Detecting Cocaine Use with Wearable Electrocardiogram Sensors

Annamalai Natarajan[1]   Abhinav Parate[1]   Edward Gaiser[2]   Gustavo Angarita[2]

Robert Malison[2]   Benjamin Marlin[1]   Deepak Ganesan[1]

[1]School of Computer Science, University of Massachusetts, Amherst

{anataraj, aparate, marlin, dganesan}@cs.umass.edu

[2]Department of Psychiatry, Yale School of Medicine, New Haven

{edward.gaiser, gustavo.angarita, robert.malison}@yale.edu

ABSTRACT

Ubiquitous physiological sensing has the potential to profoundly improve our understanding of human behavior, leading to more targeted treatments for a variety of disorders. The long term goal of this work is development of novel computational tools to support the study of addiction in the context of cocaine use. The current paper takes the first step in this important direction by posing a simple, but crucial question: Can cocaine use be reliably detected using wearable electrocardiogram (ECG) sensors? The main contributions in this paper include the presentation of a novel clinical study of cocaine use, the development of a computational pipeline for inferring morphological features from noisy ECG waveforms, and the evaluation of feature sets for cocaine use detection. Our results show that 32mg/70kg doses of cocaine can be detected with the area under the receiver operating characteristic curve levels above 0.9 both within and between-subjects.

Author Keywords

Drug Use, On-body Sensors, Classification

ACM Classification Keywords

I.2 ARTIFICIAL INTELLIGENCE : [Medicine and science]

General Terms

Algorithms; Experimentation; Measurement; Performance

INTRODUCTION

In recent years, the ability to continuously monitor the activities, health, and lifestyle of individuals using sensor technologies has reached unprecedented levels. Wearable "on-body" sensors now enable routine and continuous monitoring of a host of physiological signals including heart rate, blood pressure, respiratory rate, blood glucose and more. Commercial sensors manufactured by companies including Zephyr [23] are now commonly available to the everyday consumer for personal fitness and health applications.

Such ubiquitous physiological sensing has the potential to profoundly improve our understanding of human behavior, leading to more targeted treatments for a variety of disorders. This work focuses on the development of novel computational tools to support the study of addiction in the context of cocaine use. Current clinical studies of addictive behavior around cocaine use are limited to collecting data through subjective, retrospective self-reports [5]. In addition to being vulnerable to recall bias and intentional misrepresentation, this methodology may fail to capture important aspects of addiction that an individual patient is not consciously aware of including context-dependent triggers.

In contrast, continuous sensing technologies provide a new window into patients' daily lives - one through which an individual's behavior might be more completely, accurately, and dynamically viewed as it evolves in real-time. We posit that such an approach has the potential to yield new insights on unique aspects of an individual's addictive behavior as well as the extent to which specific environmental factors (people, places, and things) contribute to relapse, leading to highly individualized and dynamic markers of treatment response. While such an approach offers considerable promise, rigorous, well-controlled studies are required to establish the clinical utility of the technology in complex use cases like understanding cocaine addiction. The current work takes the first step in this important direction by posing a simple, but crucial question: *Can cocaine use be reliably detected using wearable on-body sensors?*

To begin to answer this question, it is important to understand the effect of cocaine on physiology, and the state-of-art in uncovering specific physiological features that could be indicative of cocaine use. By virtue of its peripheral actions on the sympathetic nervous system, it is well known that cocaine produces robust changes in primary indices of cardiovascular function (e.g., increases in heart rate) [19]. However, overall heart rate is also known to be easily confounded by a variety of other real-world behaviors such as exercise or stress. The topic of how cocaine affects morphological features of ECG waveforms is thus of great interest and the literature on the effects of cocaine provides support for a number of such features [9, 14, 15, 21, 19].

Building on this body of prior work from the cocaine literature, we select ECG as the most promising sensing modality

for the detection of cocaine use. In this work, we evaluate the ability of machine learning methods to detect cocaine use based on two fundamentally different representational frameworks. The first framework is a fully data-driven approach where the underlying detector has access to complete ECG waveforms. The key benefits of such a framework are that it eliminates the need for feature extraction from ECG waveforms, which is non-trivial both because of the noisy nature of the data, as well as waveform variations across subjects. The second framework we consider is knowledge-based in that it leverages features that have been explored in prior studies of cocaine's effect on the heart. We study six different features of ECG waveforms that have been shown to correlate with cocaine use in the literature. These features (described in detail in the following sections) include the RR interval, QT interval, QTc interval, PR interval, QRS interval, and T wave amplitude.

This paper makes the following primary contributions to wearable ECG-based sensing and cocaine use detection:
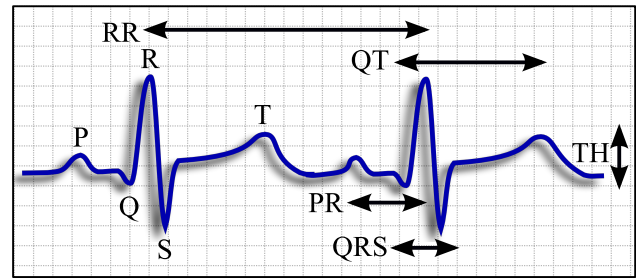
- The presentation of a novel clinical study of cocaine use conducted at the Yale Center for Clinical Investigation's Hospital Research Unit (YCCI-HRU) in which a unique set of wireless ECG data were collected from experienced cocaine users.
- The development of a computational pipeline for inferring morphological features from noisy ECG waveforms collected using a wearable ECG sensor (the Zephyr BioHarness 3).
- The evaluation of ECG feature sets from the knowledge-based and data-driven detection frameworks using both within-subject and between-subject protocols.

The clinical study involved six subjects, each of whom participated in a multi-week protocol involving multiple experimental conditions. The subjects wore the Zephyr BioHarness 3 chest band for extended time periods including baseline sessions (following several days of habituation) and cocaine use sessions with multiple dosage levels. The resulting ECG data is quite noisy and also subject to signal dropout. Standard algorithms for morphological analysis of intensive care unit (ICU) quality ECG data were found to be inadequate when applied to wireless ECG data, motivating the development of a customized feature extraction pipeline that includes significant filtering and smoothing. Our results show that wearable ECG-based cocaine use detection can be reliably performed both within and between subjects. We achieve average AUC rates over 0.9 at the 32mg/70Kg dosage level and over 0.8 at the 8 and 16mg/70Kg dosage levels using a combination of features.

BACKGROUND AND RELATED WORK
In this section, we provide background information on the physiological effects of cocaine use, and review prior work from the ubiquitous computing community that has made use of wearable ECG for personal health sensing.

Effects of Cocaine Use: Cocaine addiction is associated with largely predictable and highly characteristic physiological, behavioral, and subjective effects [17]. Such effects de-

Figure 1. This figure illustrates two ECG periods. The peaks and troughs of the PQRST complex are labeled on the left period. Several ECG features that have been reported to undergo changes in the presence of cocaine use are also illustrated including the RR interval, the QT interval, the PR interval, the QRS interval and the T wave height.

rive directly from cocaine's well-established pharmacological mechanism of action: it is an indirect agonist/monoamine reuptake inhibitor. By virtue of its peripheral actions on the sympathetic nervous system, cocaine produces robust and predictable changes in primary indices of cardiovascular and neurological function (increases in heart rate, systolic, and diastolic blood pressure and pupillary diameter). As a "psychostimulant", cocaine also produces a characteristic profile of centrally-mediated, behavioral effects including increased arousal, alertness, decreased somnolence/sleep, increased talkativeness, decreased appetite, and involuntary motor activity ranging from non-specific increases in general activity level to highly repetitive and stereotypical movements, both simple and behaviorally complex.

In this work, we focus on the acute physiological effects of cocaine use on the heart. There is substantial evidence from human studies that cocaine use causes changes in cardiovascular function that are observable in electrocardiogram (ECG) signals [9, 14, 15, 19]. Figure 1 illustrates two periods of a typical ECG signal. We can see that each period is characterized by the presence of three peaks (P, R, T) and two troughs (Q, S), which are collectively known as the PQRST complex. Traditional ECG analysis focuses on how various relationships between the locations and heights of these peaks and troughs changes in response to differing conditions. Several such changes have been observed in subjects under the influence of cocaine.

As already mentioned, cocaine use has a robust effect on heart rate, causing it to increase significantly [19]. An increased heart rate manifests as a reduced ECG period length. Due to the ease of detecting the peak of the R wave, which is typically the highest peak in the ECG signal, the ECG period is usually measured using the RR interval as shown in Figure 1. While heart rate is known to increase with cocaine use resulting in a decrease in RR interval, this feature is obviously confounded by any other activity that influences heart rate (activity, stress, etc) and thus has limited usefulness in practice. Cocaine has also been reported to have an effect on the QT interval [19]. Some research has also made use of a "corrected" QT interval, QTc, meant to partially normalize out the effect of heart rate. QTc is typically computed as the length of the QT interval divided by the square root of the length of the RR interval (Bazett's correction) [22]. However, the literature

is contradictory as to whether QTc increases or decreases in the presence of cocaine. Magnano et al., [15] suggests that QTc prolongs in the presence of cocaine while Levin et al., [14] suggests that QTc shortens. Magnano et al. have also reported changes in the height and shape of the T-wave in the presence of cocaine. Animal studies have pointed to additional effects of cocaine on the PR and QRS intervals [10, 18]. The PR, QRS and T wave height (TH) are also illustrated in Figure 1. In this work, we focus our knowledge-based detection framework around these six ECG features (RR, QT, QTc, TH, PR and QRS) since they are well supported by the literature regarding the effect of cocaine on the heart.

Wearable ECG for Health and Activity Sensing: A substantial body of work has explored the use of ECG using wearable chestbands, primarily in the context of understanding physiological stress [1, 11], assessing cognitive load [8] and detection of arrhythmias caused specifically by atrial fibrillation [2, 3, 12]. The work on physiological stress focus primarily on the use of heart rate-based features to determine an individual's stress level. Hong et al [11] aim to predict physiological stress using simple ECG-based features (average and median of heart-rate, ECG amplitude, and RR interval) combined with a set of features derived from accelerometer and galvanic skin response (GSR) sensors. Haapalainen et al [8] estimate cognitive load experienced by an individual using a combination of sensors to monitor ECG, electroencephalogram (EEG), GSR, heat-flux, eye-movements and change in pupil size. They observed that the best feature for cognitive load level classification for most of the subjects was derived from the median absolute deviation of ECG. Research on atrial fibrillation has looked at extraction of specific features from the QRS complex. For example, [2, 3] uses QRS duration and PR interval to detect atrial fibrillation.

Our work clearly builds on past research regarding the effects of cocaine on the electrophysiology of the heart by focusing our knowledge-based detection framework around a set of ECG-based features that is well supported by the literature. While most of the previous studies, involving mobile and on-body physiological sensing, used heart rate and heart rate variability as features [1, 7, 11] only few studies used limited forms of morphological features [2, 3]. To the best of our knowledge our ECG feature set is the most extensive ever used. Unlike past studies from both the cocaine literature and the mobile sensing literature, we also consider a data-driven detection framework where detectors are built directly from complete ECG waveforms.

COCAINE USE STUDY DESIGN

As part of an ongoing National Institute on Drug Abuse-approved cocaine use study at Yale University, we have collected data from six medically healthy, non-treatment seeking, experienced cocaine users. Subjects typically participate in the study for a two week period. The study consists of multiple components that we describe in this section.

Dry-Out Period: When subjects are first admitted to the unit, they undergo a "dry-out" period to ensure that the acute influence of previous drug use does not affect the results of the

study. All subjects undergo a "dry-out" period that lasts for several days.
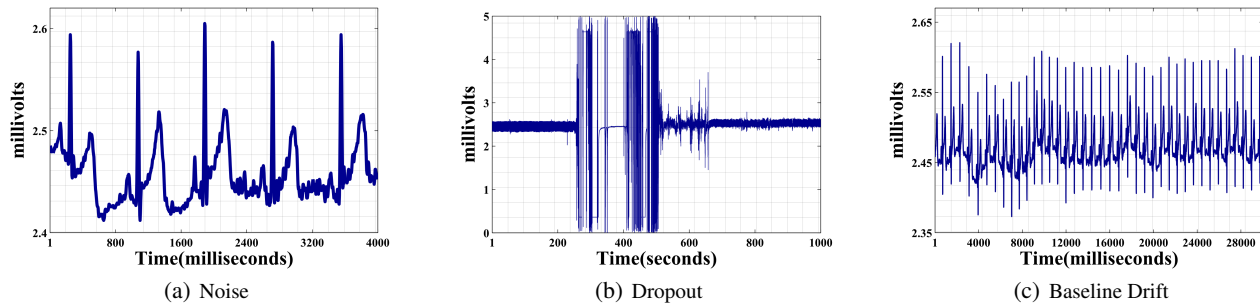
On-Body Sensor System: During the study protocol, the subjects wear a Zephyr Bioharness 3 chest band [23] which provides several raw (ECG, chest band diameter, accelerometer) and derived (RR interval, heart rate, respiratory rate) measurements. These chest bands are designed to be comfortable and less intrusive to wear than Holter monitors. We focus on the raw ECG data only in this work.

Our system encompasses two levels of data logging. The first level is on the sensor itself. The second is on a smartphone. The data on the sensor is downloaded at the end of each day and uploaded to a secure server. The sensor also transmits summary packets to the phone, which are periodically transmitted wirelessly to the secure server. The redundancy offered by logging via the phone as well as on the sensing device proved extremely useful since we found that the device failed to log locally on two occasions, but we were able to retrieve data transmitted from the phone.

Cocaine Administration Sessions: Subjects participate in a single 6-hour cocaine administration experiment comprised of several sessions. In this work, we use ECG data collected from a baseline session, three fixed-dose cocaine administration sessions and three cocaine self-administration sessions. These sessions appear in the same order for all subjects. Subjects wear the sensor system for all sessions. The baseline session is conducted at the end of the dry-out period and immediately before cocaine administration. It provides ECG measurements in the complete absence of cocaine. The three fixed-dose sessions last 20 minutes each. At the start of each of the these three sessions, the subjects receive a single-bolus intravenous (IV) injection of cocaine. The three cocaine sessions use a fixed-order, ascending dose regimen of 8, 16, and 32 mg per 70kg respectively with a 100kg cap per adjusted dose. This procedure is based on extensive prior experience, which has shown these doses and procedures to be safe, well tolerated, valid, behaviorally relevant, and test-retest reliable [20].

The main purpose of the baseline and fixed-dose sessions is to assess subjects for participation in subsequent cocaine self-administration sessions. Physiological (ECG, respiration) and behavioral (visual analog ccale self-ratings) assessments are conducted at five-minute intervals throughout each session. An advanced cardiac life support certified research nurse and a basic life support certified research assistant are also present. Subjects who exhibit a heart rate greater than 160 beats per minute, diastolic blood pressure greater than 110 mm Hg, systolic blood pressure greater than 180 mm Hg, and/or have evidence of clinically significant cardiac ectopy, arrhythmia, or other dangerous symptoms are excluded from further self-administration sessions.

The three self-administration sessions give subjects some control over the amount of cocaine they can receive. Each self-administration session uses one dosage level (8mg, 16mg or 32mg). The order of the dosage levels is randomized. The subject can click a button to receive an IV cocaine infusion

Figure 2. This figure depicts some of the issues that occur when using a wearable ECG device. The data are inherently noisy compared to ICU-quality ECG that uses careful skin preparation as well as adhesive electrodes. Various forms of signal dropout occur in our data, including cases that manifest as extreme noise. The data are also subject to baseline drift even over short time scales.

at the given dosage level within each self-administration session. There is a minimum period of 5 minutes enforced between subsequent infusions. All cocaine self-administration sessions take place at the YCCI-HRU. A saline lock, or peripheral intravenous device, is used for infusions of cocaine. Saline locks are maintained by trained research personnel in accordance with local, institutional policies and procedures.

In this work, we make use of the ascending dose regime in the baseline and fixed-dose sessions to explore the ability to discriminate between ECG waveforms from the baseline condition (no cocaine), and ECG waveforms from each of the three fixed dose sessions (8, 16, and 32 mg). We also explore the ability to discriminate between the ECG waveforms from the baseline session and those obtained from the complete set of cocaine sessions (the ascending dose regime combined with the self-administration sessions).

At the time of writing this paper, six subjects have completed the protocol. While the numbers may seem small, we note that in-patient studies on drug use are considerably harder to run and much more costly than typical user studies involving wearable sensors. Also, despite the low number of subjects, we are collecting hours of ECG waveform data per subject comprising thousands of heartbeats. We believe that the volume of data we have currently obtained is sufficient to explore the problem of cocaine use detection and provide an initial answer to our motivating question of whether cocaine use can be reliably detected based on wearable ECG sensor data.

DATA CHARACTERISTICS AND PROCESSING

We begin by presenting the noise characteristics of the ECG data obtained in our study to illustrate the challenges it presents. We then describe our data processing pipeline and how it extracts the features required to support both the data-driven and knowledge-based detection frameworks in the presence of these noise sources.

Noise Characteristics: We present several examples of raw data obtained from our sensing and data acquisition pipeline in Figure 2. This figure illustrates various difficulties with the use of a sensor like the Zephyr BioHarness chest band where no skin preparation is used and the electrodes are not adhesive. Figure 2(a) gives an indication of how noisy the raw data is in the best case. We also often see ECG periods that

have larger-scale distortions where the R wave may still be evident while the other peaks and troughs are not discernible. Such distorted periods would not pose a difficulty for features based on the RR interval only (heart rate), but they do pose challenges when attempting to extract the complete PQRST complex. Fortunately, these distorted periods appear to be transients and don't frequently occur in long runs.

Figure 2(b) shows an example of signal dropout resulting in extended intervals of extreme noise. These intervals are easy to identify because their characteristics differ widely when compared to normal signals. They contain no useful information and no features can be extracted from them. They typically result from large-scale disturbances to the sensor like completely removing or readjusting the chest band.

Finally, Figure 2(c) shows the degree to which the signal baseline drifts over short time spans. The baseline is also observed to drift over longer time spans. The long-run drift is likely due to slippage of the sensor over time. It is unclear what causes the short-run drift, but it is likely a hardware issue with the sensor itself. Again, the drift is a minor issue when extracting features based on the RR interval, but needs to be accounted for when extracting and standardizing morphological features.

Data Processing and Feature Extraction: Several ECG data processing toolkits provide functions to extract the PQRST complex from raw ECG waveforms (e.g. ECGBag [4]). However, we found that these toolkits are geared to the high-quality signals typical of ICU-grade sensors. While they can often be used to extract R peak locations, they lack robustness to noise and dropout when attempting to extract morphological features from signals obtained from the class of consumer-grade wearable ECG sensors that we use in this work. On the other hand, chest band-based ECG sensors have excellent ease of use in ambulatory settings when compared to multi-lead ECG devices like Holter monitors that also use adhesive electrodes.

It is also interesting to note that while consumer-grade devices such as the Zephyr Bioharness have been available for some time, most existing work only uses the ECG signal to extract basic features based on the RR interval like heart rate and heart rate variability. This is likely due to the difficulty of
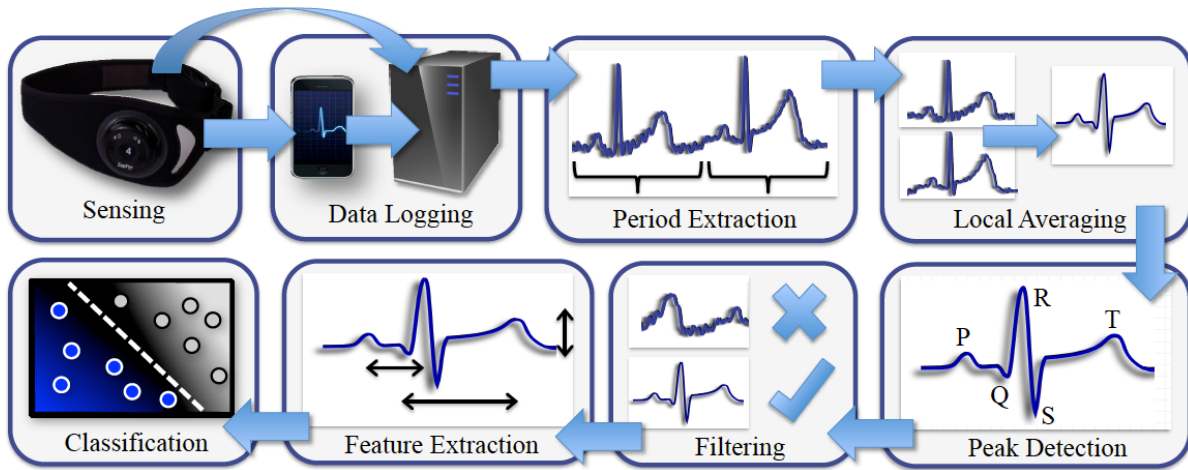
Figure 3. This figure illustrates the primary steps in our sensing, data acquisition and data processing pipeline. Raw ECG measurements are transmitted wirelessly to a smartphone and also downloaded directly to a server to provide redundancy. We first segment ECG periods using RR intervals. To deal with noise in the signals, we compute local averages over 30 second sliding windows. We apply peak detection to the smoothed waveforms and discard those that do not have the correct configuration of peaks and troughs. We apply feature extraction and standardization followed by classification. The above steps apply only to features in the knowledge-based framework. For features in the data-driven framework the local averaging step is directly followed by classification

extracting more nuanced morphological features in the presence of substantial noise. One aim in the current work is to bridge the gap between traditional ECG analysis based on features that require identification of the complete PQRST complex (as is common in the electrocardiology literature), and the analysis of ECG signals obtained from consumer-grade sensors. We illustrate the main components of the data processing and feature extraction pipeline that we have developed to deal with these issues in Figure 3. We describe the components in detail below.

Period Extraction: The easiest way to extract each complex is to split the ECG waveform into individual RR intervals. The R peak is easily localized because it clearly stands out as the highest amplitude peak in the raw ECG signal. We use the ECGBag [4] toolbox to identify the location of R wave peaks in the raw ECG trace. We use the R peak locations to segment the entire raw ECG trace into RR intervals. We remove spurious R peaks by computing the heart rate using the time between two adjacent R peaks, and discard those that fall outside normal heart beat range (50 - 150 bpm). This effectively removes much of the extreme noise due to drop out. However, the RR intervals that are extracted from the raw data are of non-uniform length since they vary inversely with heart rate. After storing the RR interval feature, we standardize all periods by resampling them to 100 samples per period, effectively removing all information about heart rate.

Smoothing by Local Averaging: To smooth out the effect of inherent noise as well as transient distorted ECG waves, we average the standardized waveforms within thirty second sliding windows. It is very important to note that this smoothing procedure is not appropriate for detecting transient morphological changes in ECG waveforms, which are often of interest in arrhythmia detection. It is appropriate in our case as the reported effects of cocaine on the electrophysiology of the heart are persistent over time scales of several minutes [13].
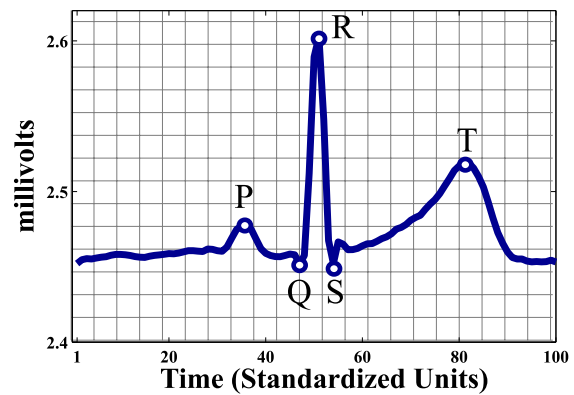

Figure 4. Example filtered and smoothed ECG period with extracted PQRST locations.

Having obtained the length-standardized and smoothed RR intervals, we remove the DC offset of each period separately to deal with baseline drift. We also divide by the standard deviation within each individual sample to normalize out any drift in overall amplitude. Figure 4 shows an example of the result of smoothing that can be contrasted with the raw data shown in Figure 2(a).

Peak Detection: Given the cleaned signal, we next need to identify the elements of the PQRST complex. To do this, we apply a standard peak detection algorithm to the absolute value of the smoothed and standardized waveforms. We take the five largest peaks and assign them the labels PQRST in temporal order. Waveforms that do not exhibit a peak-trough-peak-trough-peak sequence expected of the PQRST complex are filtered out as described below.

Filtering: When individual waves are as pronounced as the example shown in Figure 4, the PQRST complex is relatively easy to identify, as described above. However, we find that

(a) Morphological Features
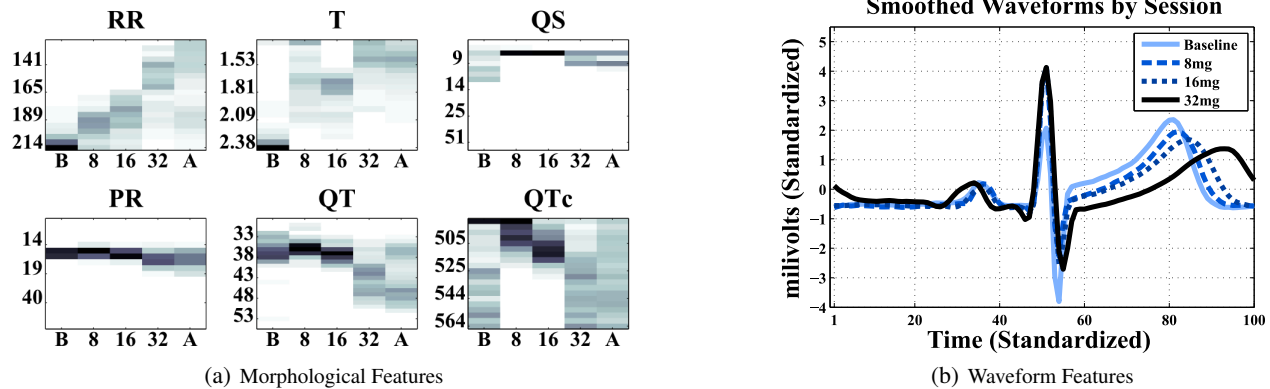

(b) Waveform Features

Figure 5. Figure (a) shows histograms represented as heat maps for each morphological feature as a function of dosage level (darker colors indicate higher frequency) for Subject 3. Figure (b) shows sample ECG waveforms, also from Subject 3, illustrating the effect of cocaine dosage level on the shape of the waveform.

some of the peaks are harder to detect than others, and residual noise in the waveform sometimes results in too many or too few peaks being detected. Since the subsequent feature extraction step is highly sensitive to incorrect identification of the PQRST complex, we apply several filtering steps to discard potentially problematic waveforms. First, we filter out any waveforms that do not result in exactly three peaks and two troughs in the correct sequence. Second, we filter out waveforms where one or more waveform values is more than three standard deviations away from the sliding window mean. Such waveforms may exhibit the correct sequence of peaks and troughs, but the presence of extreme values is likely due to noise and the detected peaks are then not reliable.

Feature Extraction: Given the standardized, smoothed and filtered periods with identified PQRST complex locations, the feature extraction step is straightforward. We have already obtained the RR interval length (prior to standardization) as a byproduct of the ECG period segmentation step. We define the QT, PR and QRS interval to be the distance between the identified locations of the Q and T peaks, P and R peaks and Q and S peaks, respectively.[1] These distances are all measured in the standardized waveform space where the RR interval length has been completely normalized away. To calculate QTc in its standard units, we then need to multiply our QT interval (where RR has be normalized away) by the square root of the original RR interval length. Finally, the T wave height is simply obtained as the height of the standardized waveform at the location identified as the peak of the T wave. Figure 5(a) shows example histograms of each of these feature values across dosage levels for one subject. The histograms are represented as heat maps. Each column of each figure represents the distribution over feature values for a single dosage level. Higher frequencies are displayed as darker colors. We can see that some features like RR, TH and QT show distinct

changes between the baseline (first column) and cocaine use (second to fifth column) sessions.

Data Volume: The table below shows the total number of samples available for each subject and each session following all data processing steps.

Subject	B	8	16	32	A
1	1224	258	1024	60	7883
2	2232	1489	915	1015	11945
3	1916	1830	999	1459	13630
4	1362	731	296	299	7950
5	131	405	1524	1454	21458
6	647	317	98	784	16148

Table 1. Number of data cases per subject for the baseline, 8mg, 16mg, and 32mg sessions, as well as for all cocaine sessions combined (fixed dose and self-administration)

COCAINE DETECTION

The final stage of our computational pipeline is detection of cocaine use. In this section we discuss the feature sets used, how we frame the cocaine use detection problem as a classification problem and describe our classification model.

Knowledge-Based and Data-Driven Detection: As described in the introduction, we consider two different detection frameworks that differ in the features they use. The previous section described at length how we compute the RR interval and the five morphological features (QT, QTc, PR, QRS, TH) that are supported by the literature on cocaine. These features form the core of our knowledge-based detection framework.

The second detection framework we consider is a purely data-driven framework where the feature representation for a period is the complete vector of ECG waveform amplitudes. For all the reasons described in the previous section, the use of raw ECG waveforms in the data-driven framework is not advisable. Instead, we use the length-normalized and smoothed waveforms produced by the first two steps of the pre-processing pipeline. In the data-driven framework, each

---

[1]We note that the electrocardiology literature recommends slightly different procedures for computing the QT, PR and QRS interval lengths based on where tangent lines to the individual waves intersect the baseline. We do not use this approach because numerical estimates of the tangent lines remain noisy even for our smoothed samples.

data case consists of a vector of 100 feature values representing the amplitude of the ECG waveform at each of the standardized time points. We refer to these features as the *waveform features* and use the abbreviation W.

The use of the complete waveform itself is motivated by the fact that we observe substantial differences in the shape of the ECG waveform across the four sessions (baseline, 8, 16, and 32 mg/70kg cocaine) in most subjects. These differences are illustrated in Figure 5(b) using examples waveforms drawn from the middle of each session. We can see that for this subject, as the cocaine dosage level increases, there are large-scale morphological changes in the ECG waveform including QT prolongation and T-wave flattening. The differences are substantial enough to suggest that discrimination between sessions should be easy given the full waveform representation, at least within subjects. Other subjects exhibit qualitatively similar changes as a function of cocaine dose.

The advantage of the fully data-driven approach is that even with the steps we have taken to reduce noise in the ECG signals, the peak detection step can still be quite sensitive. Using full waveforms allows us to avoid (or supplement) the peak detection-based features with a feature representation that does not require such processing and should thus be more robust to residual noise in the signals. The only drawback is that the number of individual features is obviously larger than when using single morphological features. However, there is no issue whatsoever fitting models to 100 features given that we have thousands of available data cases. In addition, the number of features is still low enough for detection algorithms to eventually run in real-time on a smart phone or other embedded system.

Cocaine Detection as Classification: Given a particular feature representation, we view the problem of constructing a detector for cocaine use as a standard binary classifier learning problem. We use the four experimental conditions in our study (baseline, 8, 16, and 32 mg/70kg cocaine) to form four sets of binary classification problems: baseline vs 8mg, baseline vs 16mg, baseline vs 32mg, and baseline vs all dosage levels combined. Within each problem, we construct classifiers based on each of the RR, QT, QTc, PR, QRS, TH and W features separately. In this work, we consider the problem of detecting cocaine use independently for each smoothed ECG period derived from each 30 second window. Each data case consists of a feature vector $\mathbf{x}_n$ extracted from the smoothed waveform and a corresponding class label y_n indicating which of the two sessions the data case belongs to. When using the RR, QT, QTc, PR, QRS or TH features, the feature vector $\mathbf{x}_n$ consists of one element only. When using the waveform features W, the feature vector $\mathbf{x}_n$ is length 100, containing the 100 waveform amplitudes. We also consider feature vectors containing all the morphological features (QT, QTc, PR, QRS and TH) and all of the morphological features combined with the waveform features.

Classification Model: In solving the detector construction problem, we can apply any existing binary classification model. In this work, we select a standard linear logistic regression classifier [6]. Given a feature vector $\mathbf{x}$ consisting

of D features, the binary logistic regression classifier returns the probability that the feature vector belongs to the positive class. Letting Y represent the label for the instance $\mathbf{x}$, logistic regression computes the class probability as shown below. θ is a length D vector of feature weights. It is easy to see that the classifier has a linear decision boundary specified by the weights θ.

$$P(Y = 1) = \frac{1}{1 + \exp\left(-(\theta^\top \mathbf{x} + b)\right)} \quad (1)$$

The default classification rule when using linear regression is to predict that the data case belongs to the positive class if $P(Y = 1) > 0.5$. Learning the weights of the logistic regression classifier is accomplished by maximizing the log likelihood of the training data using numerical optimization [6]. This is a continuous, convex optimization problem with no constraints. It can be solved using any gradient-based optimizer. In this work, we use the limited memory Broyden Fletcher Goldfarb Shanno (BFGS) algorithm [16]. Given a data set $\mathcal{D} = \{(y_n, \mathbf{x}_n)\}_{1:N}$, the log likelihood function is defined as shown below. We assume the labels for the two classes are -1 and 1.
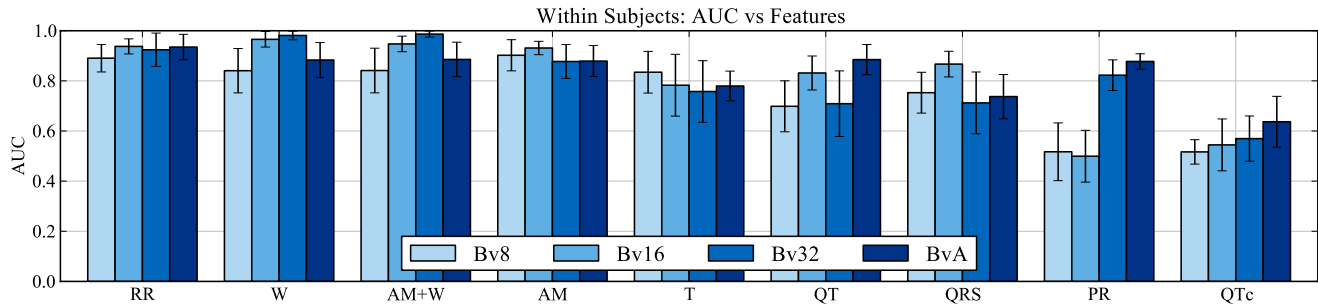
$$\mathcal{L}(\theta, b | \mathcal{D}) = -\sum_{n=1}^{N} \log \left(1 + \exp(-y_n(\theta^\top \mathbf{x}_n + b))\right) \quad (2)$$

The term within the sum is known as the *logistic loss*. It upper bounds the zero-one loss and is very similar to the hinge-loss function used in support vector machines (SVMs) [6]. While the slightly different properties of the logistic loss mean it does not yield a support vector property, this is irrelevant in the linear case since the model is explicitly parameterized in weight-space. Further, there is known to be very little difference between a learned linear logistic regression model and a learned linear SVM when the amount of available data is large. One advantage of linear logistic regression over the linear SVM is that it directly outputs probabilities, which are often more desirable in medical domain applications such as ours.

We do expect that non-linear classifiers including kernel logistic regression or kernel support vector machines will improve cocaine detection performance further. However, we leave an investigation of alternative classification models for future work. In this work, we instead focus on an evaluation of the utility of different features and feature combinations when used with the linear logistic regression model.

EXPERIMENTS AND RESULTS

In this section, we present experiments and results comparing multiple feature sets on the four classification problems introduced in the previous section: baseline vs 8mg (Bv8), baseline vs 16mg (Bv16), baseline vs 32mg (Bv32), and baseline vs all dosage levels combined (BvA). We consider two experimental conditions for each problem. In the within-subjects condition, we train and test detectors separately for each subject. In the between-subjects condition, we use each subject as the test subject in turn, and train the detector on data from the remaining subjects. We assess detection performance using the area under the receiver operating characteristic curve

Figure 6. This plot shows the results of within-subjects cocaine detection for all feature sets and four classification tasks. We see that the within-subjects detection task can be reliably solved with average AUC values approaching the maximum of 1.0 **at the higher dosage levels for some feature sets.**

(AUC) since there are sample imbalances between the classes both within and between subjects. Table 2 summarizes the feature sets that we use in our analysis. KB indicates features from the knowledge-based framework while DD indicates features from the data-driven framework.

Identifier	Description	Framework
RR	RR interval length	KB
QT	QT interval length	KB
QTc	Corrected QT interval	KB
PR	PR interval length	KB
QRS	QS interval length	KB
TH	Height of T wave	KB
W	Waveform Features	DD
AM	All morphological features (QT, QTc, PR, QRS, TH)	KB
AM+W	All morphological features plus waveform features	DD/KB

Table 2. Summary of Feature Sets

Within-Subject Cocaine Detection: In this section, we look at within subject classification where the classifier is trained and tested on data from the same subject. While training a classifier for each individual user of such a detection system is clearly not practical, studying within-subject classification sheds light on which features work best if we ignore between-subject variability in baseline ECG waveform structure, habituation, and cardiac response to cocaine.

In the within-subjects setting, we split the data for each of the four classification problems (Bv8, Bv16, Bv32 and BvA) into a training and testing set for each each subject. Splitting the data for each subject randomly is not a valid protocol and will result in inflated accuracy since the features and labels are highly correlated in time. Instead, we divide each subject's data for each session into two equal halves. We use the first half of both session to train the classifier and the second half of both session to test the classifier. We train and test detectors separately for each subject using this protocol. We report the mean AUC averaged across subjects as well as the standard error of the mean in Figure 6 for each of the four classification problems and each of the nine feature sets. The features are ordered from highest average AUC across the four tasks to lowest average AUC.

First, we see that the detection problem can often be more accurately solved when the dosage level is higher than when it is
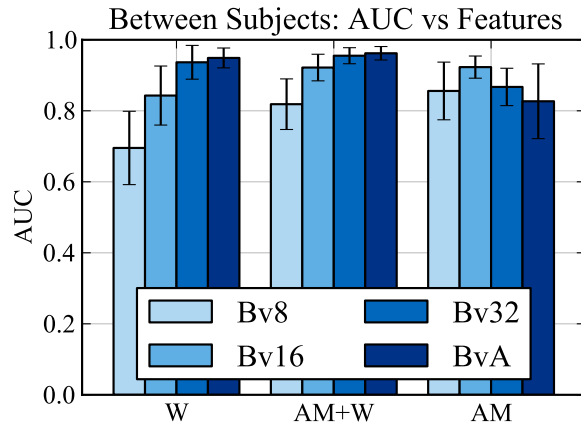
lower. This is certainly the case with the waveform features. It is consistent with the pattern of increasing morphological deformation as a function of dosage presented in Figure 5(b). Determining why this pattern is not observed consistently for all features will require further investigation.

In terms of the feature sets, we see that the RR interval length performs the best on average among the sets of features we have investigated, followed closely by the waveform features and the combination of the waveform features and morphological features. The AUC differences between these three feature sets are not statistically significant as evidenced by the overlapping error bars. Importantly, the average AUC across all four classification problems is above 0.8 out of a maximum of 1.0, and for the waveform features, the AUC for baseline vs 32mg is nearly equal to 1.0. We note that the performance ranking of the morphological features generally agrees with the degree of overlap in the feature distributions across dosage levels as seen in Figure 5(a). Among the morphological features, T and QT perform best. These two features along with RR exhibit the least overlap in feature distributions.

Before moving on to the between-subjects case, we pause to consider the potential of the RR feature as a basis for cocaine detection outside of the clinical setting. While the RR interval has very good performance in the clinical setting, it is obviously confounded by any other activity that results in an increase of base heart rate, as mentioned previously. The fact that the waveform features yield essentially the same accuracy while completely removing the effect of heart rate is thus very encouraging. It implies that there are significant differences in the shape of the ECG waveform in the presence and absence of cocaine. Since gross heart rate is so easily confounded, we do not consider the RR interval to be a viable feature for practical use in a cocaine detection system.

Between-Subject Cocaine Detection: We now turn to between-subjects cocaine detection. In these experiments we use all of each subject's data as test data, training on all of the data from the remaining subjects. We report the mean AUC averaged over all subjects as well as the standard error of the mean. We restrict our attention to the three best sets of features identified in the within-subjects experiment. We

Figure 7. This plot shows the results of between-subjects cocaine detection for the three best feature sets found in the within subjects analysis and all four classification tasks. We see that the AUC results are again approaching the maximum of 1.0 **at the higher dosage levels.**

exclude the RR interval from consideration for the reasons discussed above.

These results again show that the average AUC is always above 0.8 for the best feature set and again approaches the maximum possible value of 1.0 at the higher dosage levels. In this case, we see that the combined feature set AM+W performs best. The performance gain of AM+W over W in this case indicates that there is some benefit to combining the waveform and morphological features in the between-subjects case. This is not surprising as there are some significant differences between the characteristics of the ECG waveforms from different subjects (one subject has an average R peak height nearly 1.5 times higher than the other subjects, for example.)

DISCUSSION AND CONCLUSIONS

The results presented in the previous section support a very positive answer to the motivating question driving our current research: we can indeed reliably detect cocaine use based on data from wearable ECG sensors using appropriately chosen features. While results obtained in the within-subjects setting are quite remarkable, collecting data to learn a customized detector for every individual is clearly not feasible outside of the clinical setting. However, the fact that high detection accuracy was also obtained in the between-subjects setting at higher dosage levels is extremely promising. We expect that the between-subjects results can be further improved given training data from additional subjects. We also expect that clustering subjects according to their ECG waveforms and building detectors at the cluster level will help to account for between-subjects variability in basic ECG waveform characteristics.

There are several additional exciting areas for future research. At the sensing level, we are interested in exploring what additional information we might obtain from other modalities such as galvanic skin response that may also correlate with cocaine use. At the data processing level, we plan to investi-

gate solving the ECG period segmentation and PQRST complex extraction problems as joint inference problems in a unified probabilistic model. We expect that this will result in improvements in our solutions to both sub-problems. At the detection level, we plan to investigate non-linear classification methods including kernel logistic regression and kernel support vector machines, which will likely lead to improvements in detector reliability. We also plan to re-cast the cocaine use detection problem as a structured prediction problem as opposed to many independent classification problems. This will allow us to smooth predictions over longer time intervals, likely leading to further performance improvements.

In the clinical setting, there are several additional problems that we are interested in adapting our current methods to. A more difficult problem than detecting whether an individual is under the influence of cocaine is identifying the precise instant the drug was administered. This will require data from additional subjects since, while we have many ECG periods, there are relatively few infusion events per subject. Another problem is predicting the cumulative concentration of the drug through time. While there are discrete infusion events, cocaine diffuses out of the bloodstream with a half-life measured in tens of minutes. Learning to predict effective concentrations is difficult because we currently lack ground truth training data on the concentration of cocaine in the bloodstream. However, it may be possible to apply currently available mathematical models of cocaine diffusion and metabolization to compute estimated effective concentrations given the discrete time course of infusions. Finally, our current study also includes the collection of survey-type data from subjects at five minute intervals during self administration including questions that probe an individual's perception of their level of cocaine intoxication. We plan to explore how well these self assessments correlate with estimated and predicted effective concentrations.

Finally, looking toward the use of this technology outside of the clinical setting presents several additional challenges. First, the subjects were seated for the duration of the cocaine administration sessions since the infusions were delivered intravenously. If this constraint were removed, we would expect to see additional noise in the ECG data due to increased motion of the sensor. Our pre-processing pipeline has not been evaluated in such settings yet, but we expect that it would easily handle more frequent noise of the type we have already encountered. There is also an opportunity to use accelerometer data from the chest band to assist with filtering out ECG samples with potential artifacts due to motion of the subject. Second, there are significant differences between the intravenous cocaine administration method used in our study and real-world self administration methods and forms of the drug. This may result in different lag times between administration of the substance and it's observable effect on the cardiovascular system. However, we expect the effects to be similar to those we have observed in our study at equivalent effective concentrations.

REFERENCES

1. Alamudun, F., Choi, J., Gutierrez-Osuna, R., Khan, H., and Ahmed, B. Removal of subject-dependent and activity-dependent variation in physiological measures of stress. In *Pervasive Computing Technologies for Healthcare, Proceedings of the 6th International Conference on* (2012), 115–122.

2. Bouhenguel, R., and Mahgoub, I. A risk and incidence based atrial fibrillation detection scheme for wearable healthcare computing devices. In *Pervasive Computing Technologies for Healthcare, Proceedings of the 6th International Conference on* (2012), 97–104.

3. Bouhenguel, R., Mahgoub, I., and Ilyas, M. An energy efficient model for monitoring and detecting atrial fibrillation in wearable computing. In *Body Area Networks, Proceedings of the 7th International Conference on* (2012), 59–65.

4. Clifford, G. D. Ecgtools.
   `http://www.robots.ox.ac.uk/~gari/code.html`.

5. Elman, I., Krause, S., Breiter, H., Gollub, R. L., Heintges, J., Baumgartner, W. A., Rosen, B. R., and Gastfriend, D. R. The validity of self-reported drug use in non-treatment seeking individuals with cocaine dependence: correlation with biochemical assays. *The American Journal on Addictions 9*, 3 (2000), 216–221.

6. Friedman, J., Hastie, T., and Tibshirani, R. *The elements of statistical learning*, vol. 1. Springer Series in Statistics, 2001.

7. Gaggioli, A., Pioggia, G., Tartarisco, G., Baldus, G., Corda, D., Cipresso, P., and Riva, G. A mobile data collection platform for mental health research. *Personal Ubiquitous Computing 17*, 2 (2013), 241–251.

8. Haapalainen, E., Kim, S., Forlizzi, J. F., and Dey, A. K. Psycho-physiological measures for assessing cognitive load. In *Ubiquitous computing, Proceedings of the 12th ACM international conference on* (2010), 301–310.

9. Haigney, M. C., Alam, S., Tebo, S., Marhefka, G., Elkashef, A., Kahn, R., Chiang, C., Vocci, F., and Cantilena, L. Intravenous cocaine and qt variability. *Journal of cardiovascular electrophysiology 17*, 6 (2006), 610–616.

10. Hale, S. L., Lehmann, M. H., and Kloner, R. A. Electrocardiographic abnormalities after acute administration of cocaine in the rat. *The American journal of cardiology 63*, 20 (1989), 1529–1530.

11. Hong, J.-H., Ramos, J., and Dey, A. K. Understanding physiological responses to stressors during physical activity. In *Ubiquitous Computing, Proceedings of the 2012 ACM Conference on* (2012), 270–279.

12. Hu, S., Shao, Z., and Tan, J. A real-time cardiac arrhythmia classification system with wearable electrocardiogram. In *Body Sensor Networks, Proceedings of the 2011 International Conference on* (2011), 119–124.

13. Kolbrich, E., Barnes, A., Gorelick, D., Boyd, S., Cone, E., and Huestis, M. Major and minor metabolites of cocaine in human plasma following controlled subcutaneous cocaine administration. *Journal of analytical toxicology 30*, 8 (2006), 501–510.

14. Levin, K., Copersino, M., Epstein, D., Boyd, S., and Gorelick, D. Longitudinal ECG changes in cocaine users during extended abstinence. *Drug Alcohol Depend 95*, 1-2 (2008), 160–163.

15. Magnano, A., Talathoti, N., Hallur, R., Jurus, D., Dizon, J., Holleran, S., M., B. D., Collins, E., and Garan, H. Effect of acute cocaine administration on the QTc interval of habitual users. *The American journal of cardiology 97*, 8 (2006), 1244–1246.

16. Nocedal, J., and Wright, S. J. *Numerical optimization*. Springer verlag, 1999.

17. O'Brien, C. P. Evidence-based treatments of addiction. *Focus 9*, 1 (2011), 107.

18. Schwartz, A., Janzen, D., Jones, R., and Boyle, W. Electrocardiographic and hemodynamic effects of intravenous cocaine in awake and anesthetized dogs. *Journal of electrocardiology 22*, 2 (1989), 159–166.

19. Schwartz, B. G., Rezkalla, S., and Kloner, R. A. Cardiovascular effects of cocaine. *Circulation 122*, 24 (2010), 2558–2569.

20. Sughondhabirom, A., Jain, D., Gueorguieva, R., Coric, V., Berman, R., Lynch, W. J., Self, D., Jatlow, P., and Malison, R. T. A paradigm to investigate the self-regulation of cocaine administration in humans. *Psychopharmacology 180*, 3 (2005), 436–446.

21. Vongpatanasin, W., Taylor, A. J., and Victor, R. G. Effects of cocaine on heart rate variability in healthy subjects. *The American journal of cardiology 93*, 3 (2004), 385–388.

22. Wikipedia. QT_c interval.
    `http://en.wikipedia.org/wiki/QT_interval`.

23. Zephyr. Bioharness 3. `http://www.zephyr-technology.com/products/bioharness-3/`.