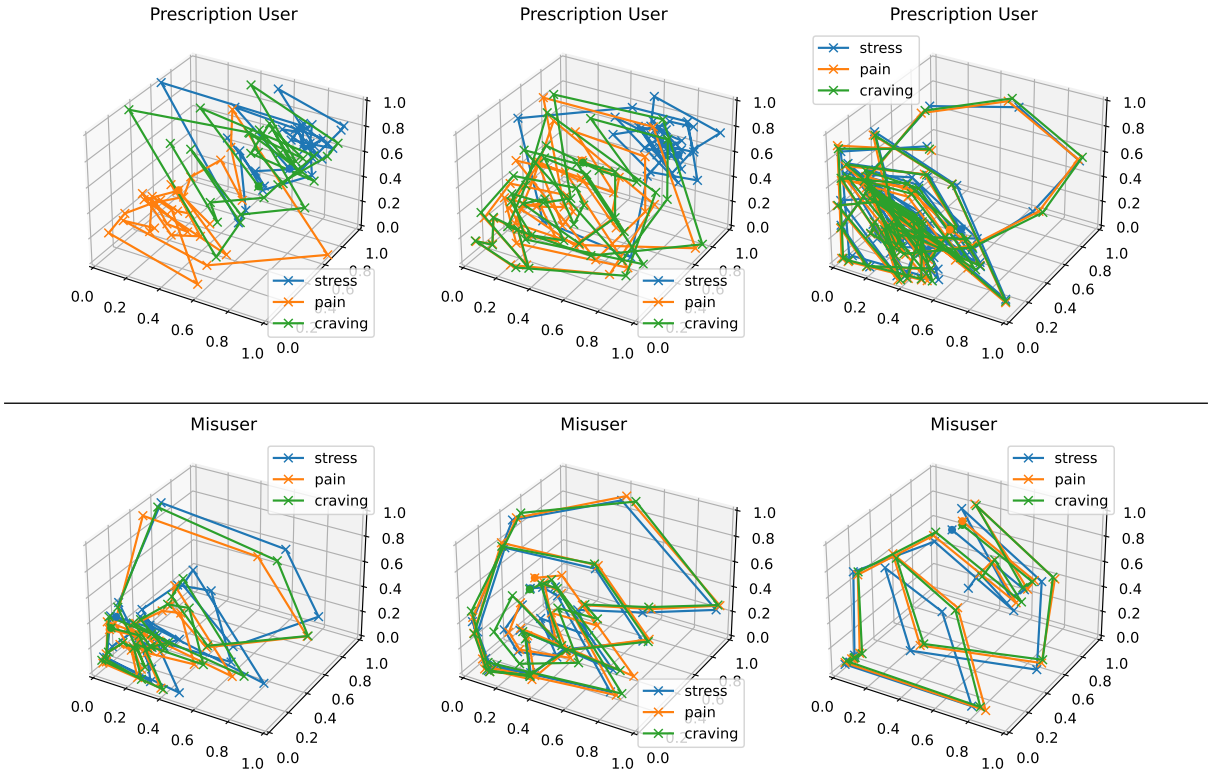

Figure 4: 3D visualization of Takens' embedding from the trajectories of affective states. Opioid users who follow the prescription are more likely to have less predicatable trajectories of affective states along the day comparing to those of misusers. The 3 coordinates indicate respectively inferential level of affective states at time t (i.e., current time), t-1 and t-2 which is also widely known as Takens embedding. The difference of such chaotic pattern is statistically significant as shown in Appendix B in Table 3.

References

[1] Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W.-C., Wang, C.-B., Bernardini, S.: The covid-19 pandemic. Critical reviews in clinical laboratory sciences **57**(6), 365–388 (2020)

[2] NIDA: Drug Overdose Death Rates (2023). `https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates`

[3] Brian Mann, M.B. Aneri Pattani: In 2023 fentanyl overdoses ravaged the U.S. and fueled a new culture war fight (2023). `https://www.npr.org/2023/12/28/1220881380/overdose-fentanyl-drugs-addiction`

[4] Garland, E.L., Froeliger, B., Zeidan, F., Partin, K., Howard, M.O.: The downward spiral of chronic pain, prescription opioid misuse, and addiction: cognitive, affective, and neuropsychopharmacologic pathways. Neuroscience & Biobehavioral Reviews **37**(10), 2597–2607 (2013)

[5] Lazarus, R.S.: Stress and Emotion: A New Synthesis. Springer, ??? (2006)

[6] Price, D.D.: Psychological and neural mechanisms of the affective dimension of pain. Science **288**(5472), 1769–1772 (2000)

[7] Giuliani, N.R., Berkman, E.T.: Craving is an affective state and its regulation can be understood in terms of the extended process model of emotion regulation. Psychological inquiry **26**(1), 48–53 (2015)

[8] Kuppens, P.: It's about time: A special section on affect dynamics. Emotion Review **7**(4), 297–300 (2015)

[9] Vanhasbroeck, N., Ariens, S., Tuerlinckx, F., Loossens, T.: Computational models for affect dynamics. Springer (2021)

[10] Schupp, H.T., Junghöfer, M., Weike, A.I., Hamm, A.O.: The selective processing of briefly presented affective pictures: an erp analysis. Psychophysiology **41**(3), 441–449 (2004)

[11] Buu, A., Cai, Z., Li, R., Wong, S.-W., Lin, H.-C., Su, W.-C., Jorenby, D.E., Piper, M.E.: The association between short-term emotion dynamics and cigarette dependence: A comprehensive examination of dynamic measures. Drug and alcohol dependence **218**, 108341 (2021)

[12] Freeman, W.J.: Societies of Brains: A Study in the Neuroscience of Love and Hate. Psychology Press, ??? (2014)

[13] Haken, H.: Synergetics of brain function. International journal of psychophysiology **60**(2), 110–124 (2006)

[14] Apkarian, A.V., Baliki, M.N., Geha, P.Y.: Towards a theory of chronic pain. Progress in neurobiology **87**(2), 81–97 (2009)

[15] McEwen, B.S.: Allostasis and allostatic load: implications for neuropsychopharmacology. Routledge (2013)

[16] Thayer, J.F., Lane, R.D.: Claude bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. Neuroscience & Biobehavioral Reviews **33**(2), 81–88 (2009)

[17] Thayer, J.F., Lane, R.D.: A model of neurovisceral integration in emotion regulation and dysregulation. Journal of affective disorders **61**(3), 201–216 (2000)

[18] Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y.H., Koo, B.-H.: Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry investigation **15**(3), 235 (2018)

[19] Alinia, P., Sah, R.K., McDonell, M., Pendry, P., Parent, S., Ghasemzadeh, H., Cleveland, M.J.: Associations between physiological signals captured using wearable sensors and self-reported outcomes among adults in alcohol use disorder recovery: development and usability study. JMIR Formative Research **5**(7), 27891 (2021)

[20] Gullapalli, B.T., Natarajan, A., Angarita, G.A., Malison, R.T., Ganesan, D., Rahman, T.: On-body sensing of cocaine craving, euphoria and drug-seeking behavior using cardiac and respiratory signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **3**(2), 1–31 (2019)

[21] Bertz, J.W., Epstein, D.H., Preston, K.L.: Combining ecological momentary assessment with objective, ambulatory measures of behavior and physiology in substance-use research. Addictive behaviors **83**, 5–17 (2018)

[22] Määttänen, I., Henttonen, P., Väliaho, J., Palomäki, J., Thibault, M., Kallio, J., Mäntyjärvi, J., Harviainen, T., Jokela, M.: Positive affect state is a good predictor of movement and stress: Combining data from esm/ema, mobile hrv measurements and trait questionnaires. Heliyon **7**(2) (2021)

[23] Murray, D.W., Ridenour, T.A., Swingler, M.M., Morgan, A., Hegarty-Craver, M.: Feasibility of combining biosensor and ecological momentary assessment to measure stress experiences among economically disadvantaged adolescents. Stress and Health **39**(3), 684–689 (2023)

[24] Jiang, M., Mieronkoski, R., Rahmani, A.M., Hagelberg, N., Salanterä, S., Liljeberg, P.: Ultra-short-term analysis of heart rate variability for real-time acute pain monitoring with wearable electronics. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1025–1032 (2017). IEEE

[25] Kalkman, G.A., Brink, W., Pierce, M., Atsma, F., Vissers, K.C., Schers, H.J., Dongen, R.T., Kramers, C., Schellekens, A.F.: Monitoring opioids in europe: the need for shared definitions and measuring drivers of opioid use and related harms. European Addiction Research **28**(3), 231–240 (2022)

[26] Gudin, J.A., Mogali, S., Jones, J.D., Comer, S.D.: Risks, management, and monitoring of combination opioid, benzodiazepines, and/or alcohol use. Postgraduate medicine **125**(4), 115–130 (2013)

[27] Gullapalli, B.T., Carreiro, S., Chapman, B.P., Ganesan, D., Sjoquist, J., Rahman, T.: Opitrack: a wearable-based clinical opioid use tracker with temporal convolutional attention networks. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies **5**(3), 1–29 (2021)

[28] Benyamin, R., Trescot, A.M., Datta, S., Buenaventura, R.M., Adlaka, R., Sehgal, N., Glaser, S.E., Vallejo, R.: Opioid complications and side effects. Pain physician **11**(2S), 105 (2008)

[29] Tejeda, H.A., Bonci, A.: Dynorphin/kappa-opioid receptor control of dopamine dynamics: Implications for negative affective states and psychiatric disorders. Brain research **1713**, 91–101 (2019)

[30] Luo, Y., Deznabi, I., Shaw, A., Simsiri, N., Rahman, T., Fiterau, M.: Dynamic clustering via branched deep learning enhances personalization of stress prediction from mobile sensor data. Scientific Reports **14**(1), 6631 (2024)

[31] Stark, J., Broomhead, D.S., Davies, M.E., Huke, J.: Takens embedding theorems for forced and stochastic systems. Nonlinear Analysis: Theory, Methods & Applications **30**(8), 5303–5314 (1997)

[32] Yan, Y., Huang, Y.-C., Zhao, J., Liu, Y.-S., Ma, L., Yang, J., Yan, X.-D., Xiong, J., Wang, L.: Topological nonlinear analysis of dynamical systems in wearable sensor-based human physical activity inference. IEEE Transactions on Human-Machine Systems **53**(4), 792–801 (2023) `https://doi.org/10.1109/THMS.2023.3275774`

[33] Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A.: Determining lyapunov exponents from a time series. Physica D: nonlinear phenomena **16**(3), 285–317 (1985)

[34] Hu, K., Ivanov, P.C., Chen, Z., Carpena, P., Stanley, H.E.: Effect of trends on detrended fluctuation analysis. Physical Review E **64**(1), 011114 (2001)

[35] Qian, B., Rasheed, K.: Hurst exponent and financial market predictability. In: IASTED Conference on Financial Engineering and Applications, pp. 203–209 (2004). Proceedings of the IASTED International Conference Cambridge, MA

[36] Garland, E.L., Froeliger, B., Howard, M.O.: Allostatic dysregulation of natural reward processing in prescription opioid misuse: Autonomic and attentional evidence. Biological psychology **105**, 124–129 (2015)

[37] Garland, E.L., Bryan, C.J., Nakamura, Y., Froeliger, B., Howard, M.O.: Deficits in autonomic indices of emotion regulation and reward processing associated with prescription opioid use and misuse. Psychopharmacology **234**, 621–629 (2017)

[38] Garland, E.L., Howard, M.O.: Prescription opioid misusers exhibit blunted parasympathetic regulation during inhibitory control challenge. Psychopharmacology **238**, 765–774 (2021)

[39] Roberts, R.L., Garland, E.L.: Association between opioid use disorder and blunted heart rate variability among opioid-treated chronic pain patients. Addiction biology **27**(6), 13230 (2022)

[40] Koob, G.F., Le Moal, M.: Drug addiction, dysregulation of reward, and allostasis. Neuropsychopharmacology **24**(2), 97–129 (2001)

[41] Shurman, J., Koob, G.F., Gutstein, H.B.: Opioids, pain, the brain, and hyperkatifeia: a framework for the rational use of opioids for pain. Pain Medicine **11**(7), 1092–1098 (2010)

[42] Elman, I., Borsook, D.: Common brain mechanisms of chronic pain and addiction. Neuron **89**(1), 11–36 (2016)

[43] Friedman, B.H.: An autonomic flexibility–neurovisceral integration model of anxiety and cardiac vagal tone. Biological psychology **74**(2), 185–199 (2007)

[44] Ottaviani, C., Medea, B., Lonigro, A., Tarvainen, M., Couyoumdjian, A.: Cognitive rigidity is mirrored by autonomic inflexibility in daily life perseverative cognition. Biological psychology **107**, 24–30 (2015)

[45] Eddie, D., Price, J.L., Bates, M.E., Buckman, J.F.: Substance use and addiction affect more than the brain: The promise of neurocardiac interventions. Current addiction reports **8**(3), 431–439 (2021)

[46] Zhu, Z., Bonanno, G.A.: Affective flexibility: Relations to expressive flexibility, feedback, and depression. Clinical Psychological Science **5**(6), 930–942 (2017)

[47] De Guzman, K.R., Snoswell, C.L., Taylor, M.L., Gray, L.C., Caffery, L.J.: Economic evaluations of remote patient monitoring for chronic disease: a systematic review. Value in Health **25**(6), 897–913 (2022)

[48] El-Rashidy, N., El-Sappagh, S., Islam, S.R., M. El-Bakry, H., Abdelrazek, S.: Mobile health in remote patient monitoring for chronic diseases: Principles, trends, and challenges. Diagnostics **11**(4), 607 (2021)

[49] Eddahmani, I., Pham, C.-H., Napoléon, T., Badoc, I., Fouefack, J.-R., El-Bouz, M.: Unsupervised learning of disentangled representation via auto-encoding: A survey. Sensors **23**(4), 2362 (2023)

[50] Ghojogh, B., Ghodsi, A., Karray, F., Crowley, M.: Factor analysis, probabilistic principal component analysis, variational inference, and variational autoencoder: Tutorial and survey (2021)

[51] Ding, Z., Xu, Y., Xu, W., Parmar, G., Yang, Y., Welling, M., Tu, Z.: Guided variational autoencoder for disentanglement learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7920–7929 (2020)

[52] Thompson, J.M.T., Stewart, H.B., Turner, R.: Nonlinear dynamics and chaos. Computers in Physics **4**(5), 562–563 (1990)

[53] Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., Wen, J.-R.: A Survey of Large Language Models (2023)

[54] Huang, J., Chang, K.C.-C.: Towards Reasoning in Large Language Models: A Survey (2023)

[55] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., Zhou, D.: Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of the 35th International Conference on Neural Information Processing Systems, pp. 24824–24837. Curran, New Orleans, LA, USA and online (2022).

[56] Wang, B., Min, S., Deng, X., Shen, J., Wu, Y., Zettlemoyer, L., Sun, H.: Towards understanding chain-of-thought prompting: An empirical study of what matters. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) ACL:2023:long, pp. 2717–2739. acl, Toronto, Canada (2023). https://doi.org/10.18653/v1/2023.acl-long.153

[57] Luo, Y., Liu, Y., Cai, R., Rahman, T.: Start simple: Progressive difficulty multitask learning. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 4: Student Research Workshop), pp. 48–55 (2024)

[58] Garland, E.L., Gullapalli, B.T., Prince, K.C., Hanley, A.W., Sanyer, M., Tuomenoksa, M., Rahman, T.: Zoom-based mindfulness-oriented recovery enhancement plus just-in-time mindfulness practice triggered by wearable sensors for opioid craving and chronic pain. Mindfulness **14**(6), 1329–1345 (2023)

[59] Myllymaki, T.: Stress and recovery analysis method based on 24-hour heart rate variability (2014)

[60] Charlton, P.H., Bonnici, T., Tarassenko, L., Clifton, D.A., Beale, R., Watkinson, P.J.: An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. Physiological measurement **37**(4), 610 (2016)

[61] Föhr, T., Tolvanen, A., Myllymäki, T., Järvelä-Reijonen, E., Rantala, S., Korpela, R., Peuhkuri, K., Kolehmainen, M., Puttonen, S., Lappalainen, R., *et al.*: Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: a cross-sectional study. Journal of Occupational Medicine and Toxicology **10**, 1–9 (2015)

[62] Madison, A., Vasey, M., Emery, C.F., Kiecolt-Glaser, J.K.: Social anxiety symptoms, heart rate variability, and vocal emotion recognition in women: evidence for parasympathetically-mediated positivity bias. Anxiety, Stress, & Coping **34**(3), 243–257 (2021)

[63] Myllymäki, T., Kyröläinen, H., Savolainen, K., Hokka, L., Jakonen, R., Juuti, T., Martinmäki, K., Kaartinen, J., Kinnunen, M.-L., Rusko, H.: Effects of vigorous late-night exercise on sleep quality and cardiac autonomic activity. Journal of sleep research **20**(1pt2), 146–153 (2011)

[64] Palesh, O., Scheiber, C., Kesler, S., Gevirtz, R., Heckler, C., Guido, J.J., Janelsins, M., Cases, M.G., Tong, B., Miller, J.M., *et al.*: Secondary outcomes of a behavioral sleep intervention: A randomized clinical trial. Health Psychology **38**(3), 196 (2019)

[65] Rusko, H., Rönkä, T., Uusitalo, A., Kinnunen, U., Mauno, S., Feldt, T., Kinnunen, M., Martinmäki, K., Hirvonen, A., Hyttinen, S., *et al.*: Stress and relaxation during sleep and awake time, and their associations with free salivary cortisol after awakening. In: Nordic Ergonomics Society Congress (2006)

[66] Teisala, T., Mutikainen, S., Tolvanen, A., Rottensteiner, M., Leskinen, T., Kaprio, J., Kolehmainen, M., Rusko, H., Kujala, U.M.: Associations of physical activity, fitness, and body composition with heart rate variability–based indicators of stress and recovery on workdays: a cross-sectional study. Journal of Occupational Medicine and Toxicology **9**, 1–9 (2014)

[67] Uusitalo, A., Mets, T., Martinmäki, K., Mauno, S., Kinnunen, U., Rusko, H.: Heart rate variability related to effort at work. Applied ergonomics **42**(6), 830–838 (2011)

[68] Butler, S.F., Budman, S.H., Fernandez, K.C., Houle, B., Benoit, C., Katz, N., Jamison, R.N.: Development and validation of the current opioid misuse measure. Pain **130**(1-2), 144–156 (2007)

[69] Tobin, R.M., Graziano, W.G., Vanman, E.J., Tassinary, L.G.: Personality, emotional experience, and efforts to control emotions. Journal of personality and social psychology **79**(4), 656 (2000)

[70] Hu, Y., Liu, Y., Zhang, S., Zhang, T., Dai, B., Peng, B., Yang, H., Dai, Y.: A cross-space cnn with customized characteristics for motor imagery eeg classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 1554–1565 (2023) https://doi.org/10.1109/TNSRE.2023.3249831

[71] Bansal, S., Gowda, K., Kumar, N.: Multilingual personalized hashtag recommendation for low resource indic languages using graph-based deep neural network. Expert Systems with Applications **236**, 121188 (2024) https://doi.org/10.1016/j.eswa.2023.121188

[72] Jiang, M., Rosio, R., Salanterä, S., Rahmani, A.M., Liljeberg, P., da Silva, D.S., de Albuquerque, V.H.C., Wu, W.: Personalized and adaptive neural networks for pain detection from multi-modal physiological features. Expert Systems with Applications **235**, 121082 (2024) https://doi.org/10.1016/j.eswa.2023.121082

[73] Chang, J., Zhang, C., Hui, Y., Leng, D., Niu, Y., Song, Y., Gai, K.: Pepnet: Parameter and embedding personalized network for infusing with personalized prior information. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '23, pp. 3795–3804. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3580305.3599884 . https://doi.org/10.1145/3580305.3599884

[74] Guo, P., Lee, C.-Y., Ulbricht, D.: Learning to Branch for Multi-Task Learning (2020)

[75] Jang, E., Gu, S., Poole, B.: Categorical Reparameterization with Gumbel-Softmax (2017)

[76] Park, J.S., O'Brien, J.C., Cai, C.J., Morris, M.R., Liang, P., Bernstein, M.S.: Generative Agents: Interactive Simulacra of Human Behavior (2023)

[77] Wang, G., Liu, X., Ying, Z., Yang, G., Chen, Z., Liu, Z., Zhang, M., Yan, H., Lu, Y., Gao, Y., *et al.*: Optimized glycemic control of type 2 diabetes with reinforcement learning: a proof-of-concept trial. Nature Medicine **29**(10), 2633–2642 (2023)

[78] Muzammil, M.: Finetuning EnDevSols/tinyllama-2.5T-Clinical model on Clinical Dataset. (2021). https://huggingface.co/muzammil-eds/tinyllama-2.5T-Clinical-v2 Accessed 2023-03-19

[79] Wang, J., Shi, E., Yu, S., Wu, Z., Ma, C., Dai, H., Yang, Q., Kang, Y., Wu, J., Hu, H., Yue, C., Zhang, H., Liu, Y., Li, X., Ge, B., Zhu, D., Yuan, Y., Shen, D., Liu, T., Zhang, S.: Prompt Engineering for Healthcare: Methodologies and Applications (2023)

Appendix A Final List of Heart Rate Variablity (HRV) Features

Domain	Feature Names
Time	hr_mean, hr_std, sdnn, rmssd
Frequency	fft_peak_VLF, fft_peak_LF, fft_peak_HF, fft_abs_VLF, fft_abs_LF, fft_rel_VLF, fft_rel_LF, fft_rel_HF, fft_log_LF, fft_log_HF, fft_norm_VLF, lomb_peak_VLF, lomb_peak_LF„ lomb_peak_HF, lomb_rel_VLF, lomb_rel_LF„ lomb_log_VLF, lomb_log_LF, lomb_log_HF, ar_rel_LF, ar_ratio
Non-linear	sd_ratio, dfa_alpha1, dfa_alpha2

with default setting of frequency filtering bands:

- VLF: [0.00Hz - 0.04Hz]
- LF: [0.04Hz - 0.15Hz]
- HF: [0.15Hz - 0.40Hz]

Appendix B Exploratory Data Analysis

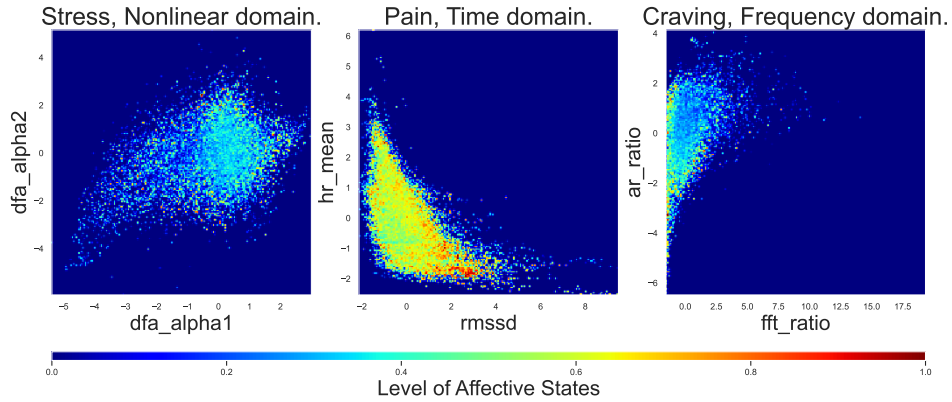

Figure 5: Illustrates how affective state levels are distributed in the feature space. Each point in the heat map indicate the average level of affective states given the pair of HRV features. 'dfa_alpha' are the nonlinear dynamic features extracted from detrended fluctuation analysis. 'RMSSD' and 'HR_mean' are the time domain HRV features. 'fft_ratio' and 'ar_ratio' are the ratio between LF and HF as frequency domain features extracted from Fourier transform and autoregressive method. From left, to right, the plots were made with respect to stress, pain and craving respectively.

Before conducting statistical analysis, we first conduct typical user segmentation using K-Means. The optimal numbers of demographic groups are found using elbow rule as shown in Figure 6. Each sub figures shows the final distortion values when running K-Means with different number of cluster centers. We then conduct pair-wise statistical comparison of the distribution of levels of stress, pain, and craving among different demographic groups. The statistical results are shown below,

The values in the cells in the table below follow the following order from top to down: alternative hypothesis of relationship from row i to column j, followed by permutation tests' p-values of distributions of stress, pain, and craving. The statistics is the raw value of average level of row j subtracted from the average level of row i.

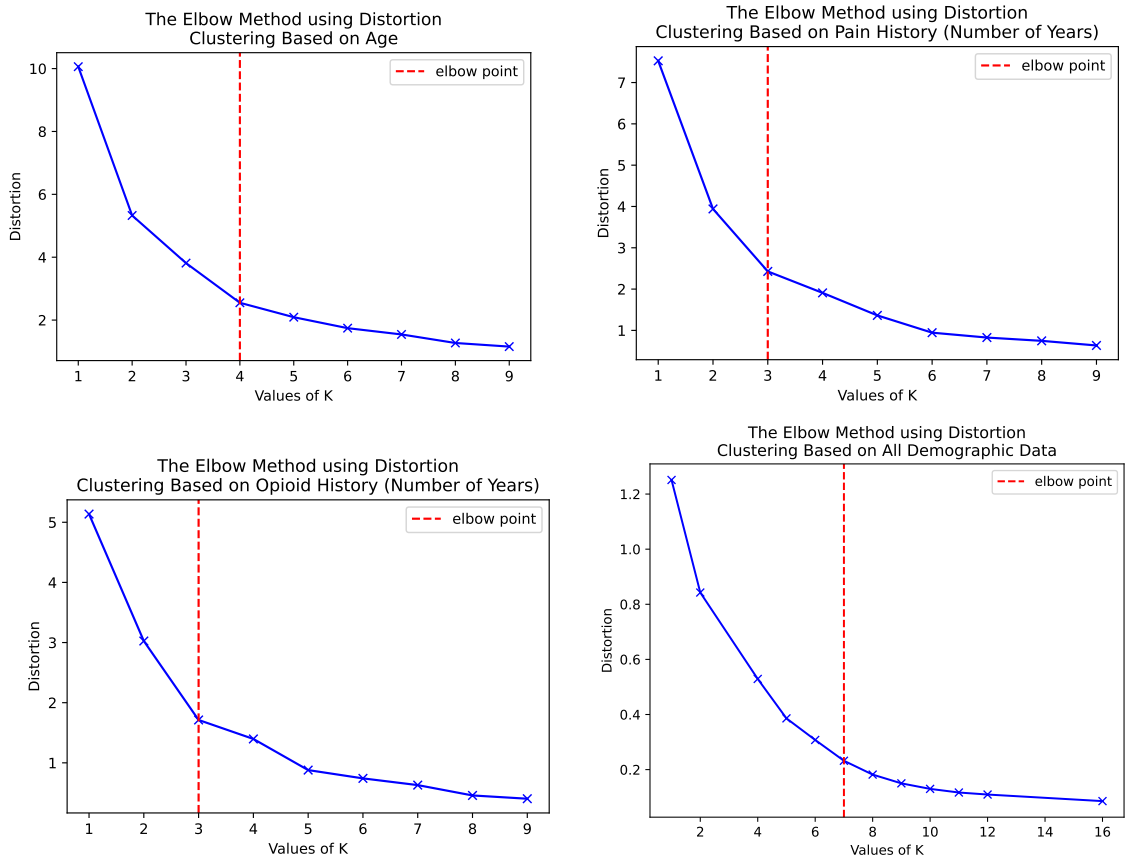

Figure 6: **Demonstration of using elbow rule to identify optimal number of groups when clustering based on different demographic information.**

Age Groups	38 n=12	48 n=17	61 n=15	74 n=7
38	-	greater $P < .001$ $P < .001$ $P < .001$	greater $P < .001$ $P < .001$ $P < .001$	greater $P < .001$ $P < .001$ $P < .001$
48	-	-	greater $P < .001$ $P = .983$ $P = 1.0$	greater $P < .001$ $P < .001$ $P = 1.0$
61	-	-	-	greater $P = .099$ $P < .001$ $P = .999$

where the age number of each group stands for the center value, the exact ranges are: [27, 43], [44, 53], [55, 56], [70, 79] respectively. From the table above we observe that the younger the age, the higher the self-reported stress where we could saw a gradually decreased trend. Older users tends to have lower level of craving which can be shown in the first row that the ratings from age around 38 are statistically greater than any other age groups.

Education Groups	1 n=32	0 n=19
1	-	less $P = .018$ $P < .001$ $P = .031$

where 1 represents having completion of college, and 0 otherwise. If we use 0.05 as significant threshold, then users who complete college got lower ratings in all the moods comparing to users without college completion.

Pain Duration Groups (years)	8 n=16	17 n=22	31 n=13
8	-	greater $P < .001$ $P = 0.008$ $P < .001$	greater $P < .001$ $P < .001$ $P < .001$
17	-	-	greater $P < .001$ $P < .001$ $P = 0.008$

where the years of pain duration of each group stands for the center value, the exact ranges are: [2, 12], [13, 23], and [25, 40] respectively. With threshold of 0.05, each pain duration group is statistically greater than the next group with longer duration.

Opioid Duration Groups (years)	2 n=33	10 n=13	22 n=5
2	-	greater $P = 0458$ $P = 0.188$ $P = 0.006$	greater $P < .001$ $P < .001$ $P < .001$
10	-	-	greater $P < .001$ $P < .001$ $P < .001$

where the years of opioid usage duration of each group stands for the center value, the exact ranges are: [0.25, 6], [7, 15], and [18, 30] respectively. For opioid usage duration, only craving level demonstrate a clear decreasing trend. For stress and pain, we can only conclude that users with shorter opioid usage duration have statistically higher rating of moods.

Mental Dx Groups	1 n=16	2 n=27	3 n=8
1	-	less $P < .001$ $P = .890$ $P = .938$	less $P < .001$ $P < .001$ $P < .001$
2	-	-	less $P < .001$ $P < .001$ $P < .001$

where group 1 stands for no other mental disorder being diagnosed; group 2 stands for MDD Dx only; and group 3 stands for MDD Dx and PTSD Dx and/or GAD Dx. We observe that the more the number of mental disorders being diagnosed, the higher the stress. And no mental disorders have significantly less self-reported pain and craving levels.

	Prescription User n=17		Misuser n=34		
Affective State	**Mean**	**SD**	**Mean**	**SD**	***P***
Stress H0	4.24	0.51	4.03	0.58	**.003**
Stress H1	1.67	0.77	1.51	0.71	.064
Pain H0	4.19	0.60	4.02	0.56	**.021**
Pain H1	1.69	0.73	1.52	0.67	**.040**
Craving H0	4.19	0.59	4.05	0.57	**.036**
Craving H1	1.78	0.69	1.50	0.67	**.002**

Table 3: Results from permutation test. The statistics being used is the result of subtracting the mean of misuser group from the mean of normal user group, with alternative hypothesis that the normal users have more unpredicatable (higher value) of the corresponding nonlinear dynamics features. 'H0' and 'H1' stands for persistence entropy on zero and one dimension homology respectively.

Appendix C Ablation Study

	Stress		Pain		Craving	
Methods	**MAE ↓**	**PCC ↑**	**MAE ↓**	**PCC ↑**	**MAE ↓**	**PCC ↑**
DCNN	2.213±0.084	0.300±0.082	1.857±0.05	0.221±0.043	2.193±0.099	0.323±0.076
DCNN.d	2.238±0.075	0.312±0.04	1.854±0.033	0.240±0.036	2.212±0.085	0.316±0.047
DCNN.p	**1.648±0.052**	**0.662±0.028**	**1.517±0.063**	**0.611±0.043**	**1.466±0.093**	**0.779±0.034**
ConvTran	2.298±0.079	0.289±0.038	1.860±0.066	0.201±0.056	2.286±0.0921	0.304±0.081
ConvTran.p	**1.857±0.081**	**0.666±0.043**	1.892±0.116	**0.657±0.024**	**1.486±0.065**	**0.783±0.026**

Table 4: **Personalized approach.** Bold numbers representing improved performance. MAE and PCC stand for Mean-Absolute-Error and Pearson Correlation Coefficient. '.d' indicates the concatenation of demographic to the model latent space for personalized calibration. '.p' indicates personalization being added. ↑ indicates that higher values in the corresponding column are better, and ↓ indicates that lower values are preferable. The mean and standard deviation across the 5 fold cross validations are reported, with the peak performance highlighted in bold.

	Stress		Pain		Craving	
Methods	**MAE ↓**	**PCC ↑**	**MAE ↓**	**PCC ↑**	**MAE ↓**	**PCC ↑**
DCNN.p	1.648±0.052	**0.662±0.028**	1.517±0.063	0.611±0.043	1.466±0.093	0.779±0.034
DCNN.p.bd	1.662±0.044	0.650±0.037	1.340±0.076	0.669±0.039	1.309±0.107	0.777±0.040
DCNN.p.b	**1.646±0.048**	0.654±0.020	1.359±0.046	0.660±0.023	1.359±0.061	0.763±0.033
DCNN.p.b.da	1.677±0.028	0.628±0.023	**1.325±0.062**	**0.673±0.027**	1.357±0.059	0.759±0.024
DCNN.p.b.db	1.658±0.061	0.649±0.035	1.351±0.084	0.660±0.034	**1.280±0.084**	**0.797±0.032**

Table 5: **Fuse demographic information. Bold numbers represent the best performance.** '.b' stands for having a branching module, and '.bd' stands for the fixed branch assignment based on clustering from demographic data. '.da' and '.db' stand for fuse demographic data after and before the branching module respectively.

	Stress		Pain		Craving	
Methods	**MAE ↓**	**PCC ↑**	**MAE ↓**	**PCC ↑**	**MAE ↓**	**PCC ↑**
PMB.bd	1.555±0.046	0.686±0.0213	1.289±0.051	0.693±0.024	1.233±0.070	0.800±0.031
PMB.db	1.479±0.060	0.689±0.027	1.251±0.068	0.716±0.034	1.143±0.137	**0.810±0.043**
PMB.db.r	1.459±0.052	0.694±0.023	1.258±0.070	0.718±0.033	1.158±0.104	0.805±0.040
PMB.db.rr	**1.440±0.064**	**0.695±0.026**	**1.213±0.061**	**0.723±0.030**	**1.126±0.117**	**0.810±0.038**

Table 6: **Results of Actual Multitask Learning and Novel Temperal Fusion.** Bold numbers represent the best performance. PMB stands for Personalized Multitask Learning Model with Branching. '.r' indicates the inclusion of temporal fusion with recency as weights, and '.rr' signifies the presence of relevance scores as well.

Metric/aggregation method	w/ NNRF	Mean Pool	Max Pool
Nonlinear Dynamical Classifier, AUC ROC	**0.630 ± 0.155**	0.546 ± 0.181	0.537 ± 0.212
Nonlinear Dynamical Classifier, AUC PRC	**0.790 ± 0.114**	0.765 ± 0.110	0.755 ± 0.117

Table 7: **Ablation study: Performance with and without NNRF** The result shows that representation fusion with the proposed NNRF significantly outperforms typical aggregation rule.

Filter	**ROC AUC**	**PRC AUC ↑**
All Data	0.734	0.860
1 most relevant trajectory	0.638	0.815
5 most relevant trajectory	0.692	0.836
8 most relevant trajectory	0.623	0.817
1 least relevant trajectory	0.557	0.767
5 least relevant trajectory	0.587	0.804
most relevant and least relevant	0.637	0.765
5 relevant and least relevant	0.702	0.832
most relevant and 5 least relevant	0.491	0.697

Table 8: **Ablation study: Top-k (least) Relevant Trajectories**

Feature Type	**Feature Name**	**PRC AUC ↑**
Time Domain	mean, std, max/min, skew, kurtosis	0.757
Frequency Domain	centroid, spread, max freq, mean power	0.745
Time Domain & Frequency Domain	mean, std, max/min, skew, kurtosis, centroid, spread, max freq, mean power	0.709
Nonlinear Dynamics	Lyapunov Exponent (LE)	**0.796**
	Detrended Fluctuation Analysis (DFA)	**0.795**
	Persistence Entropy (PE)	**0.808**
Nonlinear Dynamics & Time Domain	PE, mean, std, max/min, skew, kurtosis	0.680
Nonlinear Dynamics & Frequency Domain	PE, centroid, spread, max freq, mean power	0.752
Nonlinear Dynamics	PE+LE	**0.809**
Nonlinear Dynamics	PE+DFA	**0.795**

Table 9: Ablation study on different combinations of non linear dynamics features.

Appendix D R-squared following Tables Result Tables

	Stress	Pain	Craving
Methods	$R^2 \uparrow$	$R^2 \uparrow$	$R^2 \uparrow$
DCNN	0.048±0.066	-0.027±0.027	0.067±0.052
DCNN.p	**0.401±0.037**	**0.265±0.073**	**0.481±0.082**
ConvTran	0.049±0.026	-0.023±0.055	0.056±0.076
ConvTran.p	**0.299±0.101**	-0.055±0.091	**0.481±0.056**

Table 10: R^2 **scores of adding personalized module.**

	Stress	Pain	Craving
Methods	$R^2 \uparrow$	$R^2 \uparrow$	$R^2 \uparrow$
DCNN.p	0.401±0.037	0.265±0.073	0.481±0.082
DCNN.p.bd	0.400±0.058	0.434±0.060	0.588±0.066
DCNN.p.b	0.399±0.029	0.421±0.038	0.540±0.051
DCNN.p.b.da	0.370±0.021	**0.443±0.039**	0.549±0.038
DCNN.p.b.db	**0.402±0.045**	0.425±0.053	**0.598±0.053**

Table 11: R^2 **scores following Table 5.**

	Stress	Pain	Craving
Methods	$R^2 \uparrow$	$R^2 \uparrow$	$R^2 \uparrow$
PMB	0.430±0.069	0.473±0.053	0.631±0.087
PMB.r	**0.436±0.061**	0.466±0.058	0.623±0.076
PMB.rr	**0.436±0.060**	**0.507±0.051**	**0.641±0.075**

Table 12: R^2 **scores following Table 6.**

Appendix E Feature Contributions

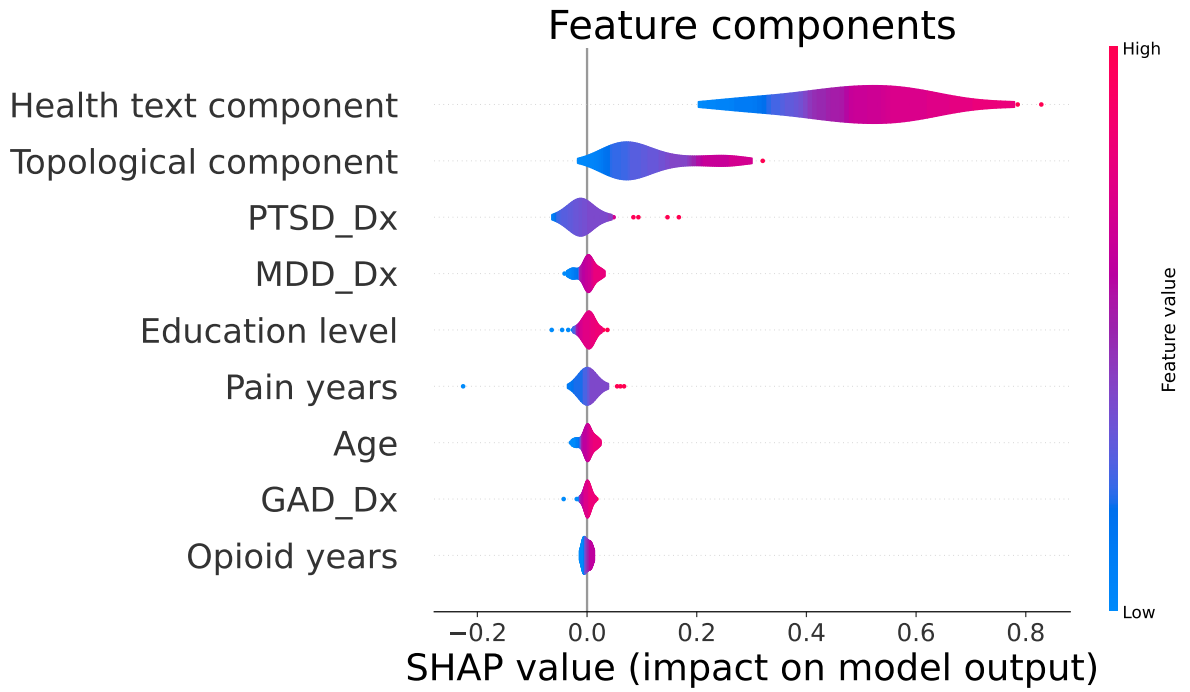

Figure 7: **SHAPLEY analysis.** The health text component contains unstructured health related information including text based description of symptoms and opioid prescriptions; Topological component contains the representation from nonlinear dynamic analysis; And the rest are items in the structured tabular data. From this result, we observe that health information and topological representations contribute the most to the inference.

Deep learning time series feature	Nonlinear Dynamical Feature	Quantitative Health Info	Text based Health Info	AUC ROC	AUC PRC	Rel Diff
✗	✗	✓	✗	0.674 ± 0.193	0.824 ± 0.110	-
✓	✗	✓	✗	0.573 ± 0.092	0.782 ± 0.064	-9.55%
✗	✓	✓	✗	0.717 ± 0.134	0.871 ± 0.071	+6.01%
✗	✓	✗	✗	0.630 ± 0.155	0.790 ± 0.114	-5.21%
✗	✗	✓	✓	0.807 ± 0.122	0.913 ± 0.062	+14.82%
✗	✓	✓	✓	0.863 ± 0.108	0.940 ± 0.053	+4.83%

Table 13: **Feature contribution to inference,** showing the influence on performance from adding or removing each component.

Appendix F External Validation

F.1 External Dataset 1

Regarding justifying the modeling methodology on a broader scope, we've transferred the model and re-run the experiments on another dataset. The new dataset is from 111 opioid users with chronic pain who were engaged in cognitive tasks for a duration of approximately 10 minutes. During the task, an ECG signal was collected. Brief sample characteristics of this dataset are presented in Table 14.

Factors (N=111)	Domain	Statistics
Age, Mean (SD), years		56.0 (12.5)
Gender, n (%)	Male Female	50 (45) 61 (55)
Opioid Misuse Status, n (%)	Positive for Opioid Misuse Negative for Opioid Misuse	58 (52) 53 (48)
Data interval, Mean (SD),minutes		10.8 (11.0)

Table 14: **Demographic factors and statistics of sample population (N = 111).**

Due to relatively short duration of the physiological signal data available in this dataset, we leverage ultra-brief HRV which is calculated from inter-beat-interval data extracted from ECG within a window size of 30 seconds instead of 5 minutes as we used for our main dataset in the paper. Our experiments conducted on this dataset focus on the influence of different components (including raw HRV, nonlinear dynamical affective states features, and the clinical and demographic information) to the performance of opioid misuse prediction. All the rest of the configurations of model structure and training settings remain the same.

We conduct similar ablation studies on this secondary dataset, and the similar behavior and trends of performance are observed from Table 15.

HRV Features	Nonlinear Dynamical Feature	Clinical and Demographic Information	AUC ROC	AUC PRC	Rel Diff
✗	✓	✗	0.513±0.11	0.582±0.076	-
✓	✗	✗	0.626±0.089	0.693±0.102	-
✓	✓	✗	0.679±0.05	0.729±0.039	+8.47%
✗	✗	✓	0.490±0.016	0.518±0.016	-
✗	✓	✓	0.538±0.095	0.592±0.072	+9.80%
✓	✗	✓	0.626±0.089	0.693±0.102	-
✓	✓	✓	0.679±0.05	0.729±0.039	+8.47%

Table 15: **Feature contribution to inference,** showing the influence on performance from adding or removing each component.

From the table we observe that adding nonlinear dynamical features always brings statistically significant improvement in the performance. We also conducted a permutation statistical test on the persistence entropy as we did in the original dataset. Informed by the findings from the main study dataset, we hypothesized that people with opioid misuse will show less entropy (more predictable change in inferential affective states from the HRV signal) than people who take opioids as prescribed. We observed that the persistence entropy at 1 dimension homology computed from the inferential trajectory of pain levels rejects the null hypothesis with statistical significance ($P < .05$).

Finally, we conduct SHAP analysis on this secondary dataset, and the results are shown in Figure 8. The features in this Figure are sorted by their importance from top to bottom. We observe that the model relies heavily on the persistence entropy of the inferential trajectories of levels of craving and pain at zero dimension homology. The clinical and demographic text information components did not help with the performance in this dataset, so we exclude them from the figures for cleanliness.

F.2 External Dataset 2

HRV Features	Nonlinear Dynamical Feature	AUC ROC	AUC PRC	Rel Diff
✗	✓	0.617 ± 0.256	0.829 ± 0.138	-
✓	✗	0.700 ± 0.292	0.818 ± 0.185	-
✓	✓	**0.750 ± 0.316**	**0.851 ± 0.199**	+7.14%

Table 16: **Feature contribution to inference,** showing the effect of including (✓) or excluding (✗) specific components on model performance.

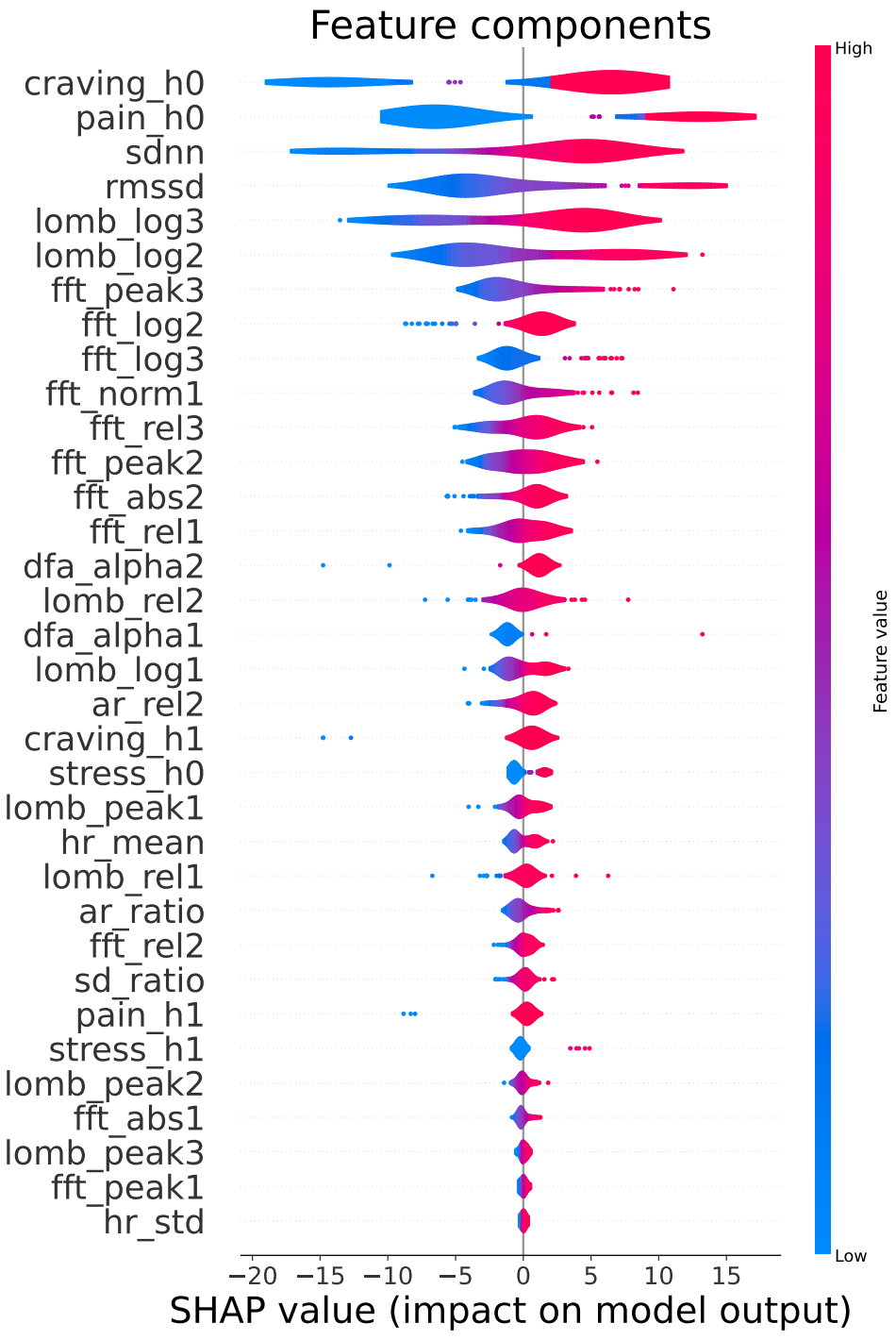

Figure 8: **SHAPLEY analysis on External Dataset 1.**

Appendix G Supplementary Content and Visualization of Model's Components

During the training of the learning-to-branch module, all the parameters in the model are optimized through backpropagation along with gradient descent method. Regarding the branching module specifically, with the above formula 2 as reference describing the branching mechanism, we have the gradient being defined as:

$$\frac{\partial B}{\partial d_j} = \frac{\exp((\log \theta_j + \epsilon)/\tau)}{\sum_k \exp((\log \theta_{k,j} + \epsilon_k)/\tau)} \tag{6}$$

where $\frac{\partial B}{\partial d_j}$ will be further utilize for optimizing the parameter of each branch layer F_{branch} which is implemented by typical multi-layer perceptron. We also have the gradient for the θ, which is the probability distribution over the branches, defined as:

$$\frac{\partial \theta_{i,j}}{\partial d_j} = \frac{B_i \cdot \exp((\log \theta_{i,j} + \epsilon)/\tau)}{\theta_{i,j}\tau \sum_k \exp((\log \theta_{k,j} + \epsilon_k)/\tau)} - \frac{B_i \cdot \exp((\log \theta_{i,j} + \epsilon)/\tau)}{\theta_{i,j}\tau (\sum_k \exp((\log \theta_{k,j} + \epsilon_k)/\tau))^2} \tag{7}$$

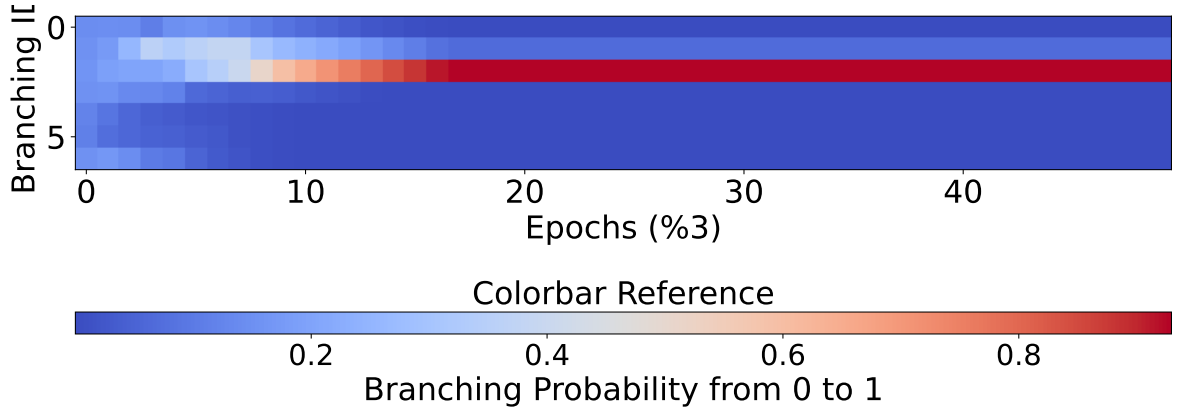

Figure 9: Evolution of branching parameters during the training. X-axis is the number of epochs. Each row of y-axis represent the changing of probabilities of each branch. The higher the probability, the brighter the cell color. The values of the probability is calculated according to formula 2 without sampling the noise term ϵ.

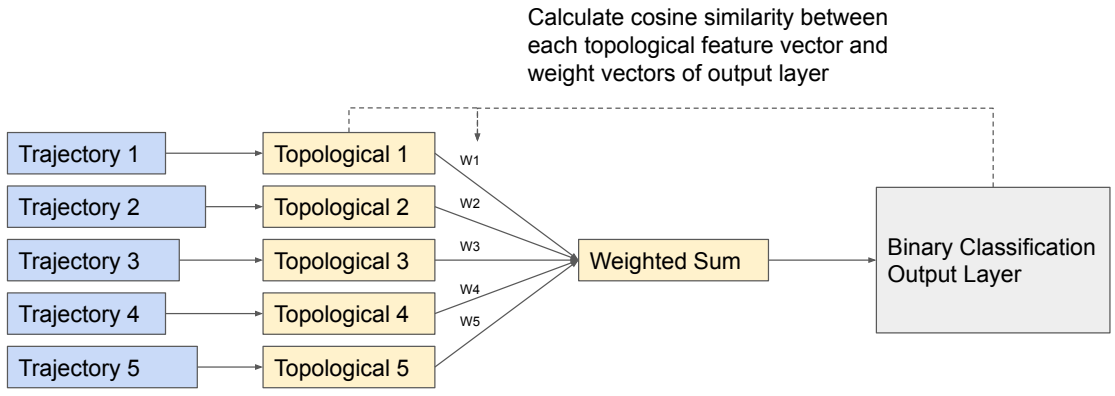

Figure 10: Novel Temporal Fusion Mechanism. The recency of representations from each time step and their relevance to the objective task are taking into consideration during fusion.

	Attributes			Stress		Pain		Craving	
Methods	P	B	D	MAE ↓	PCC ↑	MAE ↓	PCC ↑	MAE ↓	PCC ↑
ConvTran	✗	✗	✗	2.30±0.08	0.29±0.04	1.86±0.07	0.20±0.06	2.29±0.09	0.30±0.08
DCNN	✗	✗	✗	2.21±0.08	0.30±0.08	1.86±0.05	0.22±0.04	2.19±0.10	0.32±0.08
DCNN w/ D	✓	✗	✓	2.24±0.08	0.31±0.04	1.85±0.03	0.24±0.04	2.21±0.09	0.32±0.05
PDCNN w/o D (Ours)	✓	✓	✗	1.65±0.05	0.65±0.02	1.36±0.05	0.66±0.02	1.36±0.06	0.76±0.03
PDCNN (Ours)	✓	✓	✓	1.66±0.06	0.65±0.04	1.35±0.08	0.66±0.03	1.28±0.08	0.80±0.03
PMB w/o B (Ours)	✓	✗	✗	1.65±0.05	0.66±0.03	1.52±0.06	0.61±0.04	1.47±0.09	0.78±0.03
PMB (Ours)	✓	✓	✓	**1.44±0.06**	**0.70±0.03**	**1.21±0.06**	**0.72±0.03**	**1.13±0.12**	**0.81±0.04**

Table 17: Personalized approach. For each metric the best value achieved is indicated as bold. MAE and PCC stand for Mean-Absolute-Error and Pearson Correlation Coefficient. 'P' stands for belonging to the category of personalized approaches. 'B' indicates the inclusion of having Learning to Branch for dynamic user segmentation. 'D' stands for fuse information from demographic data. PDCNN stands for Personalized DCNN, and PMB stands for Personalized Multitask Learning Model with Branching.