OpiTrack: A Wearable-based Clinical Opioid Use Tracker with Temporal Convolutional Attention Networks

BHANU TEJA GULLAPALLI, University of Massachusetts Amherst, USA
STEPHANIE CARREIRO, Division of Medical Toxicology, Department of Emergency Medicine University of Massachusetts Medical School, USA
BRITTANY P CHAPMAN, Division of Medical Toxicology, Department of Emergency Medicine University of Massachusetts Medical School, USA
DEEPAK GANESAN, University of Massachusetts Amherst, USA
JAN SJOQUIST, University of Massachusetts Medical School, USA
TAUHIDUR RAHMAN, University of Massachusetts Amherst, USA

Opioid use disorder is a medical condition with major social and economic consequences. While ubiquitous physiological sensing technologies have been widely adopted and extensively used to monitor day-to-day activities and deliver targeted interventions to improve human health, the use of these technologies to detect drug use in natural environments has been largely underexplored. The long-term goal of our work is to develop a mobile technology system that can identify high-risk opioid-related events (i.e., development of tolerance in the setting of prescription opioid use, return-to-use events in the setting of opioid use disorder) and deploy just-in-time interventions to mitigate the risk of overdose morbidity and mortality. In the current paper, we take an initial step by asking a crucial question: Can opioid use be detected using physiological signals obtained from a wrist-mounted sensor? Thirty-six individuals who were admitted to the hospital for an acute painful condition and received opioid analgesics as part of their clinical care were enrolled. Subjects wore a noninvasive wrist sensor during this time (1-14 days) that continuously measured physiological signals (heart rate, skin temperature, accelerometry, electrodermal activity, and interbeat interval). We collected a total of 2070 hours ($\approx$ 86 days) of physiological data and observed a total of 339 opioid administrations. Our results are encouraging and show that using a Channel-Temporal Attention TCN (CTA-TCN) model, we can detect an opioid administration in a time-window with an F1-score of 0.80, a specificity of 0.77, sensitivity of 0.80, and an AUC of 0.77. We also predict the exact moment of administration in this time-window with a normalized mean absolute error of 8.6% and R^2 coefficient of 0.85.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing**;

Additional Key Words and Phrases: Opioid administration, Physiological signal, Depthwise convolutions, Temporal convolutional network, Channel and Temporal Attention

Authors' addresses: Bhanu Teja Gullapalli, University of Massachusetts Amherst, Amherst, MA, 01002, USA, bgullapalli@cs.umass.edu; Stephanie Carreiro Division of Medical Toxicology, Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA, 01655, USA, Stephanie.Carreiro@umassmed.edu; Brittany P Chapman Division of Medical Toxicology, Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA, 01655, USA, Brittany.Chapman@umassmed.edu; Deepak Ganesan, University of Massachusetts Amherst, Amherst, MA, 01002, USA, dganesan@cs.umass.edu; Jan Sjoquist, University of Massachusetts Medical School, Worcester, MA, 01655, USA, Jan.Sjoquist@umassmed.edu; Tauhidur Rahman, University of Massachusetts Amherst, Amherst, MA, 01002, USA, trahman@cs.umass.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2021 Association for Computing Machinery.
2474-9567/2021/9-ART102 $15.00
https://doi.org/10.1145/3478107

ACM Reference Format:
Bhanu Teja Gullapalli, Stephanie Carreiro, Brittany P Chapman, Deepak Ganesan, Jan Sjoquist, and Tauhidur Rahman. 2021. OpiTrack: A Wearable-based Clinical Opioid Use Tracker with Temporal Convolutional Attention Networks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 102 (September 2021), 29 pages. https://doi.org/10.1145/3478107

1 INTRODUCTION

Opioids are a class of drugs that bind to receptors in the brain and body to produce a variety of physiologic effects [66]. They include prescription opioids (i.e., morphine and oxycodone) as well as illicit drugs, such as heroin. When taken into the body, opioids attach to receptors located in the brain, spinal cord, and other parts responsible for the feeling of pain [70]. As a result, they block pain signals sent from the brain to other parts of the body and vice versa, leading to pain relief (analgesia). Other short-term effects include feeling relaxed and drowsy (sedation), shallow breathing (respiratory depression), and euphoria. Intake of opioids also causes large amounts of dopamine release, stimulating reward pathways involved in addiction. Repeated use of opioids can lead to an individual developing uncomfortable symptoms of withdrawal when the drug is discontinued (physical dependence) and requiring increasingly large amounts to experience the same effect (tolerance). So while opioids can be used therapeutically to treat moderate-to-severe pain, they also carry the risk of long-term consequences such as misuse, dependence, tolerance, and addiction [5]. For this reason, opioid misuse is often preceded by legitimate use for medical purposes. This can make healthcare providers reluctant to treat pain and cause patient suffering. In the year 2018, roughly 70% of drug overdose deaths in the United States (US) involved an opioid [96]. A growing number of these deaths are related to opioid use disorder (OUD), commonly referred to as "opioid addiction."

The current standard of care for monitoring opioid use includes patient self-report and testing for opioid metabolites in biological specimens (typically urine); both methods suffer from significant limitations [46, 60]. Self-report is subject to underreporting due to recall bias or intentional concealment of use. Urine drug screens are retrospective, have limited windows of detection (24-48 hours for many opioids), can be manipulated and tampered with, and do not provide a detailed timeline of use patterns [29, 36]. Currently, no objective measure exists to monitor opioid use (or misuse) in real-time. Provided with such a technology, we could prevent opioid-related deaths and increase the safety of opioid prescribing.

Wearable sensors present a particularly attractive possibility given their unobtrusive nature and increasing acceptability in the general population. In recent years, there has been increasing interest in leveraging commercially available wearable and mobile sensing devices for both detecting substance use in real-time and enabling just-in-time interventions [15, 33, 63, 65]. However, much of the prior work has focused on detecting tobacco, alcohol, and cocaine use [15, 33, 65] as opposed to opioids. A recent effort studied opioid overdose detection [63], but opioid overdose is very different from broader opioid use. The prior existing work on detecting opioid use with wearables is based on a small amount of data that centered around pre-opioid exposure [53].

In this paper, we comprehensively explore the possibility of using a wrist-mounted sensor that captures physiological signals continuously and passively to detect opioid administrations in subjects admitted to the hospital for an acute painful condition. Our work has several key contributions:

- **We collect a unique large dataset of opioid administrations.** We collected a series of longitudinal datasets (observing multiple opioid events in the same subjects over a period of 1-14 days) from a total of 36 subjects. In total, our data consisted of 2070 hours of wearable sensor data containing a total of 339 intravenous opioid administrations. To our knowledge, a dataset of this magnitude and size has never been collected in the context of opioid use before.
- **We detect changes occurring in physiological signals during an opioid administration.** We find features from the previously reported signals to be significant during opioid administration, reproducing the results from prior works. Along with these changes, we also observed the features of electrodermal

activity and heart rate variability derived from interbeat interval to demonstrate significant changes during opioid use.

- **The physiologic changes occurring during opioid use are distinct from changes noted with other routine daily activities.** Although the study environment was controlled (in-hospital), subjects were free to perform limited activities of daily living throughout the data collection period (i.e., get out of bed to a chair, walk in the hall, interact with other people). The results from our current work support the notion that detection of opioid use can occur unobtrusively and with high accuracy despite this background noise. These findings are encouraging as we move toward systems that can detect opioid use in real-time in natural environments beyond clinical settings.
- **We propose a novel Channel-Temporal Attention TCN (CTA-TCN) architecture, which jointly predicts whether and when an opioid administration occurred.** To handle time-sequences of large size, we used a TCN architecture, and to aid the model in attending to only the important time points, a temporal-based self-attention block was used. To better handle the inter-modality differences and to weigh their contributions, we used Depthwise convolutions and Channel attention block. We also devised a hybrid loss function consisting of weighted cross-entropy and weighted kappa, specific to our problem.
- **We systematically analyze the performance of our model across different demographic subgroups.** We observed slightly improved performance of the model in males and in individuals who were opioid-naive compared to those with a history of chronic use. The performance of the model was found to be uniform across all the body mass index (BMI) subgroups. Lastly, an attribution-based explainable AI technique has been used to get further insight into the proposed temporal convolutional networks.

To our knowledge, this is the first work to demonstrate that by attending to unique opioid-related changes in multimodal physiological signals from a wrist-mounted sensor, opioid administration moments can be accurately detected and precisely labeled in a longitudinal dataset collected from a clinical population. This work provides new opportunities to create technologies for real-time opioid use (or misuse) monitoring and just-in-time interventions for opioid addiction mitigation.

2 RELATED WORK

In the context of health and wellness, wearable devices have already made a significant impact in detecting and monitoring various health conditions. For example, researchers have studied activity recognition [4, 43] and stress detection [18, 76] over the past decade using smartwatches, chest bands, and mobile devices embedded with location, electrocardiogram (ECG), and inertial sensors. A significant challenge in these studies has involved gathering high-quality labeled data. However, with recent machine learning advancements, research has shifted toward using self-supervised [74, 75] and unsupervised [48, 97] based methods to optimize the use of unlabeled data. In regards to monitoring health conditions, [50, 68] have used multimodal wearable technologies for early diagnosis of Parkinson's disease and [44, 77, 92] developed cardiac monitoring systems that notify the user or their physician of cardiac abnormalities such as hypertension or arrhythmia. Various other works have also looked into the applicability of wearable devices for the early detection of migraine attacks [47] and Alzheimer's disease [93]. In regards to behavioral health and wellbeing, [35, 72] used physiological signals from wearable sensors to detect different emotion states, and [3, 10] were able to classify different stages of sleep for sleep quality assessments. In parallel, the impact of wearable devices has diversified and grown. Our current work focuses on using a wrist-mounted sensor to detect opioid administrations in a clinical setting using an architecture based on temporal convolution network (TCN) with channel-and-temporal attention. Usage of TCN's in wearable devices is an active ongoing research area, and it is already used in the contexts of activity recognition detection [61] and hand movement classification via surface electromyographic signal [100]. Unlike these past works, in our case, we consider a multi-modality-based signal captured from different sensors of the wearable device . We

use channel-and-temporal based attention modules to help the model in attending to the different modalities and different time-steps based on their importance [6, 94, 98]. This section reviews and summarizes the existing literature available on detecting substance use with wearable sensors.

Opioid Use: The current literature surrounding the detection of opioid use is largely focused on return-to-use events ("relapse") during treatment for opioid use disorder and monitoring adherence during chronic opioid therapy. Both utilize self-reporting, and laboratory testing methodologies, which are subject to recall bias and limited by imprecision [27, 42]. Recent work by Nandakumar et al.[63] described detection of opioid toxicity and overdose using a contactless sensor system. Their study converted a smartphone into an active sonar monitoring system to detect opioid-induced respiratory depression among hospitalized surgical patients and individuals using drugs in a supervised injection facility. To our knowledge, the only existing work using wearable sensors to detect opioid use was previously conducted on 30 emergency department (ED) patients who received opioid analgesics as part of their clinical care [53]. The authors used a wristband sensor (Q sensor, Affectiva, Waltham, MA) to collect physiological data before and after a single dose of an opioid analgesic was administered. A decision tree classifier was applied to identify physiologic patterns associated with opioid use. While promising, the generalizability of these results was limited due to a low number of opioid administrations captured (one per subject) and a small amount of data recorded that centered only around pre-opioid exposure. On the contrary, in our current work, we collect multiple days of continuous physiological data from each subject, containing numerous opioid administrations. This provides a more robust dataset to study the physiological signals associated with opioid use.

Tobacco Use: The detection of tobacco use has previously been achieved using wearable sensors containing inertial measurement units (IMUs) and Respiratory Inductive Plethysmograph (RIP) bands. Many of these works use features associated with the act of smoking (i.e., hand-to-mouth gestures, chest wall contraction and expansion) as opposed to biomarkers or physiologic changes associated with nicotine use [1, 69, 73, 80, 83, 89]. While two of the works reported promising results using RIP bands in a laboratory-based setting, their models performed poorly once tested in the natural environment [1, 80]. This was attributed to a high susceptibility to noise when subjects were performing concurrent activities. Later works attempted to address these shortcomings by incorporating IMUs, RIP bands, and a "smart cigarette lighter" [40, 81]. They also used deep learning methods to extract features and better combine multimodal information automatically. Significant improvement was observed in cigarette smoking detection both in and out of the lab. Unlike opioids, the effect smoking cigarettes has on the human body (change in respiration) is immediate and well-studied [25, 57, 90], which motivated many works to develop models based on hand-crafted features. Cigarette smoking sessions also have a hand-to-mouth gesture associated with it, making tobacco use easier to detect than opioids.

Alcohol Use: Electrochemical detection strategies seem to be at the forefront when it comes to sensor-based methods for alcohol use. These transdermal sensors work to detect target analytes in sampled biofluids (i.e., sweat, interstitial fluid) for passive, non-invasive monitoring of alcohol use and intoxication. While the concentration of analytes in interstitial fluid and sweat have been closely correlated to concurrent blood alcohol levels, detection is delayed due to the time it takes for alcohol to be expelled in the biofluids measured [12]. While both alcohol and opioids affect the physiology of the body when taken, strong biomarkers exist in alcohol, making its detection immediate and easier compared to opioids.

Cocaine Use: Cocaine is a common drug of abuse that is mainly sought out for its stimulant and euphoric effects. When used, it acts on the central nervous system to inhibit the reuptake of catecholamines resulting in supra-physiologic potentiation of sympathetic nervous system (SNS) activity [28]. Physiologic effects include vasoconstriction (narrowing of the blood vessels), mydriasis (dilated pupils), increased heart rate, and increased

blood pressure [95]. Prior literature has demonstrated the ease and effectiveness of measuring several of these physiologic changes using wearable sensors. Hossain et al. [38] used a device that included ECG and accelerometry sensors to develop a physiologically-informed computational model for automated detection of cocaine use in controlled and natural environments. Natarajan et al. [64] later expanded on this work to strengthen the lab-to-field generalization performance of models. Carreiro et al. [13] used a wearable sensor that measured skin temperature, electrodermal activity, and locomotion to detect physiologic markers of SNS arousal. In their study of 15 subjects, they used sensor data to identify distinct episodes of cocaine use that were missed by both self-report and urine drug screening.

3 USER STUDY AND PRELIMINARY ANALYSIS

3.1 General Study Protocol

This was an observational study approved by the Institutional Review Board of the University of Massachusetts Medical school. A convenience sample of 36 adults admitted to the University of Massachusetts Medical center was utilized. Subjects were identified while receiving care in the emergency department, prior to hospital admission, and screened for inclusion/exclusion criteria. Once a potential subject was deemed eligible, a research assistant approached them, and informed consent was obtained.

Inclusion criteria were: 18 years of age or older, admission to the hospital for pancreatitis (inflammation of the pancreas), clinical plan of care including opioid analgesics for pain management, able to speak English, and capable of providing informed consent. Exclusion criteria were: physical inability to wear a wrist-mounted sensor (i.e., upper extremity amputation or fracture), being under police custody, or currently being pregnant.

This study focused only on adult subjects as pediatric subjects may have physiologic differences in drug response and should be studied separately. Pancreatitis is a common medical condition characterized by severe pain and is routinely treated in the hospital with opioid analgesics, thus represented an excellent target population for the study. Adults lacking capacity, pregnant women, and prisoners are considered vulnerable populations and are routinely excluded from such studies. Overall details of the subjects recruited for the study are shown in the below table,

Table 1. Demographic factors and statistics of 36 subjects. The statistics is presented either in the form of mean and standard deviation (SD) values or count (n) and percentage values (%).

Factors		**Statistics**
Gender, n (%)	Male	21 (58)
	Female	15 (42)
	Other	0 (0)
Age, mean (SD)		50.6 (14.6)
BMI, n (%)	Normal	12 (34)
	Overweight	11 (32)
	Obese	12 (34)
Opioid use classification, n (%)	Naive	14 (39)
	Occasional	7 (19)
	Chronic	15 (42)

3.2 Hardware

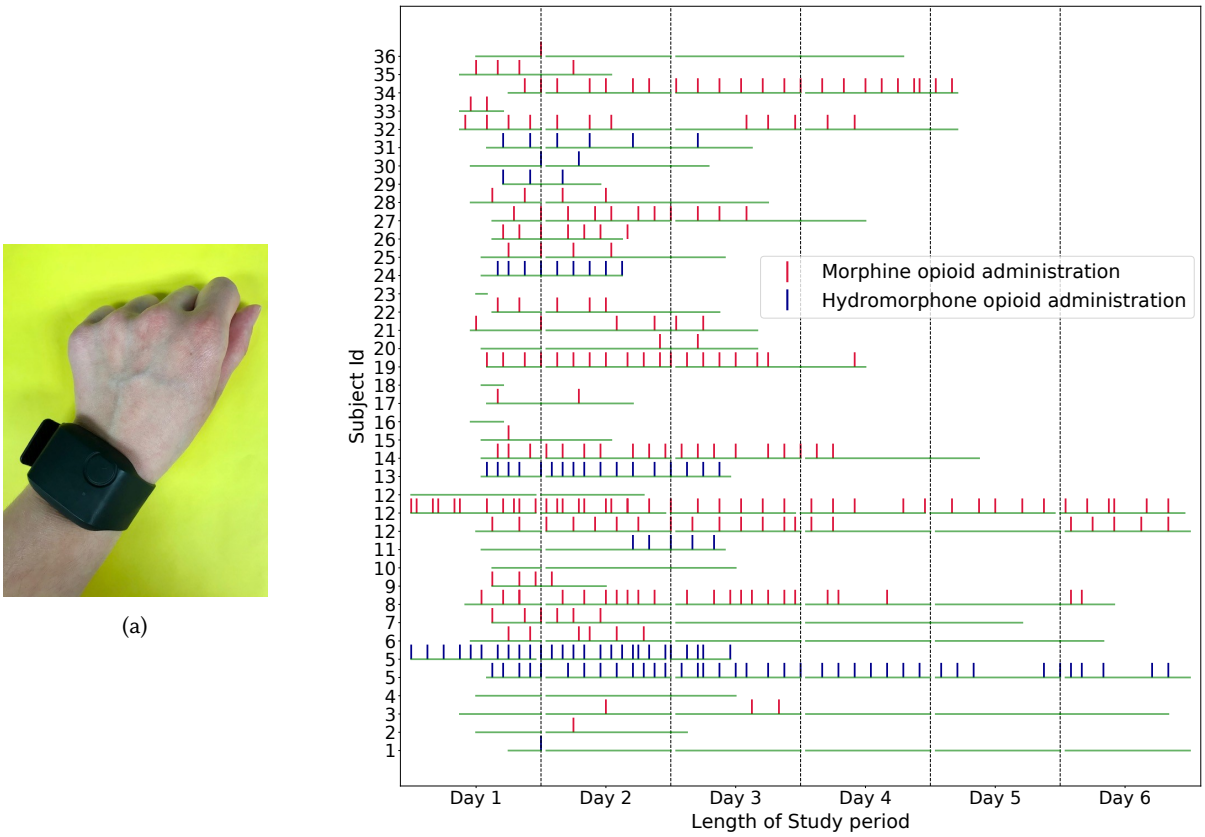
The Empatica E4 (Empatica, Milan, Italy) [31] is a commercially available, noninvasive, research-grade device worn on the wrist. It continuously measures and records various physiologic parameters using the following four sensors at pre-defined sampling frequencies: 1) a photoplethysmogram (PPG) sensor sampled at 64 Hz (instantaneous heart rate sampling frequency of 1 Hz); 2) an electrodermal activity (EDA) sensor sampled at 4 Hz; 3) a 3-axis accelerometer (ACC) sampled at 32 Hz; and, 4) an infrared (IR) thermophile sampled at 4 Hz. The PPG sensor measures Blood Volume Pulse (BVP), from which Emptaica E4 derives interbeat interval (IBI) using a standard tacogram, this subsequently is used to compute heart rate variability (HRV). The EDA sensor is used to capture electrical conductance across the skin and is characterized by the skin conductance level (EDA tonic) and skin conductance response (EDA phasic). A 3-axis accelerometer detects body motion, and an IR thermopile measures peripheral skin temperature. These raw signals captured from the E4 wristband have been validated in different research settings and are shown to have a performance comparable to the clinically used devices [55, 58, 78]. All data collected from the E4 was initially stored in the sensor's on-board integrated memory until connected to a computer via USB and transmitted to Empatica's HIPAA-compliant cloud-based server (Empatica Connect). All the data files were downloaded in comma-separated values (CSV) format for analysis.

3.3 Data Collection

Subjects were enrolled for the duration of their hospital stay, during which time they wore the E4 on their non-dominant wrist at all times. Daily in-person check-ins were completed by the research assistant who would exchange the sensor for a fully charged one, troubleshoot any technical issues, and download data collected in the previous 24 hours. Demographic and historical information (including medical history, current medications, opioid use history, and substance use history) were collected upon enrollment. Electronic medical records (EMRs) were queried throughout the study period to obtain information about the timing, dose, and route of all medication administrations, and subjects were instructed to press the E4's event marker button to "tag" and event whenever they received an opioid pain medication. Thus for every opioid administration, we created two opportunities to obtain ground truth data- one as the subject annotates the event and the second when the nurse documents the medication administration in the EMR.

Thirty-six subjects were enrolled over the course of three years. A total of 2070 hours of sensor data were collected, during which time 339 intravenous (IV) opioid administrations were captured. Five subjects were not administered any IV opioids during the study period despite having them ordered. The majority of administrations were morphine (68%), and the remaining were hydromorphone. The type of opioid administered was based upon clinician preference/judgment. The amount of opioid administered (dose) was converted to morphine milligram equivalents (MME), which is a standardized way to compare the relative potency of different types of opioids [88]. The timeline of different opioid administration events over the study period for all the participants is shown in figure1b.

Subjects' opioid use history was classified by a medical toxicologist as either *naive* (no provider-prescribed opioid use within the past six months and no lifetime history of opioid misuse), *chronic* (currently taking provider-prescribed opioids, ongoing opioid misuse, or less than five years of recovery from an opioid use disorder), or *occasional* (not meeting criteria for naive or chronic). All opioid administrations that occurred during the study were coded as either "confirmed" or "EMR-based." *Confirmed* administrations were defined by the subject annotation (tag on the sensor) within ten minutes of the EMR-documented time of administration. The sensor tag time was used as the opioid administration time. *EMR-based* administrations did not have an associated sensor tag from the subject, and the EMR-documented time was used as the opioid administration time. All study data collected were managed using the Research Electronic Data Capture (REDCap) [34] web application and entered into a biometric data repository.



(a) (b)

Fig. 1. (a) Picture of E4 sensor. (b) Plot showing the distribution and timeline of opioid administration events per subject (hourly-basis). If a subject was enrolled for more than 6 days (subjects 5 and 12), we used multiple rows. The Green horizontal bar indicates the start, end, and total duration of sensing data collected.

3.4 Preliminary Statistical Analysis of Changes in Wearable Signals Due to Opioid

Before attempting to model opioid use events with the wearable sensor data, we ran a preliminary analysis on the cleaned sensor data (explained in 4.1) to: 1) determine whether our wrist-worn sensor is sufficiently sensitive to register physiological changes associated with opioid administration; and 2) whether these changes in the sensor data are statistically significant. While the physiologic effects of opioid use have been previously described in the medical domain, the application of wearable sensors to identify and measure these changes has not been extensively studied. To this end, we performed a paired t-test between physiological signals by applying different statistical functions (listed in Table 2) on a 15-minute time window of a pre- and post-opioid administration. A list of time-domain and frequency-domain-based statistical functions for different signals, including accelerometry, electrodermal activity, skin temperature, heart rate, and interbeat interval, are considered for this statistical analysis. Throughout the paper, the accelerometry signal refers to the magnitude of net acceleration computed from the 3-axis accelerometer signal. The choice of statistical functions for this analysis was based on prior knowledge. Some functions were previously utilized in measuring physiologic effects associated with opioid use

Table 2. List of functions applied on physiological signals for preliminary statistical analysis.

Feature Type	Physiological Signal	Statistical function
Time-Domain	Accelerometry (ACC), Electrodermal Activity (EDA), Skin temperature (TEMP), Heart rate (HR)	Minimum Maximum Mean Median Standard Deviation Skewness Kurtosis Inter-quartile-range
Time-Domain	Interbeat interval (IBI)	MeanNN SDNN RMSSD SDSD NN50 pNN50
Frequency-Domain	ACC	Dominant frequency Spectral Entropy Spectral Energy Minimum Maximum Mean Standard Deviation
Frequency-Domain	IBI	VLF LF HF 'LF/HF ratio' LF (nu) HF (nu)'

[14], while others have been employed in other well-known problem domains (e.g., activity recognition, stress detection, etc.) [11, 18, 30, 56, 76]. All signals were standardized for this analysis.

We observed that several heart rate functions (mean, maximum) decreased in the post-opioid administration period, which is consistent with the medical literature [16]. The changes in skin temperature and accelerometer like increase in skin temperature max and drop-in motion after the opioid intake are consistent with recent work by Carreiro et al. [14]. In the context of opioid use detection, the heart rate variability (HRV) has not been extensively studied by prior work. Several of the statistical functions that we have extracted from IBI can capture HRV information. In our data, we found that the difference between pre- and post-opioid administration values of several time- and frequency-domain functions of IBI to be statistically significant. All the functions that are statistically significant from the paired t-test experiment are presented in table 10 of Appendix A.
Of note, several of the functions that showed significant differences had not been previously reported in the literature. Overall, these results validate that the physiological signals captured by the wristband sensor are sensitive to physiological changes due to opioid administration. While encouraging, this result does not directly guarantee a robust and high fidelity opioid use tracking system with the wearable sensor. Such a system must learn to recognize the opioid use-related physiological changes from the non-opioid events while being resilient to inter-individual differences. In the next section, we will present a temporal convolutional network with a channel-temporal attention mechanism that can combine the multimodal sensor data and uniquely recognize the opioid-related physiological changes to detect opioid administration events.

4 METHODOLOGY

In this section, we describe how the data collected from our study is pre-processed, how the features were extracted. Finally, we present our model architecture for detecting opioid use or administration.

4.1 Data Cleaning

The first step of the data cleaning process involved detecting and removing data recorded when the subject was loosely/not wearing the sensor. This step was accomplished by removing portions of data consisting of values that would be considered incompatible with human life (i.e., skin temperature at or below 20°C, zero EDA signal, and brief spikes (150-200 beats per minute) in heart rate). ≈ 6.5% of the sensor data was removed in this process.

The data still contained a low signal-to-noise ratio, so pre-processing of the data was required to remove outliers, noise artifacts, etc. We used a Butterworth low-pass filter of order 5 and 1 Hz cut-off frequency to filter EDA, accelerometry, heart rate, and skin temperature signals. For filtering noise in IBI, we used a threshold-based artifact correction algorithm [91]. We first computed a local average in this algorithm by taking a median filter on the IBI time-series. We looked at the difference between the IBI value and the local average value for each timestep. If the absolute difference was greater than a threshold (0.35 seconds), we marked it as an artifact. The identified artifacts were replaced by interpolated values using cubic spline interpolation.

4.2 Model Architecture

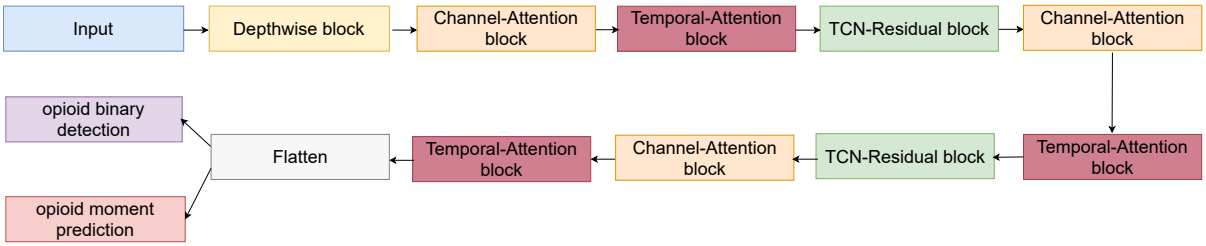


Fig. 2. Channel-Temporal Attention TCN (CTA-TCN) model architeture used for jointly predicting opioid administration and opioid moment prediction.

As mentioned previously, we collected multiple days of physiological data from each subject. To create the dataset for our OpiTrack system , we partitioned this continuous sensor data into time-windows of fixed size and considered a sliding window. The choice of time-window size and sliding-window size was a hyperparameter, and the optimal values for these are discussed in the next section. Given physiological signal information of multiple modalities from a time-window, our model detects if an opioid administration occurred in this time-window or not. If an administration is detected, it then predicts the exact moment (time-minute) of administration in the time-window.

Our model consists of four main components: 1) Depthwise block; 2) Residual blocks based on Temporal convolutional networks (TCN) architecture; 3) Channel-Attention block; and, 4) Temporal-Attention block. An overview of the model architecture is shown in figure 2 and descriptions of each component can be found below.

Depthwise Block: The input to the model is a time-windowed multimodal physiological signal. All these signals have been denoised as mentioned in section 4.1. In section 3.4, we discussed how different features from different modalities show different trends due to opioid administration (e.g., heart rate decreases due to opioid while skin temperature tends to increase). To capture these unique modality-specific trends, our model at first uses Depthwise convolution to extract features from each modality separately. No parameters is shared between the convolutions used for different modalities. An overview of the depthwise block architecture is shown in figure 4a. Since depthwise convolution extracts features from each modality separately and does not capture cross-modality interaction, we introduce TCN residual block that performs convolutions across modalities.

TCN-Residual Block: A relatively large time-sequences as input is required for opioid use modeling due to the inherent time delay for the opioid to impact physiology (i.e., opioid metabolism). It can take approximately 60-90 minutes for an IV opioid administration to have an effect on the human body [8] . Commonly used temporal models like Recurrent Neural Networks (RNNs), which include Long-Short Term Memory networks (LSTM) and Gated Recurrent Units (GRUs), may not be optimal for our application as they are known to suffer from exploding and vanishing gradients for long input time series[37]. We approached this problem by constructing our model based on Temporal convolutional networks (TCNs) [49, 67].

TCNs, a type of convolution network that convolve over time-domain, combine various aspects of RNNs and CNNs architecture while overcoming exploding and vanishing gradients by avoiding the backpropagation path defined by the input sequence. TCNs are also shown to outperform RNNs on different standard benchmark time-series datasets, and tasks [7]. There are two key characteristics of TCN. First, like RNNs, the model can take an input sequence of any length and produce the same length output sequence during any model stage. This is accomplished by using a 1D fully-convolutional network (FCN) and zero-padding input layers based on the kernel size. Second, while CNNs use standard convolutions, the convolutions used in TCNs are causal dilated. Causal convolutions allow no leakage of information from the future to the past while predicting for the present. Dilated convolutions in TCN will enable the model's receptive field (the size in history the model can look at while predicting the present) to be exponential of the network's depth. Figure 3 shows the differences between various convolutions. For a 1-D input sequence x $\in \mathbb{R}^n$ and filter f:{0,1,..k-1} $\rightarrow \mathbb{R}$, the result of causal dilated convolution on any element s of the output sequence can be defined as.

$$F(s) = \sum_{i=0}^{k-1} f(i).x_{s-d.i}$$

Here, d is the dilation factor, k is the filter size and the term $x_{s-d.i}$ reflects the causal convolution and the receptive field. To ensure we can handle larger input sequences and deeper networks with high dilation factors, residual connections are added to the TCN. These connections help to skip the TCN if it negatively impacted the model's performance. We observed a 3% drop in F1-score with the removal of residual connection. The TCN-Residual block architecture we used in our model is shown in figure 4b.

Channel-attention Block: The input to the model is time-series data captured by the E4 sensor. This data includes time-series information of heart rate, Skin temperature, EDA, and IBI. We view each of these physiological signals as a channel to the model. This input is passed through several convolution blocks, and the filters in each block generate subsequent feature maps that learn different information associated with the input. The feature maps generated are different from one another and will have varying levels of importance for predicting the target class. Therefore, providing the model a way to prioritize the feature maps/channels that are important can improve our system's performance.

In section 3.4, when we performed statistical analysis on various functions of physiological signals during an opioid administration, we observed different signals vary at different rates and magnitudes (Table 10 from

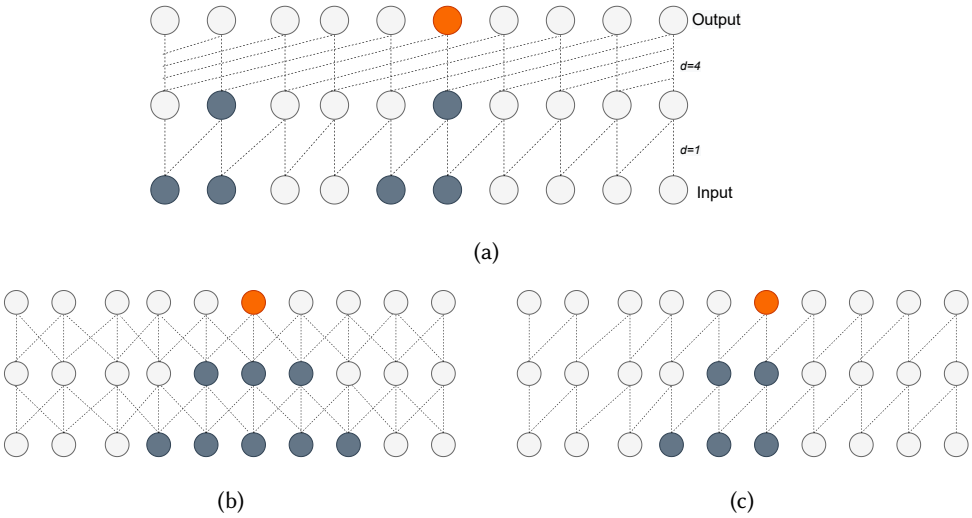


Fig. 3. (a) Causal dilated convolutions with dilation factors d=1,4 and filter size k=2. (b) Standard convolutions with filter size k=2. (c) Causal convolutions with filter size k=2. This figure also shows the input information (grey) used to predict a certain output time-step (orange). We can see how dilation increases the model's receptive field.

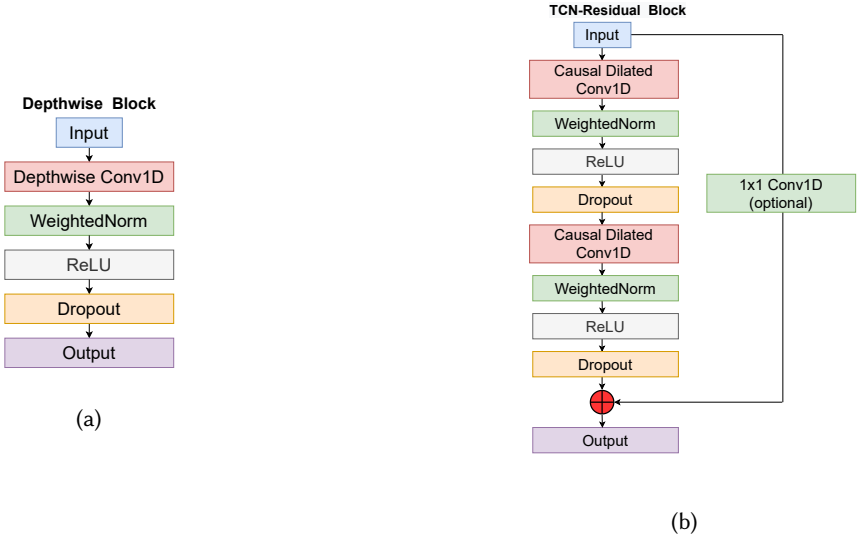


Fig. 4. (a) Depthwise block used in our model. Each channel (modality) is convolved independent of other channels. (b) TCN-Residual block consisting of dilated convolutions to ensure the input and output size stay the same while gradually increasing the receptive field of the model.

Appendix A). So, keeping channel attention in our architecture seemed a natural choice to consider. We addressed channel-attention using a squeeze-excitation (SE) module [39]. The main reasons for using SE are: 1) the number

of learnable parameters is very low; and 2) they have very little overhead, making them easily pluggable into any of the architectures. In general, SE networks are used with 2D data which include images and videos [17, 23, 51]. Here, we modified it slightly and used it on 1D time-series data. The architecture of our channel-attention is shown in figure 5a.

Channel-attention (SE) block consists of two main components-*Squeeze*, and *Excitation*. The Channel-attention's block input is a TxC vector where T is the length of the time-sequence, and C is the number of channels. This input is first passed through the *Squeeze* module. Ideally, for each channel to attend to other channels, it requires looking at all the information present in the channels, which would necessitate a total of TxC parameters. As we go deeper into the network and consider longer time-sequences, the number of parameters will be a large overhead for the model. These issues are addressed in the *Squeeze* module where the input of TxC is squeezed along the temporal-direction using a global average pooling function to generate a $1xC$ vector. For a input $x \in \mathbb{R}^{TxC}$, the squeeze operation along a channel c would be

$$S_c = \frac{\sum_{i=0}^{T-1} x_{i,c}}{T}$$

where $x_{i,c}$ is the input along the cth channel. This vector is then passed through the *Excitation* module. In this module, we learn the non-linear interaction between channels, which is achieved by first compressing the $1xC$ vector into $1x\frac{C}{r}$ and then expanding it back to $1xC$ vector using two fully-connected (FC) layers. r is the reduction factor and is a hyperparameter which controls the capacity and computational cost of the SE block. We use a value of 2 for r in our model. The *Excitation* operation can be defined by the equation

$$E = \sigma(W_2\delta(W_1S))$$

where S is the output after the *Squeeze* module, $W_1 \in \mathbb{R}^{\frac{C}{r}xC}$ and $W_2 \in \mathbb{R}^{Cx\frac{C}{r}}$ are the weight matrices of the FC layers, δ is the non-linear operation which in our case is ReLU[62] function and σ is the sigmoid operation. The output from the *Excitation* module is passed through a sigmoid activation to keep all the channel scores in the 0-1 range. Finally, the output after sigmoid activation is broadcasted back to the original input using an element-wise multiplication.

Temporal-attention Block: Since the physiological changes due to opioid use are temporally localized (as discussed in section 3.4 and take place over a specific time period after administration, we use self-attention based mechanism to attend to the values from the informative time regions within a window. The self-attention block in our architecture is based on several recent works [20, 94] and is shown in figure 5b. The Temporal-attention block takes the input X processed by Channel-attention block. The first step of attention includes generating *Query*, *Key*, and *Value* from X, where *Query*=W_QX, *Key*=W_KX, and *Value*=W_VX. $W_Q \in \mathbb{R}^{Cx\frac{C}{5}}$, $W_K \in \mathbb{R}^{Cx\frac{C}{5}}$ and $W_V \in \mathbb{R}^{CxC}$ are the corresponding weight matrices. We then multiply *Query* and *Key* to generate a compatibility score between different temporal moments. The compatibility scores are normalized by a softmax function to have a total sum of one. In the next step, we multiply normalized compatibility scores with *Value* to generate a temporal attention-weighted vector A. Finally, we add back the input vector X to A to generate output O. γ is a learnable parameter,

$$O = \gamma A + X$$

.

4.3 Loss Function

Given a time-window consisting of physiological signals from a wearable wrist-borne sensor, our model detects if an opioid has been administered in this time-window or not, and the moment of administration, if it occurred.

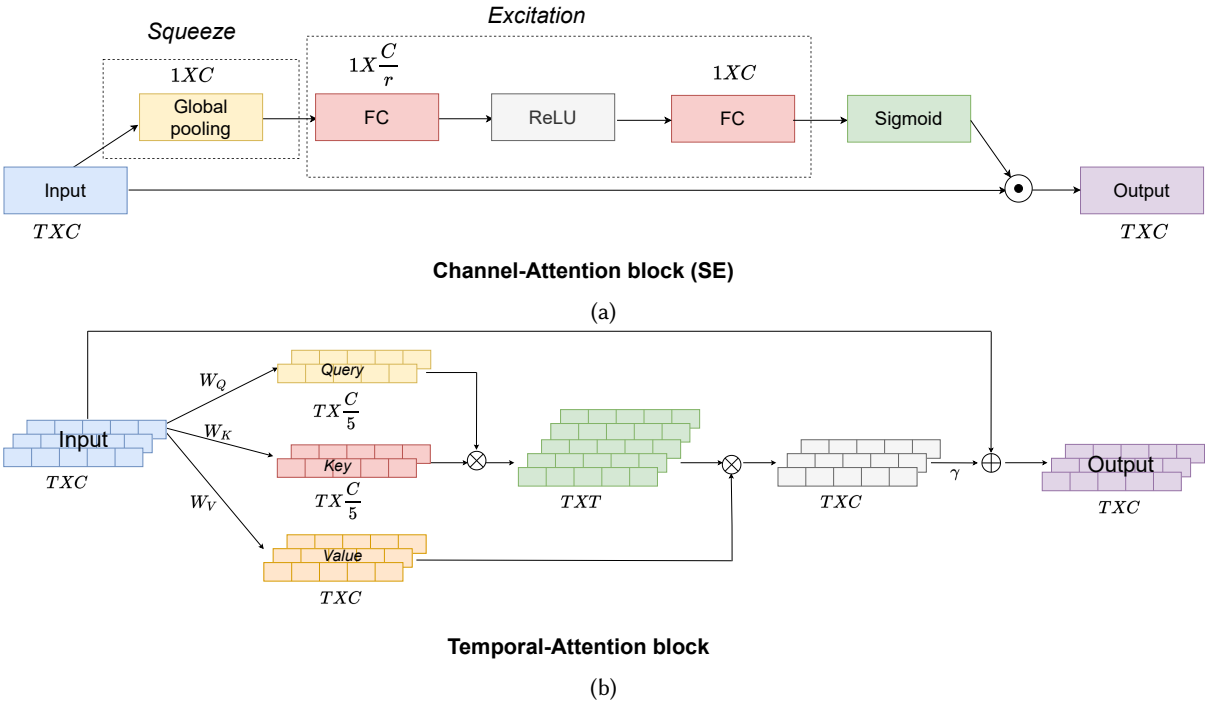


Fig. 5. (a) Channel-attention (SE) block of our model. ⊙ represents element-wise matrix multiplication. (b) Temporal-attention block used in our model. ⊗ represents a matrix product and ⊕ requires matrix sum.

To jointly solve both of these tasks, we use a hybrid loss function. The equation below describes the hybrid loss function used.

$$L = \lambda L_{WCE} + (1 - \lambda) L_{KAPPA} \tag{1}$$

L_{WCE} is the standard binary-weighted cross-entropy loss between actual administration class and predicted administration. For every time-window, the actual administration value is 0/1 depending on whether an opioid administration occurred in it or not. Predicted administration is the output of the 'opioid binary detection' block of figure 2. To account for the class imbalance, we used weighted cross-entropy loss (WCE) instead of a standard cross-entropy.

L_{KAPPA} is the weighted kappa index [19] used as a loss function. While L_{WCE} term helps detect an opioid use event in a time-window, L_{KAPPA} is used for detecting the exact moment of administration in the time-window. The actual administration moment is specified as a one-hot encoding vector with a size same as the time-window length. A '1' value indicates the exact time of administration in the window and '0' values represent all other minutes in the time-window. The predicted moment from the "opioid moment prediction" block of figure 2 has the same size as the actual administration moment vector while containing softmax probabilities for the moment of administration. Weighted kappa as loss function is explained in detail in the Appendix B. For time windows with no opioid administration, the weighted kappa loss term will be absent. λ is a hyperparameter to control the importance given to both the loss terms. We use a value of 0.4 in our model.

5 RESULTS

Using the model architecture described in section 4.2, we model opioid administration, which includes detecting opioid use and predicting the exact moment of administration. Along with the Channel-Temporal attention TCN (CTA-TCN) model, we explored models of varying degrees of complexity. These include logistic regression, bidirectional LSTM (BiLSTM) [32], CNN-LSTM [24], and LSTM with Fully Convolutional Network (LSTM-FCN) [45]. In the CNN-LSTM model, CNN is used to extract the features from physiological signals, and LSTM uses these features to make a time-series prediction. LSTM-FCN consists of two blocks: an LSTM and an FCN block. The outputs from these two blocks are merged, and prediction is made based on that.

Along with using denoised physiological signals as input to the deep learning models and letting them do the feature extraction, we also explored the usage of hand-crafted features from physiological signals, which included extracting time- and frequency-domain features (PSTAT) discussed in section 3.4. We also considered the features Mahmud et al. [53] described in the only known prior similar work and used the same model considered there. These approaches are compared against a baseline model whose predictions are not based on physiological signals and require only subject information and time of the day. All the results reported are with a Leave-One-Subject-Out Cross-Validation (LOSOXV) experiment.

To account for the class imbalance, a weighted F1-score is used to measure the performance of the binary classification of opioid administration at the window level. In addition, we report specificity, sensitivity, and AUC scores. For predicting the opioid administration moment or time, the performance is measured in terms of mean-absolute error (MAE) and R^2 coefficient between actual and predicted administration time points. We also estimated normalized mean-absolute error (NMAE) by normalizing the MAE with respect to the window length. Since the MAE depends on the window length, NMAE allows us to get a measure that is invariant to the choice of window length.

Table 3. Performance of opioid administration detection model trained with different feature subsets and models from Leave-One-Subject-Out (LOSO) crossvalidation experiments. Performance was measured in terms of F1-score (weighted), specificity, sensitivity, and Area Under Curve (AUC). Channel-Temporal Attention TCN outperformed all the models.

Features	**Model**	**F1-score**	**specificity**	**sensitivity**	**AUC**
Demographical+ Time of the day	Baseline-Logistic	*0.44 ± 0.26*	*0.50 ± 0.46*	*0.50 ± 0.45*	*0.48 ± 0.07*
[53] [1]	Decision-Tree	*0.57 ±0.14*	*0.70 ± 0.12*	*0.41 ±0.21*	*0.51 ±0.17*
Physiological Statistical (*PSTAT*)	Logistic	*0.64 ± 0.13*	*0.65± 0.14*	*0.48 ± 0.25*	*0.55± 0.17*
Physiological Statistical (*PSTAT*)	BiLSTM	*0.70 ±0.1*	*0.71 ± 0.2*	*0.57 ±0.3*	*0.7 ±0.14*
Input signal	BiLSTM	*0.66 ±0.1*	*0.63 ±0.15*	*0.63 ±0.20*	*0.62 ±0.09*
Input signal	TCN	*0.73 ±0.11*	*0.72 ± 0.14*	*0.74 ±0.18*	*0.74 ±0.11*
Input signal	**CTA-TCN**	***0.80 ±0.1***	***0.77 ± 0.14***	***0.80 ±0.17***	***0.77 ±0.1***
Input signal	CNN-LSTM	*0.72±0.11*	*0.65±0.17*	*0.82 ± 0.12*	*0.76±0.12*
Input signal	LSTM-FCN	*0.70±0.08*	*0.71±0.16*	*0.69 ± 0.14*	*0.72±0.15*

A 100-minute long sliding time-window with window shift of 30 minutes was used in our model as it provided optimal performance with respect to opioid use detection sensitivity and exact moment prediction NMAE. We observed that using a smaller time-window (than 80 mins) resulted in low sensitivity for opioid use event detection as most of the inputs do not contain any opioid administration, and the windowed sensor time series fails to capture the full physiological signal dynamics due to opioid. Similarly, a very large time-window leads to an

[1] This work considered Frequency-domain features of Accelerometer and Time-domain features of Accelerometer, EDA, and skin temperature.

Table 4. Performance of opioid moment prediction model trained with different feature subsets and models from Leave-One-Subject-Out (LOSO) crossvalidation experiments. Performance was measured in terms of Mean Absolute Error (MAE) and R^2 coefficient. Channel-Temporal Attention TCN (CTA-TCN) outperformed all the models.

Features	**Model**	**MAE (mins)**	**NMAE (%)**	**R^2**
Demographical + Time of the day	Baseline-Logistic	*18.5 ± 2.11*	*18.5 ± 2.11*	*0.37*
[53]	Decision-Tree	*22.7 ±3.2*	*22.7 ±3.2*	*0.22*
Physiological Statistical (*PSTAT*)	Logistic	*18.1*±4.41	*18.1*±4.41	*0.34*
Physiological Statistical (*PSTAT*)	BiLSTM	*17.8 ±5.8*	*17.8 ±5.8*	*0.31*
Input signal	BiLSTM	*23.1 ± 6.1*	*23.1 ± 6.1*	*-0.36*
Input signal	TCN	*10.1 ±2.8*	*10.1 ±2.8*	*0.84*
Input signal	**CTA-TCN**	***8.6 ±2.4***	***8.6 ±2.4***	***0.85***
Input signal	CNN-LSTM	*25.7±6*	*25.7±6*	*-0.54*
Input signal	FCN-LSTM	*14.4±5.6*	*14.4±5.6*	*0.62*

increase of mean absolute error for the opioid moment prediction since the model now has to solve a significantly more challenging task of picking an accurate opioid use moment from more candidate moments available in a larger time-window. Figure 6 shows how the model's performance varies across different window lengths.

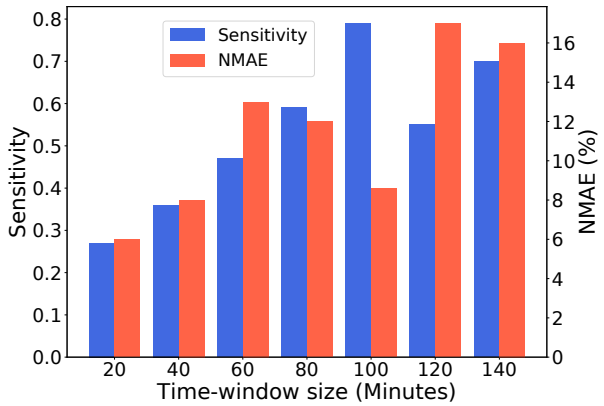


Fig. 6. Barplot showing the performance of our model with different input time-window sizes. We can see time-window size of 100 is the optimal in terms of High Senstivity and Low NMAE score.

All signals have been downsampled to have five samples per minute and the input data is standardized for each modality. The exact specifications of our model, which includes the dilation factors, number of channels for each layer, kernel size, and hidden layer sizes, are provided in Appendix D. Table 3 and 4 shows the performance of the models in opioid administration detection and opioid moment prediction. All the models that used physiological signals outperformed the baseline model, which used no sensor information, by a significant margin. The Channel-Temporal attention TCN (CTA-TCN) outperformed all other models in both tasks.

For opioid administration detection after LOSOXV, CTA-TCN attains an overall mean F1-score of 0.80, specificity of 0.77, sensitivity of 0.80, and AUC score of 0.77. For opioid moment prediction, it attained a R^2 coefficient of 0.85 and a mean MAE score of 8.6 (mins). As we considered a 100-minute input, the normalized mean absolute error (NMAE) is 8.6%. Our next best model after CTA-TCN in terms of all performance metrics is a TCN model

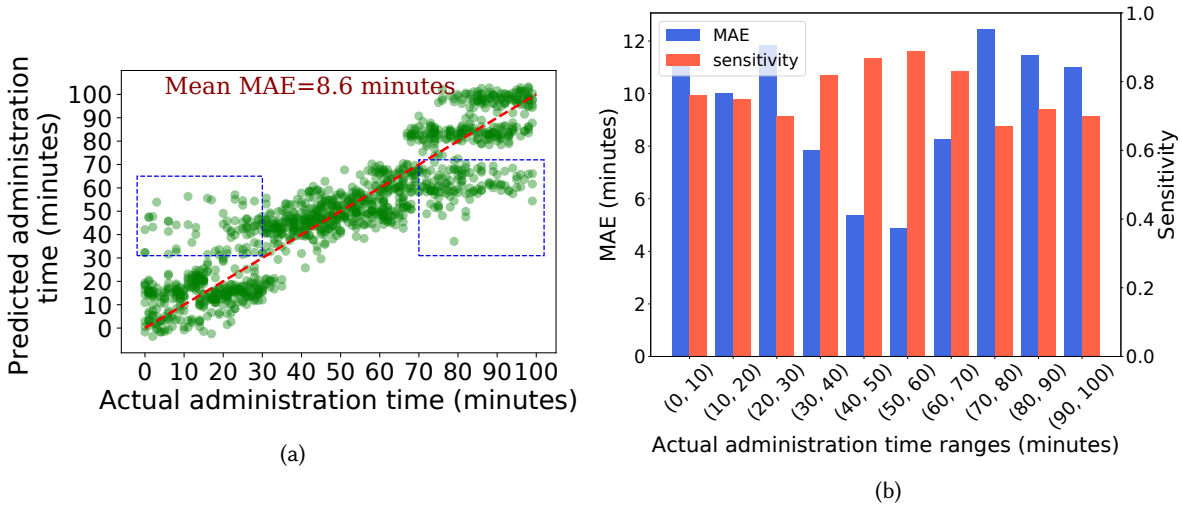


Fig. 7. (a) Scatter plot of actual administration moment and predicted administration moment by the CTA-TCN model. Small Gaussian noise is added to the predicted times to enhance visualization. (b) Barplot showing the model's MAE and sensitivity performance across different time ranges of ground truth administration moment.

similar to CTA-TCN but with no channel and temporal-attention blocks. Attention blocks increased the F1-score by ≈9.5% and decreased the NMAE by 1.5% over TCN with no attention blocks.

Figure 7a shows a scatter plot between the actual administration moment and predicted administration moment of our CTA-TCN model. If our model made perfect predictions, all the points would lie along the diagonal red-dashed line. While the predicted administration moment generally follows the actual administration moment, we see that predictions are far off from the ground-truth (points inside the blue-dashed rectangle) when the administration occurs during the start/end of the time-window. This observation motivated us to make a barplot showing MAE of opioid moment prediction task and sensitivity of opioid detection task for every 10-minute time range shown in figure 7b. In this figure, we can see that the lowest MAE and highest sensitivity are achieved when an opioid administration occurs in the middle of the time-window. This result illustrates that the model works most accurately when the windowed sensor data adequately captures information from both pre- and post-phase of an opioid administration. Consequently, the characteristic physiological changes of opioid use can be observed by the model.

5.1 Ablation Study for CTA-TCN Architecture

In this section, we validate the choices we made while designing our model architecture. We do this using an ablation-based study where we remove/modify certain blocks of the architecture while the rest is untouched to see how it impacted the model's performance. The following are the set of experiments carried out:

- Replacing depthwise convolutions with standard convolutions. This experiment aims to see if using convolution filters specific to a modality provides any benefit over convolution filters shared by all the modalities.
- Removing the TCN architecture by replacing all causal dilated convolutions with standard convolutions- This experiment aims to see how the model would perform when its receptive field is decreased from exponential, when using causal dilated convolutions, to linear, when using standard convolutions.

- Removing channel-attention blocks. This experiment aims to validate how weighing channels based on their importance for target prediction impacts the performance of our model.
- Removing temporal-attention blocks. This experiment aims to validate how our model's performance is impacted by attending over the temporal direction.

Table 5. The change in CTA-TCN performance after replacing/removing certrain blocks of the architecture while keeping rest of model as-is.

From	To	Change in F1-score	Change in NMAE score
Depthwise convolutions	Standard-convolutions	-12.6%	+4.1%
Causal-dilated convolutions	Standard-convolutions	-10.15%	+2.2%
Channel-attention	Removed	-13%	+3.1%
Temporal-attention	Removed	-5.5%	+1.7%

The results of these experiments are shown in table 5. Removing the Channel-attention/ Depthwise convolutions negatively impacted the performance of our model the most. This explains the importance of letting the model extract features specific to each modality in the network's initial layer and helping the model attend to the feature maps based on their importance rather than treating them equally. Removing temporal attention caused the least change among all the experiments. We believe this is due to how efficient the TCN-Residual blocks are already in handing the temporal information. The change in the opioid moment prediction performance was not significant compared to opioid detection for all the experiments.

5.2 Performance Breakdown Across Different Demographic/Opioid-type Subgroups

As part of the study protocol, demographic and historical information was collected for each subject. Variables of particular interest include gender, age, body mass index (BMI), and opioid use history. The demographic factors and their distribution are listed in table 1.

In this section, we discuss the performance of our model when stratified by demographic variables of interest. Our goal here was to understand if our model performance is uniform across different sub-groups or skewed towards a particular population group. If the performance is skewed, what might have caused this?

Peformance Based on Gender: Our CTA-TCN model's performance based on gender is shown in table 6. While the *model generally performed well in detecting opioid use across both genders, the performance in males (as reflected in the F1-score) was marginally better.* However, we interpret these results with caution, as this difference may be attributable to females having more opioid administrations (62%) than males but a smaller sized dataset (43%). As a result, specificity was much lower in females than males. Our model achieved equal sensitivity and predicted the opioid moment with similar NMAE in both gender groups. Previous works have cited no gender differences in opioid metabolism for morphine or hydromorphone [85, 86].

Table 6. Performance of our CTA-TCN model based on gender

Gender	NMAE	F1-score	Specificity	Sensitivity
Female	8.63	0.75	0.70	0.77
Male	8.67	0.82	0.79	0.78

Peformance Based on BMI Categories: Our CTA-TCN model's performance on various BMI categories is shown in table 7. We can see the *model performed opioid detection/opioid moment prediction uniformly across all BMI categories*

from the table. While there are previous works discussing the effect of opioid in altering the subject's BMI over time [26, 59, 87], there is less work on the opposite relationship (the effect of BMI on opioid pharmacokinetics and pharmacodynamics). Obesity and resultant health problems create complex interactions with opioid metabolism that are expected to create differences in obese individuals compared to those with normal BMI [52]; some studies indicate relatively similar pharmacokinetic parameters despite obesity [21], but obesity had been shown to be an independent risk factor for opioid-induced respiratory arrest [41]. Our data suggest that in this cohort, the accuracy of detection was not changed by BMI.

Table 7. Performance of our CTA-TCN model across different BMI categories.

BMI category	NMAE	F1-score	Specificity	Sensitivity
Normal	9.31	0.79	0.76	0.78
Overweight	8.07	0.80	0.77	0.80
Obese	7.73	0.81	0.79	0.79

Peformance Based on Opioid Use History: Repeated opioid exposure over time leads to opioid tolerance, or the requirement of higher doses of the drug to achieve the same clinical effect. Because of this well-established phenomenon, the impact of subjects' opioid use history is of particular clinical interest when considering our ability to detect opioid use. Our CTA-TCN model's performance on various population groups based on their opioid use history is shown in table 8. *The model performed better in opioid-naive subjects and those with occasional use (those without a significant history of opioid use) compared to those with a history of chronic use, which is expected.*

Table 8. Performance of our CTA-TCN model based on opioid use history.

Opioid history	NMAE	F1-score	Specificity	Sensitivity
Naive	7.97	0.82	0.79	0.79
Occasional	9.23	0.82	0.78	0.78
Chronic	9.02	0.76	0.72	0.78

Peformance Based on Type of Opioid Administered: We also considered the impact of opioid type on model performance. The two types of opioids captured in our dataset (morphine and hydromorphone) are both commonly prescribed full opioid agonists, although hydromorphone is approximately eight times more potent than morphine. Our CTA-TCN model's performance on the two types of opioids considered in this study is shown in table 9. *Performance was similar between the two categories indicating that, despite the differences in potency, the physiologic effects are similar enough for the opioid type not to imact the model's performance.*

Table 9. Performance of our CTA-TCN model across based on opioid type.

Opioid Type	NMAE	F1-score	Specificity	Sensitivity
Morphine	8.70	0.785	0.77	0.77
Hydromorphone	8.66	0.79	0.77	0.79

5.3 Feature Relevance and Explainability

In order to evaluate the relevance of each modality and to identify which one contains the most information for opioid use detection, we drop a certain modality from full set multimodal input signal and train the CTA-TCN model with the remaining modalities. The decrease in F1-score for opioid administration detection and increase in NMAE for opioid moment prediction reflects the importance of that modality.

As shown in figure 8, the heart rate and interbeat interval (IBI) signals are the two most important modalities for opioid administration detection. Recent studies have documented the effect of opioids on the cardiovascular system [9, 16] which point towards the informativeness of cardiac-related physiological biomarkers. While both HR and IBI will capture heart rate information, IBI can extract complementary information, including heart rate variability. Both EDA and ACC yield a moderate level of F1-score drop, which shows that they contain some complementary information. For opioid moment prediction, HR, EDA, and ACC contribute the most.

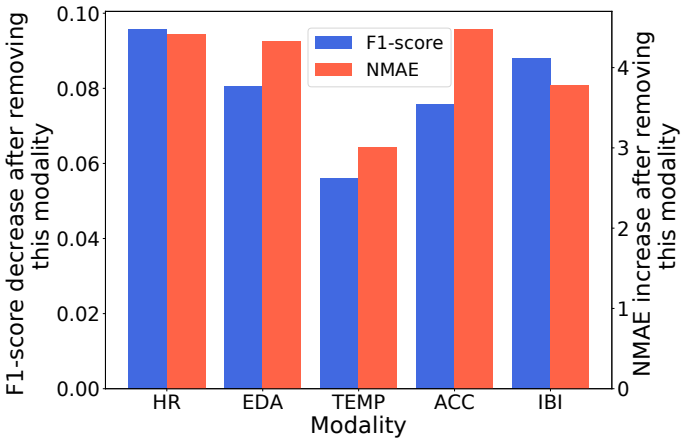
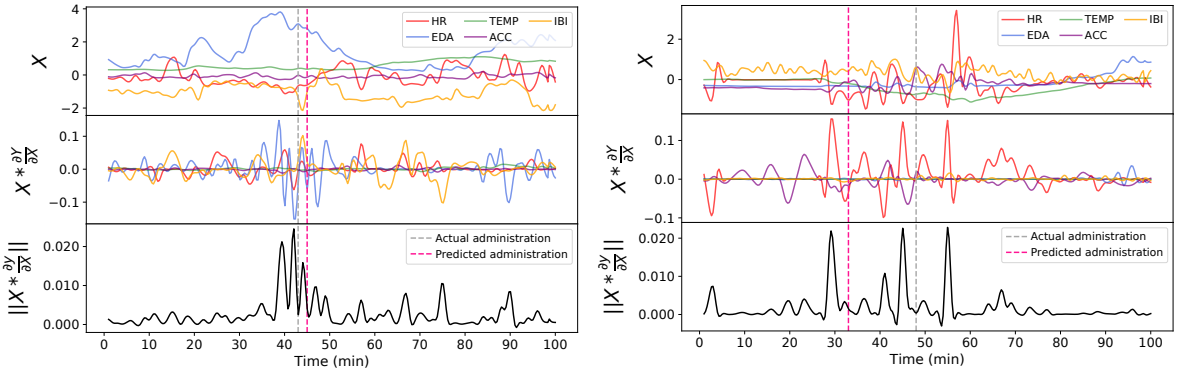


Fig. 8. Barplot showing the importance of different modalities. The F1-score decrease and NMAE increase due to the removal of a modality shows its importance for opioid detection/moment prediction.

Skin-temperature was the least important of all the modalities for both opioid detection and moment prediction. We hypothesize that this is due to the several competing ways in which opioids affect skin temperature; for example, opioids are known to cause hypothemia [101], but are also assocaited with variable degrees of histamine release, causing peripheral vasodialtion and thus a temporary increase in skin temperature. The degree to which an individual subejct experiences either of these effects is idosyncratic, making temperature changes inconsistnet and less reliable for predicition.

While the feature relevance analysis was able to show the overall importance of different modalities, now we want to investigate how the proposed model placed importance on different modalities and time instances for individual predictions. In recent times, explainable AI has received significant attention and it has been used to interpret the prediction/output of deep neural networks (DNNs) for a certain instance/data [2, 79, 102]. Among different explainable AI techniques, attribution-based methods are computationally less expensive, and model agnostic, which means if the model is differentiable from end-to-end, they can easily be added to the architecture, which makes them very popular. Here we apply a backpropagation based attribution method called *Gradient*input* [84] in which we compute the importance of the input features by taking the partial derivatives of output with respect to the input and multiplying this by the input itself. The gradient $\frac{\partial Y}{\partial X}$ gives us an estimate of each input feature's impact on the prediction class, and the intuition for multiplying this to the input can be



(a) (b)

Fig. 9. Example plot showing the output from *Gradient*input* to demonstrate importance of modality and temporal information for prediction. The topmost plot shows the input X considered for this experiment, the middle plot shows the *Gradient*input* for this input, and in the bottom plot, we take the magnitude of all the features of *Gradient*input* along the temporal direction. For the sake of visualization, we used unstandardized data to plot X.

better understood if we assume the model to be linear. While the derivative in a linear model gives us only the coefficient, multiplying this with the input gives us the total contribution of each input's features.
Figure 9a and 9b illustrate two example 100-min long sensor time series with opioid administration event and shows which part of the time series for different modalities was important for the opioid detection model as it successfully detected the opioid events. For this experiment, we consider the output Y to be from the opioid detection model. Both examples had an opioid administration event within the 100-minute time window. For each example, we made three plots. The topmost plot shows the input X which is denoised time-series data from different modalities. The middle plot shows the output of $X * \frac{\partial Y}{\partial X}$, and the bottom plot shows sum of the magnitude of all the features of $X * \frac{\partial Y}{\partial X}$ along the channel direction. While the middle plot gives information about which modalities are important and when they are important. The bottom plot shows the aggregated importance across modalities and tells which time-periods were critical across all modalities.

In figure 9a, the information from EDA, HR, and IBI contributed the most for detecting opioid administration. On the other hand, the second example shown in figure 9b shows that the information within the HR and ACC modalities were more critical for the model. This highlights that the proposed temporal convolutional networks with attention mechanism could dynamically distribute its attention to the richer modalities containing more critical information for opioid use detection. By jointly looking at the topmost and middle subplots of both examples, we observe that any time a modality is showing substantial change or activities, our model is putting more importance on it. In particular, for figure 9a, near the administration time, changes in the EDA signal are the most dominant, and our model attended the highest to this signal. Similarly, in figure 9b, the model attended HR modality the highest at times when its changes were the most dominant. From both these examples, we can also confirm that our model does not attend to a particular modality that shows little or no changes. This shows that our channel-attention was able to only attend to the modalities whose changes were significant for prediction. It should also be noted that the TEMP signal is the least important in both the examples and this aligns with the previous observation we made in section 5.3 where we observed the TEMP signal to contain the least information for opioid detection/moment prediction among all the modalities. Looking at the bottom plot of both the examples, we can see that model considered information from only a few time-points to be necessary

while paying less attention to the rest of the input. This shows that our temporal attention block was able to attend to only the key time-points in the input. While it is not possible to completely understand how predictions were made in a black-box style architecture like ours, this experiment shows and gives us a brief understanding of how the model had made a certain prediction.

6 DISCUSSION

In this section, we discuss the implications of our work across different application scenarios, limitations, and future directions.

6.1 Implications for Clinical Monitoring of Opioid Use

The current gold standards for opioid detection (patient self-report and/or urine drug screen) are fraught with limitations, resulting in subpar data to make clinical decisions. The ability to objectively detect opioid use with an 0.8 F1-score in a passive and nonobtrusive way presents an unprecedented opportunity for healthcare providers. It would improve care for patients who use opioids therapeutically (to treat pain) and those who misuse them.

In the context of therapeutic opioid use, where opioids are typically prescribed to be taken as needed for pain, data from a detection system can allow a clinician to see how much opioid medication a patient is taking and how often. Such data can easily be transformed into visualizations for clinicians to identify escalating use patterns that put patients at risk for overdose and addiction, and alternative treatments can be initiated to prevent/mitigate the development of opioid tolerance or dependance. In patients who use opioids chronically, concern for misuse and/or diversion (i.e., the practice of selling or giving away opioids) often arises. In this scenario, confirmation of opioid use as prescribed improves the therapeutic relationship and builds trust; conversely, confirmation that a patient is not taking opioids as prescribed can prompt further query and reduce the diversion of prescription opioids for unintended use.

In patients who have opioid use disorder and are in treatment, such a system can be useful to detect relapse (return to drug use) or can be further developed to identify not only opioid **use** but opioid **toxicity** (overdose) as a trigger for intervention. Once mature, this technology can also be used to monitor adherence to medications used to treat opioid use disorder, such as buprenorphine and methadone. These medications (which are opioids themselves) have the potential to save lives and reduce overdose deaths, but only if taken as prescribed.

6.2 Generalizability in Non-clinical Settings

As part of the inclusion criteria, all participants enrolled in this study were receiving opioids to treat pain and not take them for recreational use. Individuals who use opioids chronically (whether for medical or non-medical purposes) commonly experience pain in the form of two phenomena- opioid-induced hyperalgesia and opioid withdrawal. Opioid-induced hyperalgesia is characterized by a paradoxical (increased) sensitization to painful stimuli due to prolonged opioid exposure [54, 99]. Opioid withdrawal is a group of physiologic symptoms people experience when they develop opioid dependence and then stop using opioids. These symptoms range in severity and include nausea, vomiting, and severe bone or joint pain [82]. In fact, many people who abuse opioids describe a phenomenon, whereas, at some point, the drive to use opioids changes from a desire to achieve euphoria ("to get high") to a desire to prevent the pain of withdrawal [71]. Given this cycle of pain and pain relief, our findings are expected to mimic real-world opioid use scenarios from a physiologic standpoint.

6.3 Limitations and Future Directions

Our work has several limitations that impact generalizability and highlight the need for ongoing research on this topic. First, our analysis considered intravenous opioids taken for prescription purposes only. However, orally taken opioids are commonly used in this population as well. Because of differences in absorption and onset/duration of action, these types of ingestion events will need to be evaluated in a separate study. Second, so

that we could ensure the accuracy of our ground truth, we used a sample of patients who were in the hospital and had opioid administrations documented in an EMR (in addition to documentation via self-report). This resulted in a more controlled environment than a natural setting (i.e., outpatient) and may somewhat limit the generalizability of our findings.

When selecting this population for this study, one of our primary considerations was collecting a reliable ground-truth dataset. In general, individuals using opioids for non-medical purposes in the natural environment would not have been an optimal population to study in this respect due to lack of directly observed administrations, poor follow-up rates for individuals with this disease, and fear of reporting illicit use due to legal implications. We intend to expand our work into this population in the future but started with a well-labeled data set as an initial step to demonstrate the feasibility of detection.

Finally, future research needs to hone the accuracy of opioid detection, potentially by combining additional sensor modalities, user characteristics, and situational factors into the model. We explored the impact of several user characteristics in this dataset (e.g., BMI, opioid use history, etc.), but to understand the contribution of these (and other important factors such as age, race, and ethnicity), larger datasets are needed. An additional important area for exploration will be the quantification of opioid effect via sensor-based data (i.e., differentiating an analgesic dose of an opioid from an overdose). This capability would transform this application into an immediate, life-saving tool. Finally, research is needed on how to translate wearable-based opioid detection into actionable insight for healthcare providers who are already overwhelmed with data. Data visualization strategies, EMR integration, and just-in-time intervention strategies are required to fully leverage the potential of this technology.

7 CONCLUSION

This work supports that we can detect the use of opioids and the time they are administered using physiological signals from a wearable sensor. While prior work examined the physiologic changes pre-and post-administration, this work goes a step further and identifies opioid use in the context of a larger set of data using a machine learning model. It also lays the groundwork for an unobtrusive, objective system to identify opioid use for clinicians that can enhance the safety of opioid use in the context of pain management by preventing overdose and misuse.

ACKNOWLEDGMENTS

We sincerely thank the reviewers for their insightful comments and suggestions that helped improve this paper. This work is supported by the National Institutes of Health/National Institute on Drug Abuse under the grant K23DA045242 (PI: Carreiro), and start-up grant support from the College of Information and Computer Sciences and the Institute for Applied Life Sciences at the University of Massachusetts Amherst.

REFERENCES

[1] Amin Ahsan Ali, Syed Monowar Hossain, Karen Hovsepian, Md Mahbubur Rahman, Kurt Plarre, and Santosh Kumar. 2012. mPuff: automated detection of cigarette smoking puffs from respiration measurements. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. 269–280.

[2] Marco Ancona, Enea Ceolini, Cengiz Öztireli, and Markus Gross. 2017. Towards better understanding of gradient-based attribution methods for deep neural networks. *arXiv preprint arXiv:1711.06104* (2017).

[3] Anshika Arora, Pinaki Chakraborty, and MPS Bhatia. 2020. Analysis of Data from Wearable Sensors for Sleep Quality Estimation and Prediction Using Deep Learning. *Arabian Journal for Science and Engineering* 45, 12 (2020), 10793–10812.

[4] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. 2015. Physical human activity recognition using wearable sensors. *Sensors* 15, 12 (2015), 31314–31338.

[5] Kavita M Babu, Jeffrey Brent, and David N Juurlink. 2019. Prevention of opioid overdose. *New England Journal of Medicine* 380, 23 (2019), 2246–2255.

[6] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[7] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).

[8] Randall Clint Baselt and Robert H Cravey. 1982. *Disposition of toxic drugs and chemicals in man*. Vol. 8. Biomedical publications Davis, CA.

[9] Mina Behzadi, Siyavash Joukar, and Ahmad Beik. 2018. Opioids and cardiac arrhythmia: a literature review. *Medical principles and practice* 27, 5 (2018), 401–414.

[10] Alexander J Boe, Lori L McGee Koch, Megan K O'Brien, Nicholas Shawen, John A Rogers, Richard L Lieber, Kathryn J Reid, Phyllis C Zee, and Arun Jayaraman. 2019. Automating sleep stage classification using wireless, wearable sensors. *NPJ digital medicine* 2, 1 (2019), 1–9.

[11] Sansanee Boonnithi and Sukanya Phongsuphap. 2011. Comparison of heart rate variability measures for mental stress detection. In *2011 Computing in Cardiology*. IEEE, 85–88.

[12] Alan S Campbell, Jayoung Kim, and Joseph Wang. 2018. Wearable electrochemical alcohol biosensors. *Current opinion in electrochemistry* 10 (2018), 126–135.

[13] Stephanie Carreiro, Hua Fang, Jianying Zhang, Kelley Wittbold, Shicheng Weng, Rachel Mullins, David Smelson, and Edward W Boyer. 2015. iMStrong: deployment of a biosensor system to detect cocaine use. *Journal of medical systems* 39, 12 (2015), 1–8.

[14] Stephanie Carreiro, Kelley Wittbold, Premananda Indic, Hua Fang, Jianying Zhang, and Edward W Boyer. 2016. Wearable biosensors to detect physiologic change during opioid use. *Journal of medical toxicology* 12, 3 (2016), 255–262.

[15] Soujanya Chatterjee, Alexander Moreno, Steven Lloyd Lizotte, Sayma Akther, Emre Ertin, Christopher P Fagundes, Cho Lam, James M Rehg, Neng Wan, David W Wetter, et al. 2020. SmokingOpp: Detecting the Smoking'Opportunity'Context Using Mobile Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–26.

[16] Alexander Chen and Michael A Ashburn. 2015. Cardiac effects of opioid therapy. *Pain Medicine* 16, suppl_1 (2015), S27–S31.

[17] Xi Cheng, Xiang Li, Jian Yang, and Ying Tai. 2018. SESR: Single image super resolution with recursive squeeze and excitation networks. In *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 147–152.

[18] Lucio Ciabattoni, Francesco Ferracuti, Sauro Longhi, Lucia Pepa, Luca Romeo, and Federica Verdini. 2017. Real-time mental stress detection based on smartwatch. In *2017 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 110–111.

[19] Jacob Cohen. 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin* 70, 4 (1968), 213.

[20] Rui Dai, Luca Minciullo, Lorenzo Garattoni, Gianpiero Francesca, and François Bremond. 2019. Self-Attention Temporal Convolutional Network for Long-Term Daily Living Activity Detection. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 1–7.

[21] Sjoerd de Hoogd, Pyry AJ Välitalo, Albert Dahan, Simone van Kralingen, Michael MW Coughtrie, Eric PA van Dongen, Bert van Ramshorst, and Catherijne AJ Knibbe. 2017. Influence of morbid obesity on the pharmacokinetics of morphine, morphine-3-glucuronide, and morphine-6-glucuronide. *Clinical pharmacokinetics* 56, 12 (2017), 1577–1587.

[22] Jordi de La Torre, Domenec Puig, and Aida Valls. 2018. Weighted kappa loss function for multi-class classification of ordinal data in deep learning. *Pattern Recognition Letters* 105 (2018), 144–154.

[23] Dandan Ding, Junchao Tong, and Lingyi Kong. 2020. A deep learning approach for quality enhancement of surveillance video. *Journal of Intelligent Transportation Systems* 24, 3 (2020), 304–314.

[24] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. 2015. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2625–2634.

[25] Robert H Eich, Robert Gilbert, and J Howard Auchincloss Jr. 1957. The acute effects of smoking on the mechanics of respiration in chronic obstructive pulmonary emphysema. *American Review of Tuberculosis and Pulmonary Diseases* 76, 1 (1957), 22–32.

[26] Jennifer M Fenn, Jennifer S Laurent, and Stacey C Sigmon. 2015. Increases in body mass index following initiation of methadone treatment. *Journal of substance abuse treatment* 51 (2015), 59–63.

[27] Scott M Fishman, Barth Wilsey, Jane Yang, Gary M Reisfield, Tara B Bandman, and David Borsook. 2000. Adherence monitoring and drug surveillance in chronic opioid therapy. *Journal of pain and symptom management* 20, 4 (2000), 293–307.

[28] Ana Catarina Fonseca and José M Ferro. 2013. Drug abuse and stroke. *Current neurology and neuroscience reports* 13, 2 (2013), 1–9.

[29] S Fu. 2016. Adulterants in urine drug testing. *Advances in clinical chemistry* 76 (2016), 123–163.

[30] Lei Gao, AK Bourke, and John Nelson. 2014. Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical engineering & physics* 36, 6 (2014), 779–785.

[31] Maurizio Garbarino, Matteo Lai, Dan Bender, Rosalind W Picard, and Simone Tognetti. 2014. Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*.

IEEE, 39–42.
[32] Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks* 18, 5-6 (2005), 602–610.
[33] Bhanu Teja Gullapalli, Annamalai Natarajan, Gustavo A Angarita, Robert T Malison, Deepak Ganesan, and Tauhidur Rahman. 2019. 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 (2019), 1–31.
[34] Paul A Harris, Robert Taylor, Brenda L Minor, Veida Elliott, Michelle Fernandez, Lindsay O'Neal, Laura McLeod, Giovanni Delacqua, Francesco Delacqua, Jacqueline Kirby, et al. 2019. The REDCap consortium: Building an international community of software platform partners. *Journal of biomedical informatics* 95 (2019), 103208.
[35] Cheng He, Yun-jin Yao, and Xue-song Ye. 2017. An emotion recognition system based on physiological signals obtained by wearable sensors. In *Wearable sensors and robots*. Springer, 15–25.
[36] Howard A Heit and Douglas L Gourlay. 2004. Urine drug testing in pain medicine. *Journal of pain and symptom management* 27, 3 (2004), 260–267.
[37] Sepp Hochreiter. 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6, 02 (1998), 107–116.
[38] Syed Monowar Hossain, Amin Ahsan Ali, Md Mahbubur Rahman, Emre Ertine David Epstein, Ashley Kennedy, Kenzie Preston, Annie Umbricht, Yixin Chen, and Santosh Kumar. 2014. Identifying drug (cocaine) intake events from acute physiological response in the presence of free-living physical activity. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. IEEE, 71–82.
[39] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7132–7141.
[40] Masudul H Imtiaz, Volkan Y Senyurek, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov. 2019. Objective detection of cigarette smoking from physiological sensor signals. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 3563–3566.
[41] Igor Izrailtyan, Jiejing Qiu, Frank J Overdyk, Mary Erslon, and Tong J Gan. 2018. Risk factors for cardiopulmonary and respiratory arrest in medical and surgical hospital patients on opioid analgesics and sedatives. *PloS one* 13, 3 (2018), e0194553.
[42] Margaret Jarvis, Jessica Williams, Matthew Hurford, Dawn Lindsay, Piper Lincoln, Leila Giles, Peter Luongo, and Taleen Safarian. 2017. Appropriate use of drug testing in clinical addiction medicine. *Journal of addiction medicine* 11, 3 (2017), 163–173.
[43] Wenchao Jiang and Zhaozheng Yin. 2015. Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia*. 1307–1310.
[44] Priyanka Kakria, NK Tripathi, and Peerapong Kitipawang. 2015. A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. *International journal of telemedicine and applications* 2015 (2015).
[45] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen. 2017. LSTM fully convolutional networks for time series classification. *IEEE access* 6 (2017), 1662–1669.
[46] Nathaniel Katz and Gilbert J Fanciullo. 2002. Role of urine toxicology testing in the management of chronic opioid therapy. *The Clinical journal of pain* 18, 4 (2002), S76–S82.
[47] Heli Koskimäki, Henna Mönttinen, Pekka Siirtola, Hanna-Leena Huttunen, Raija Halonen, and Juha Röning. 2017. Early detection of migraine attacks based on wearable sensors: experiences of data collection using Empatica E4. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 506–511.
[48] Yongjin Kwon, Kyuchang Kang, and Changseok Bae. 2014. Unsupervised learning for human activity recognition using smartphone sensors. *Expert Systems with Applications* 41, 14 (2014), 6067–6074.
[49] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. 2017. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 156–165.
[50] Liang Li, Qian Yu, Baoteng Xu, Qifan Bai, Yunpeng Zhang, Huijun Zhang, Chengjie Mao, Chunfeng Liu, Tianyu Shen, and Shouyan Wang. 2017. Multi-sensor wearable devices for movement monitoring in Parkinson's disease. In *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 70–73.
[51] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 254–269.
[52] C Lloret-Linares, A Lopes, X Declèves, A Serrie, S Mouly, J-F Bergmann, and S Perrot. 2013. Challenges in the optimisation of post-operative pain management with opioids in obese patients: a literature review. *Obesity surgery* 23, 9 (2013), 1458–1475.
[53] Md Shaad Mahmud, Hua Fang, Honggang Wang, Stephanie Carreiro, and Edward Boyer. 2018. Automatic Detection of Opioid Intake Using Wearable Biosensor. In *2018 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 784–788.
[54] M Marion Lee, M Sanford Silverman, M Hans Hansen, M Vikram Patel, and MD Laxmaiah Manchikanti. 2011. A comprehensive review of opioid-induced hyperalgesia. *Pain physician* 14 (2011), 145–161.

[55] Cameron McCarthy, Nikhilesh Pradhan, Calum Redpath, and Andy Adler. 2016. Validation of the Empatica E4 wristband. In *2016 IEEE EMBS international student conference (ISC)*. IEEE, 1–4.

[56] Sakorn Mekruksavanich, Narit Hnoohom, and Anuchit Jitpattanakul. 2018. Smartwatch-based sitting detection with human activity recognition for office workers syndrome. In *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*. IEEE, 160–164.

[57] Dawn Milner. 2004. The physiological effects of smoking on the respiratory system. *Nursing times* 100, 24 (2004), 56–59.

[58] Nir Milstein and Ilanit Gordon. 2020. Validating measures of electrodermal activity and heart rate variability derived from the empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience* 14 (2020).

[59] Farzaneh Montazerifar, Mansour Karajibani, and Kobra Lashkaripour. 2012. Effect of methadone maintenance therapy on anthropometric indices in opioid dependent patients. *International journal of high risk behaviors & addiction* 1, 3 (2012), 100.

[60] Jessica L Moreno, Matthew S Duprey, Bryan D Hayes, Kirsten Brooks, Sabrina Khalil, Sarah E Wakeman, Russell J Roberts, Jared S Jacobson, and John W Devlin. 2019. Agreement between self-reported psychoactive substance use and urine toxicology results for adults with opioid use disorder admitted to hospital. *Toxicology communications* 3, 1 (2019), 94–101.

[61] Nitin Nair, Chinchu Thomas, and Dinesh Babu Jayagopi. 2018. Human activity recognition using temporal convolutional network. In *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction*. 1–8.

[62] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *Icml*.

[63] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E Sunshine. 2019. Opioid overdose detection using smartphones. *Science translational medicine* 11, 474 (2019).

[64] Annamalai Natarajan, Gustavo Angarita, Edward Gaiser, Robert Malison, Deepak Ganesan, and Benjamin M Marlin. 2016. Domain adaptation methods for improving lab-to-field generalization of cocaine detection using wearable ECG. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 875–885.

[65] Annamalai Natarajan, Abhinav Parate, Edward Gaiser, Gustavo Angarita, Robert Malison, Benjamin Marlin, and Deepak Ganesan. 2013. Detecting cocaine use with wearable electrocardiogram sensors. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 123–132.

[66] Lewis S. Nelson and Dean Olsen. 2019. *Opioids*. McGraw-Hill Education, New York, NY. accesspharmacy.mhmedical.com/content.aspx?aid=1163010439

[67] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499* (2016).

[68] Qi Wei Oung, M Hariharan, Hoi Leong Lee, Shafriza Nisha Basah, Mohamed Sarillee, and Chia Hau Lee. 2015. Wearable multimodal sensors for evaluation of patients with Parkinson disease. In *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*. IEEE, 269–274.

[69] Abhinav Parate, Meng-Chieh Chiu, Chaniel Chadowitz, Deepak Ganesan, and Evangelos Kalogerakis. 2014. Risq: Recognizing smoking gestures with inertial sensors on a wristband. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. 149–161.

[70] Gavril W Pasternak. 1993. Pharmacological mechanisms of opioid analgesics. *Clinical neuropharmacology* 16, 1 (1993), 1–18.

[71] Joseph V Pergolizzi Jr, Robert B Raffa, and Melanie H Rosenblatt. 2020. Opioid withdrawal symptoms, a consequence of chronic opioid use and opioid use disorder: Current understanding and approaches to management. *Journal of clinical pharmacy and therapeutics* 45, 5 (2020), 892–903.

[72] Martin Ragot, Nicolas Martin, Sonia Em, Nico Pallamin, and Jean-Marc Diverrez. 2017. Emotion recognition using physiological signals: laboratory vs. wearable sensors. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 15–22.

[73] Bethany R Raiff, Çağdaş Karataş, Erin A McClure, Dario Pompili, and Theodore A Walls. 2014. Laboratory validation of inertial body sensors to detect cigarette smoking arm movements. *Electronics* 3, 1 (2014), 87–110.

[74] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. 2019. Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–30.

[75] Pritam Sarkar and Ali Etemad. 2020. Self-supervised learning for ecg-based emotion recognition. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3217–3221.

[76] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM international conference on multimodal interaction*. 400–408.

[77] Johannes Schneider, Marc Schroth, Jörg Ottenbacher, and Wilhelm Stork. 2018. A novel wearable sensor device for continuous monitoring of cardiac activity during sleep. In *2018 IEEE Sensors Applications Symposium (SAS)*. IEEE, 1–6.

[78] Angela AT Schuurmans, Peter de Looff, Karin S Nijhof, Catarina Rosada, Ron HJ Scholte, Arne Popma, and Roy Otten. 2020. Validity of the Empatica E4 wristband to measure heart rate variability (HRV) parameters: A comparison to electrocardiography (ECG). *Journal of medical systems* 44, 11 (2020), 1–11.

[79] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[80] Volkan Y Senyurek, Masudul H Imtiaz, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov. 2019. A comparison of SVM and CNN-LSTM based approach for detecting smoke inhalations from respiratory signal. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 3262–3265.

[81] Volkan Y Senyurek, Masudul H Imtiaz, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov. 2020. A CNN-LSTM neural network for recognition of puffing in smoking episodes using wearable sensors. *Biomedical Engineering Letters* 10, 2 (2020), 195–203.

[82] Mansi Shah and Martin R Huecker. 2020. Opioid withdrawal. In *StatPearls [Internet]*. StatPearls Publishing.

[83] Muhammad Shoaib, Hans Scholten, Paul JM Havinga, and Ozlem Durmaz Incel. 2016. A hierarchical lazy smoking detection algorithm using smartwatch sensors. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 1–6.

[84] Avanti Shrikumar, Peyton Greenside, Anna Shcherbina, and Anshul Kundaje. 2016. Not just a black box: Learning important features through propagating activation differences. *arXiv preprint arXiv:1605.01713* (2016).

[85] Howard S Smith. 2009. Opioid metabolism. In *Mayo Clinic Proceedings*, Vol. 84. Elsevier, 613–624.

[86] Offie P Soldin and Donald R Mattison. 2009. Sex differences in pharmacokinetics and pharmacodynamics. *Clinical pharmacokinetics* 48, 3 (2009), 143–157.

[87] Andrew Stokes, Kaitlyn M Berry, Jason M Collins, Chia-Wen Hsiao, Jason R Waggoner, Stephen S Johnston, Eric M Ammann, Robin F Scamuffa, Sonia Lee, Dielle J Lundberg, et al. 2019. The contribution of obesity to prescription opioid use in the United States. *Pain* 160, 10 (2019), 2255.

[88] Alexander B Stone, Richard D Urman, Alan D Kaye, and Michael C Grant. 2018. Labeling morphine milligram equivalents on opioid packaging: a potential patient safety intervention. *Current pain and headache reports* 22, 7 (2018), 1–4.

[89] Qu Tang, Damon J Vidrine, Eric Crowder, and Stephen S Intille. 2014. Automated detection of puffing and smoking with wrist accelerometers. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*. 80–87.

[90] Anong Tantisuwat and Premtip Thaveeratitham. 2014. Effects of smoking on chest expansion, lung function, and respiratory muscle strength of youths. *Journal of physical therapy science* 26, 2 (2014), 167–170.

[91] Mika P Tarvainen, Juha-Pekka Niskanen, Jukka A Lipponen, Perttu O Ranta-Aho, and Pasi A Karjalainen. 2014. Kubios HRV–heart rate variability analysis software. *Computer methods and programs in biomedicine* 113, 1 (2014), 210–220.

[92] Arijit Ukil and Soma Bandyopadhyay. 2019. Automated cardiac health screening using smartphone and wearable sensors through anomaly analytics. In *Mobile Solutions and Their Usefulness in Everyday Life*. Springer, 145–172.

[93] Ramachandran Varatharajan, Gunasekaran Manogaran, Malarvizhi Kumar Priyan, and Revathi Sundarasekar. 2018. Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm. *Cluster Computing* 21, 1 (2018), 681–690.

[94] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[95] Wanpen Vongpatanasin, Yasser Mansour, Bahman Chavoshan, Debbie Arbique, and Ronald G Victor. 1999. Cocaine stimulates the human cardiovascular system via a central mechanism of action. *Circulation* 100, 5 (1999), 497–502.

[96] Nana Wilson, Mbabazi Kariisa, Puja Seth, Herschel Smith IV, and Nicole L Davis. 2020. Drug and opioid-involved overdose deaths—United States, 2017–2018. *Morbidity and Mortality Weekly Report* 69, 11 (2020), 290–297.

[97] Danny Wyatt, Matthai Philipose, and Tanzeem Choudhury. 2005. Unsupervised activity recognition using automatically mined common sense. In *AAAI*, Vol. 5. 21–27.

[98] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*. PMLR, 2048–2057.

[99] David Z Yang, Billy Sin, Joshua Beckhusen, Dawei Xia, Rebecca Khaimova, and Ilia Iliev. 2019. Opioid-induced hyperalgesia in the nonsurgical setting: a systematic review. *American journal of therapeutics* 26, 3 (2019), e397–e405.

[100] Marcello Zanghieri, Simone Benatti, Alessio Burrello, Victor Kartsch, Francesco Conti, and Luca Benini. 2019. Robust real-time embedded EMG recognition framework using temporal convolutional networks on a multicore IoT processor. *IEEE transactions on biomedical circuits and systems* 14, 2 (2019), 244–256.

[101] MR Zarrindast, A Vahedy, MR Heidari, and M Ghazi Khansari. 1994. On the mechanism (s) of morphine-induced hypothermia. *Journal of Psychopharmacology* 8, 4 (1994), 222–226.

[102] Matthew D Zeiler and Rob Fergus. 2014. Visualizing and understanding convolutional networks. In *European conference on computer vision*. Springer, 818–833.

A STATISTICAL ANALYSIS RESULTS

Table 10 shows the results of paired t-test experiments performed on statistical functions extracted from physiological signals.

Table 10. List of functions extracted from physiological signals that showed statistical significance in paired t-test analysis.

Modality	Statistical function	F value
Heart-rate	Maximum	2.72
	Mean	2.56
	Standard deviation	2.4
Skin-Temperature	Maximum	-3.04
	Standard deviation	-3.9
	Inter-quartile-range	-3.15
Electrodermal activity	Minimum	3.26
	Skewness	-4.9
	Kurtosis	-5.50
Accelerometer	Minimum	3.4
	Mean	3
	Mean frequency	-3.5
	Standard deviation frequency	-3.06
Interbeat interval	nn50	2.17
	pnn50	2.17
	sdnn	2.18
	VLF	-2.15
	LF	-2.66
	LF (nu)	-2.5

- **nn50**: The number of successive NN interval that differ by more than 50ms in the time-window (pre/post).
- **pnn50**: The percentage of successive NN interval that differ by more than 50ms in the time-window. Obtained by dividing nn50 by total number of NN intervals.
- **rmssd**: root mean square of successive differences between heartbeats in the time-window.
- **sdnn**: Standard deviation of NN intervals obtained from the time-window
- **sdsd**: Standard deviation of differences between consecutive NN intervals. All the NN intervals from the time-window are considered.
- **VLF**: Logarithmic of absolute power of very low frequency band (0-0.04 hz)
- **LF**: Logarithmic of absolute power of low frequency band (0.04-0.15 hz)
- **HF**: Logarithmic of absolute power of high frequency band (0.15-0.40 hz)
- **LF/HF**: LF power to HF power ratio
- **LF(nu)**: Normalized absoluted power of low frequency band given by equation $\frac{LF}{lF+HF+VLF}$
- **HF(nu)**: Normalized absoluted power of High frequency band given by equation $\frac{HF}{lF+HF+VLF}$

B WEIGHTED KAPPA AS LOSS FUNCTION:

Traditional loss functions for classifications such as cross-entropy loss, hinge loss, etc., weigh the disagreement between different classes equally. But in our case, where we predict the moment of administration in a time-window as a multi-class classification, we want to weigh the disagreements differently, based on how far they

are from the ground truth. For example, if the administration occurred at the tth minute in the time-window, the classifier which predicted administration at $t + 5^{th}$ minute should be preferred over the one which predicted administration at $t + 20^{th}$ minute. While both classifiers mislabeled the input, the former classifier is closer to ground truth compared to the latter.

We use weighted kappa statistic as our loss function to address this issue, which incorporates ratio-scaled degrees of disagreement (or agreement) between classes. Weighted kappa statistic has recently been used as loss functions in deep learning and general problems where the classes can be ordered using some intrinsic information [22]. In our case, it is the time of administration in the time-window.

Weighted kappa statistic is represented using three matrices, a matrix of observed scores O, a matrix of expected scores E, and weight matrix W. It is defined using equation

$$\kappa = 1 - \frac{\sum_{i,j} W_{i,j} P_{O_{i,j}}}{\sum_{i,j} W_{i,j} P_{E_{i,j}}} \tag{2}$$

where
K= Number of classes
i,j ∈ 1,2,...K
N= Total number of samples
N_i= Number of samples belonging to i^{th} class
$P_j(X_k)$= predicted probability that sample X_k belongs to j^{th} class.

W: Weight matrix to show the degree of disaggrement between classes where $w_{i,j} = \frac{(i-j)^n}{(K-1)^n}$, n=1 representns linear penalization, n=2 quadractic, and so on. For our experiments n=2 worked the best.
P_O: Observation matrix where $P_{O_{i,j}}$ is the observed proportion of samples classified as i^{th} class while their ground truth is j^{th} category.
P_E: Expected matrix obtained by doing an outer product between normalized ground truth count matrix (G) and expected count matrix obtained by chance (C). $P_{E_{i,j}} = G_i C_j$, here $G_i = \frac{N_i}{N}$, $C_j = \sum_{k=1}^{N} P_j(X_k)$
The weighted kappa statistic $\kappa \in$ [-1,1], with κ=1 indicating a strong agreement between predictions and ground truth, -1 indicating a strong disagreement and 0 indicating a random predictions. Therefore, we use $\log(1 - \kappa)$ as our loss function for opioid moment prediction.

C CTA-TCN ARCHITECTURE SETUP

In this section, we explain the architectural details of our CTA-TCN model. This includes the choice of all the hyperparameters, input/output sizes, etc. As mentioned previously, the CTA-TCN models take physiological information of Heart rate, Accelerometer, EDA, Skin temperature, and Interbeat interval from a time-window of size 100 minutes as input. All the signals are downsampled to a frequency of 5 samples per minute. Therefore the input to our model is $500x5$. We standardize the input along each modality before passing it to the CTA-TCN model. We used a dropout value of 0.2 and a leaky ReLU with a negative slope of 0.1. We optimized the loss function using Adam optimizer with a weight decay of 0.001. We used a high learning rate (lr) of 0.01 for the initial 25 epochs and used lr=0.0001 later on. We used a batch size of 32 and trained our model for 300 epochs by using an early stopping approach. If we do not see improvements over validation loss for 20 continuous epochs, we stop training our model.

D CTA-TCN PERFORMANCE WITHOUT WEIGHTED KAPPA LOSS

In this section we show the peformance of opioid moment prediction by using different type of loss functions instead of weighted kappa.

Table 11. CTA-TCN model architecture details

Architectural block	Input size	Output size	Additional hyper-parameters
Depthwise	500x5	500x10	Stride (s)=1, Dilation factor (d)=1, Padding (p)=4, Kernel size(k)=5
Channel-Attention1	500x10	500x10	reduction factor(r)=2
Temporal-Attention1	500x10	500x10	-
TCN-Residual1	500x10	500x20	s=1, d=2, p=8, k=5
Channel-Attention2	500x20	500x20	r=2
Temporal-Attention2	500x20	500x20	-
TCN-Residual2	500x20	500x30	s=1, d=8, p=35, k=5
Channel-Attention3	500x30	500x30	r=2
Temporal-Attention3	500x30	500x30	-
Flatten	500x30	15000x1	-
Opioid moment prediction	15000x1	100x1	-
Opioid binary detection	15000x1	1x1	-

Table 12. CTA-TCN model performance with different choices of loss function for opioid moment prediction

Loss function	NMAE (%)	R^2
Weighted Kappa	8.6±2.4	0.85
RMSE	14.8±2.9	0.72
Multi-class cross entropy	18.2± 3.1	0.63