HHS Public Access

Author manuscript

Int Conf Comput Netw Commun. Author manuscript; available in PMC 2021 March 16.

Published in final edited form as:

Int Conf Comput Netw Commun. 2020 February ; 2020: 445–449. doi:10.1109/icnc47757.2020.9049684.

Harmonizing Wearable Biosensor Data Streams to Test Polysubstance Detection

Joshua Rumbut[1,2], **Hua Fang**[1,2], **Honggang Wang**[3], **Stephanie Carreiro**[4], **David Smelson**[5], **Brittany Chapman**[4], **Edward Boyer**[6]

[1]Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA

[2]Computer and Information Science, University of Massachusetts Dartmouth North, Dartmouth, MA, USA

[3]Electrical and Computer Engineering, University of Massachusetts Dartmouth North, Dartmouth, MA, USA

[4]Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA USA

[5]Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA USA

[6]Department of Emergency Medicine, Harvard Medical School, Boston, USA

Abstract

Wearable biosensors, as a key component of wireless body area network (WBAN) systems, have extended the ability of health care providers to achieve continuous health monitoring. Prior research has shown the ability of externally placed, non-invasive sensors combined with machine learning algorithms to detect intoxication from a variety of substances. Such approaches have also shown limitations. The difficulties in developing a model capable of detecting intoxication generally include differences among human beings, sensors, drugs, and environments. This paper examines how approaching wireless communication advances and new paradigms in constructing distributed systems may facilitate polysubstance use detection. We perform supervised learning after harmonizing two types of offline data streams containing wearable biosensor readings from users who had taken different substances, accurately classifying 90% of samples. We examine time domain and frequency domain features and find that skin temperature and mean acceleration are the most important predictors.

Keywords

data stream; feature extraction; classification; addiction; biosensor

jrumbut@umassd.edu.

I. Introduction

Wearable biosensors are increasingly popular in healthcare, athletics, and among consumers in general. New devices, with increasing capabilities and reduced cost regularly come on the market. Continuous health monitoring, which enables clinicians to learn how their patients are faring outside the office or hospital, is among the most promising aspects of connected health (or mHealth) [1]. Wearable biosensors and mobile phones capable of communicating to the cloud or performing processing at the edge represent an unprecedented opportunity to deliver just-in-time interventions for patients facing behavioral health and substance use disorders. Furthermore, this technology empowers patients to take an active role in their self-care [2].

Like many other Internet of Things (IoT) technologies, implementation of continuous health monitoring accelerated during the fourth generation (4G) mobile communications era. Fifth generation (5G) network infrastructure as planned will have many features that could remove some of the limitations faced by wireless body area network systems (WBANs) in terms of ease of use, robustness, throughput, and flexibility for implementers [8]. 5G support for software defined networking (SDN) and mobile edge computing (MEC) may be important in enabling secure and continuous monitoring, particularly of sensitive conditions such as drug or alcohol dependence.

In the past, we have developed classifiers capable of detecting opioid [19], cocaine [24], and kratom use [20] in subjects wearing the Empatica E4 wristband. A limitation of this work has been the exclusion of individuals who simultaneously ingest multiple substances, which is a common scenario is real-world settings. Furthermore, in some cases models trained on the data streams of specific patients had difficulty making accurate predictions on previously unseen subjects. This issue is not uncommon in machine learning applications but presents a particular setback since sustaining reliability across diverse populations is a goal.

For data streams from wearable biosensors, at least three sources of inter-subject variation are present. The first is the inherent variation in different individuals' resting heart rates, and rates of metabolizing substances, etc. The second is the differences among sensors, which may grow greater over time as wear and tear and lack of calibration take effect [11]. The third is difference in environment, where one user may work in an air conditioned office and the other outdoors in all weather. These challenges must be overcome for the creation of a robust system capable of being a part of the health care decision making process.

Previous studies have shown that wearable biosensors combined with pattern recognition algorithms are capable of recognizing the effects of cocaine and other drugs when worn by users [7], [19], [24]. Other tests have shown potential for detecting other changed mental states such as psychological distress [21]. Previous studies have been limited by their focus on a single substance, often excluding individuals using more than one drug simultaneously. As continuous health monitoring develops, a key ability will be detecting a wide range of health conditions, since having a system for each health problem would result in massive duplication of effort.

In this paper, we harmonize multiple wearable biosensor data streams to create a larger, more diverse data stream from which to further develop feature extraction and classification techniques. We examine the potential of frequency domain features to improve model performance while also considering their implementation on low power systems. By summarizing the stream of tens or hundreds of thousands of sensor readings into 43 features, we are moving toward a potential system which is capable of detecting polysubstance intoxication in previously unseen individuals without burning through batteries, bandwidth, and data transfer limits.

II. Methods

In order to more rigorously test our feature extraction and model training process, we combined two big data streams. Harmonization proceeded smoothly since the same model wearable, the Empatica E4 wristband, recorded sensor data in both groups of subjects. The sensor data from each experiment was stored in a similar structure, with a subject and session ID number encoded in the folder name and then CSVs corresponding to each of the sensor measurements and some features which are derived by the device's software. There were, however, considerable differences in patterns of usage that were recorded.

The process of merging the two datasets was as follows:

Algorithm 1:

Data Harmonization Procedure

```
Result: A data frame containing features and labels from
        two studies using Empatica E4 wristband
Data: Sessions /* All sessions from both
      studies                                    */
begin
    Data ← List() ;
    foreach Session in Sessions do
        if Validate(Session)s then
            /* See Eq. 1                         */
            Session ← TrimTime(Session) ;
            /* See Eq. 3                         */
            CombinedAcc ← Magnitude(
                Session→Acceleration_x,
                Session→Acceleration_y,
                Session→Acceleration_z) ;
            Features ← Extract(
                CombinedAcc,
                Session→{Temp, EDA, HR}) ;
            Append(Data, Features) ;
        end
    end
    return Data;
end
```

For a session to be valid, it must have all features present (a small number of sessions had features missing for unknown reasons) and the readings must vary at least slightly (one session had all EDA readings equal to 0, compared to an overall mean of 2.27 and a standard deviation of 1.97).

Algorithm 2:

Session Validation Procedure

```
Result: A determination whether the session is valid for
        inclusion
Data: Session /* An object containing
      second readings                          */
    SessionType /* Kratom or ED                */
foreach Feature in Session do
    if Not Exists(Feature) or
    Min(Feature) = Max(Feature) then
        return False;
    end
end
if SessionType = "Kratom" and
Count(Session→Events) = 0 then
    return False;
end
return True ;
```

The first dataset described previously [20], consists of sensor data from three adult males in Malaysia who were frequent users of a narcotic tea made from a plant called kratom. While kratom has a long history of use in locations where the plant naturally occurs, its effects are not as well quantified as those of pharmaceutical opioids. Users have reported a variety of effects not commonly seen in other μ and κ opioid receptor agonists [23]. The participants wore the wristband while going about their routine, pressing an event marker button on the sensor to record kratom use. Twenty nine sessions were recorded, but 6 were excluded from the dataset: 3 lacked any reports of kratom use, 2 lacked accelerometry readings, and 1 had EDA readings which were consistently 0. The etiologies of these data anomalies are unknown.

The second dataset was gathered from patients who presented to a single emergency department (ED) suffering from an opioid overdose and had recently or were about receive naloxone, which is an antidote to opioid toxicity. The Empatica E4 wristband was placed on the patient's wrist as soon as possible after verbal consent was received. The wristband was removed when the patient was discharged or admitted, gave indications they may attempt to leave with the wristband, were placed on a continuous naloxone drip, or after four hours.

Patients enrolled in the ED were classified into three groups. The first are those who experienced an opioid overdose and remained breathing or responsive and a single dose of naloxone (non-recurrent). The second group (recurrent) required multiple doses of naloxone since its duration of action, estimated by elimination half-life at 60–90 minutes [10], may be insufficient for overdoses of long-lasting opioids or very large doses. The third group were primarily treated for an opioid overdose but self-reported use of additional substances from

other drug classes (polypharmacy). A medical toxicologist reviewed the clinical encounters to determine whether they were consistent with the self-reports, to reduce the chance of bias.

Totals for these groups can be seen in Table I. Additionally since there were only two samples, we grouped the recurrent and non-recurrent opioid overdose sessions together, creating one class for all opioid incidents and another for polypharmacy patients.

See Fig. 1 for a visualization of the relative distances between different sessions projected into a two dimensional space by Sammon mapping. We can see most of the opioid and polypharmacy sessions are relatively compact and overlapping. A subset of kratom sessions and a single opioid session are much more dispersed, with relatively large distances indicating they do not much similarity to the others.

A. Feature Extraction

The raw features used were acceleration in the X, Y, and Z dimensions (measured in $\frac{1}{64}^{th}$ g), electrodermal activity (EDA, also called galvanic skin response) measured in microsiemens, skin temperature which was measured in degrees Celsius, and heart rate in beats per minute which was measured by a photoplethysmograph. Each of the 4 sensors can begin and end at different times. The start times, following Unix time format, are integers. Following Empatica's documentation we assumed that recording began at the exact start of the second [12]. The ending time (t_{end}) was calculated to be

$$t_{end} = \min(t_{start} + \lfloor N_s T_s \rfloor) \forall s \in Sensors \quad (1)$$

where $Sensors$ is the set of all sensors t_{start} is the latest start time among the sensors, N_s is the number of readings for a given sensor, and T_s is the sampling period of the sensor,

$$T = \frac{1}{f_s} \quad (2)$$

where f is the sampling frequency of the sensor in hertz. All readings with timestamps outside the interval $[t_start, t_end]$ were discarded. In general, the amount trimmed was between 10 and 20 seconds.

The Empatica E4 is capable of streaming realtime sensor readings to a mobile phone or other device via the Bluetooth Low Energy protocol [13], which may process the data at the edge or forward it to a server in the cloud. Many cell phone carriers offer plans that charge $10-$15 per GB of data in the United States currently [25]. When all the sensor readings are loaded into an object in the R statistical computing environment, they require 1022.4 MB of memory for 57 sessions (18 MB per session). Compression and other techniques could help, but over the course of years transferring these large collections of sensor readings would become an onerous expense. Sending this stream directly a cloud server for processing would be wasteful. Users will benefit from the development of a suitable way to summarize the wearable biosensor data in a way that preserves the ability of classifiers to detect disturbances in their health.

With the goal of keeping processing power requirements low, we examined a large set of relatively simple to calculate features from the time and frequency domains. Our prior work has shown that accelerometer data stream has some predictive power, but there has been inconsistency in which dimension is most valuable. In some of our previous work, smoothing of the sensor signal has been done by Hilbert transform [9]. This process represents a simplification over previous efforts in that sensor readings are taken as is and the representation stays the same size regardless of the size of the stream (although aggregating more readings naturally takes more time and memory).

When looking exclusively at cocaine use, the z-axis of the accelerometer readings were the most predictive. A possible interpretation of this is that users were less likely to remain seated for long periods of time while under the influence of a stimulant, since the Z axis would often be perpendicular to the floor. We treated the accelerometer values as components of a three dimensional vector:

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}\,. \tag{3}$$

For each of the sensor inputs, we calculated the features seen in Table I. The features extracted from the time domain signal are straightforward enough, and can all be calculated in $O(n)$ time [3]. For the frequency domain, the Discrete Fourier Transform was computed using a Fast Fourier Transform (FFT) algorithm, transforming the vector of readings X into its frequency domain representation. Different FFT algorithms have been proposed and have been used in environments where processing speed and power consumption are constrained, however none have are below $O(n \log n)$ time complexity [11].

B. Classification

Four different classification algorithms were considered. Two of them produce highly interpretable models, logistic regression and naive Bayes, since such a model could have the additional value of shedding additional light on polysubstance intoxication if a good fit could be found to the data stream. Additionally, we tested two ensemble learning algorithms. Ensemble learning works by training many models and performs classification by combining their predictions [17]. The first is random forest, which grows large decision trees (which are low bias models) and attempts to reduce variance by only considering a subset of predictors at each split in the decision tree [5]. Gradient boosting machines (GBM, specifically eXtreme Gradient Boosting) creates many small trees, with each seeking to reduce the loss of the previously trained models in the ensemble [16].

Parameter tuning was performed by 10-fold cross validation, except for the random forest classifier where model selection was performed using out-of-bag error repeated 10 times. The original data was split 80%−20% between training and testing with outcome variable proportions approximately preserved. Grid search was carried out over a range of parameters in order to find good values. Models were trained and tuned using the caret package for R [18] on a desktop computer running Windows 10 with a Intel(R) Core(TM) i7–4790K CPU @ 4.00GHz and 32 GB of memory.

III. Results

Training with cross validation took 130 seconds for the GBM model, 2 seconds for the random forest model, 1 second for KNN, and less than a second for naive Bayes. While it may take longer, it would be feasible to train and tune any of these models on a late model smart phone. This application of MEC would create a model accustomed to the pecularities of the individual, their sensor, and their accustomed activities and environment. Additionally, the user would maintain control of their data. It would, however, require users to manually label enough data to train the model and prevent the development of a database of sensor recordings that could be useful for other purposes.

The random forest and GBM classifiers achieve the best performance on these data in terms of accuracy, sensitivity, and specificity, see Table II. Both models made the same predictions, which are shown in Table III. The ensemble classifiers were able to differentiate kratom users from those in the ED overdose dataset, but struggle to separate a pure opioid overdose from the polypharmacy patients (see Table II). KNN and naive Bayes were not able to do either.

As has occurred in our previous work, acceleration and temperature are highly important predictors [6], [19]. This has become a consistent pattern, although in the past it was accompanied by EDA [14], while that did not occur this time. A possible interpretation of this is that the different substances do not have readily detectable differences in the way they affect EDA, and so it cannot be used to differentiate different substances the way that it can be used to determine when a user is intoxicated. Given that EDA failed to record during one of the sessions, there is a possibility that one of the devices or the way that a user wore it interfered with the sensor and reduced the utility of EDA as a feature. Detecting and compensating for the failure or miscalibration of a sensor, as is certain to occur over longer usage, is a challenge that will be faced by all continuous health monitoring projects. Expanded variable importance scores for both random forest and GBM can be seen in Table IV.

IV. Conclusion

Although merging the datasets allowed for a broader population, this work remains limited by small sample sizes and lack of long term observation. Additionally, drug use was recorded differently in the two samples. The kratom user data may include time before taking the drug or after it worn off. We have estimated that kratom begins affecting the user 5 minutes after consumption and remains in effect for 2 hours, but this may vary among different users. For the ED patients, the timing of the dose and the exact opioid or other drug taken are unclear, complicating efforts to quantify the extent to which users were intoxicated throughout their recording sessions. Future work will continue to more closely analyze the individual’s response to ingestion of opioids and exploring how emerging wireless communication technology can enable continuous health monitoring.

Acknowledgments

This research was supported by the Seed Funding program from the Office of the Provost at the University of Massachusetts Dartmouth to Dr. Fang, a Fulbright Research Scholarship to Dr. Boyer, and NIH/NIDA K23DA045242 to Dr. Carreiro.

References

[1]. Aileni RM, Suciu G, Suciu V, Pasca S, and Strungaru R, ”Health Monitoring Using Wearable Technologies and Cognitive Radio for IoT” in Cognitive Radio, Mobile Communications and Wireless Networks, Rehmani MH and Dhaou R Eds. Springer, 2019, pp. 143–165.

[2]. Anderson K, Burford O, and Emmerton L, ”Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences, PLoS One, vol. 11, no. 5, 2016.

[3]. Blum M, Floyd RW, Pratt V, Rivest RL, and Tarjan RE, ”Linear time bounds for median computations,” In Proc of the fourth annual ACM symposium on Theory of computing, 1982, pp. 119–124.

[4]. Boyer EW, Babu KM, Adkins JE, McCurdy CR, and Halpern JH, ”Self-treatment of opioid withdrawal using kratom (Mitragynia speciosa korth),” Addiction, vol. 103, no. 6, pp. 10481050, 6 2008. doi:10.1111/j.1360-0443.2008.02209.x.

[5]. Breiman L, ”Random forests,” Machine learning, vol. 45, no. 1, pp. 5–32, 2001. doi:10.1023/A:1010933404324.

[6]. Carreiro S, Fang H, Zhang J, Wittbold K, Weng S, Mullins R, Smelson D, and Boyer EW, ”iMStrong: Deployment of a biosensor system to detect cocaine use,” Journal of Medical System, vol. 39, no. 12, p. 186, 12 2015. doi: 10.1007/s10916-015-0337-9.

[7]. Carreiro S, Smelson D, Ranney M, Horvath KJ, Picard RW, Boudreaux ED, Hayes R, and Boyer EW, ”Real-time mobile detection of drug use with wearable biosensors: a pilot study, J Med Toxicol, vol. 11, no. 1, pp. 73–9, 3, 2015. [PubMed: 25330747]

[8]. Chen M, Yang J, Zhou J, Hao Y, Zhang J, and Youn C-H, ”5G-Smart Diabetes: Toward Personalized Diabetes Diagnosis with Healthcare Big Data Clouds,” IEEE Communications Magazine, vol. 56, no. 4, pp. 16–23, 2018.

[9]. Chintha KK, Indic P, Chapman B, Boyer EW, and Carreiro S, ”Wearable Biosensors to Evaluate Recurrent Opioid Toxicity After Naloxone Administration: A Hilbert Transform Approach, In Proc Annu Hawaii Int Conf Syst Sci, vol. 2018, pp. 3247–3252, 1, 2018.

[10]. Clarke SF, Dargan PI, and Jones AL, ”Naloxone in opioid poisoning: walking the tightrope, Emerg Med J, vol. 22, no. 9, pp. 612–6, 9, 2005. [PubMed: 16113176]

[11]. Dargie W, ”Analysis of Time and Frequency Domain Features of Accelerometer Measurements,” in 2009 Proceedings of 18th International Conference on Computer Communications and Networks, 2009, pp. 1–6.

[12]. Empatica Inc. ”Session start-time format and synchronization.” Empatica.com. https://support.empatica.com/hc/en-us/articles/202800715-Session-start-time-format-and-synchronization- (accessed Jul. 24, 2019).

[13]. Empatica Inc. ”Stream your signals using the BLE connection.” Empatica.com. https://support.empatica.com/hc/en-us/articles/205683919-Stream-your-signals-using-the-BLE-connection (accessed Jul. 24, 2019).

[14]. Fang H, Wittbold M, Zhang J, Weng S, Carreiro S, Mullins R, and Boyer E, ”Describing realtime substanceuse detection from big biosensor data: a case study of cocaine users,” Health Behavior Methods and Measures paper session, Society of Behavioral Medicine, San Antonio, TX.

[15]. Fink M, Zaks A, Sharoff R, Mora A, Bruner A, Levit S, and Freedman AM, ”Naloxone in heroin dependence, Clin Pharmacol Ther, vol. 9, no. 5, pp. 568–77, Sep-Oct, 1968. [PubMed: 5676799]

[16]. Friedman J, ”Greedy function approximation: a gradient boosting machine,” The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 10 2001. Retrieved from http://www.jstor.org/stable/2699986.

[17]. Friedman J, Tibshirani R, and Hastie T, The Elements of Statistical Learning : Data Mining, Inference, and Prediction, ser. Springer Series in Statistics. New York: Springer-Verlag, 2017.

[18]. Kuhn M, ”Building Predictive Models in R Using caret Package,” Journal of Statistical Software, vol. 208, no. 5, 2008.

[19]. Mahmud MS, Fang H, Wang H, Carreiro S, and Boyer E, ”Automatic Detection of Opioid Intake Using Wearable Biosensor, in 2018 International Conference on Computing, Networking and Communications (ICNC), 2018, pp. 784–788.

[20]. Rumbut J, Fang H, Carreiro S, Singh D, Wang H, and Boyer EW, ”Detecting Kratom Intoxication in Wearable Biosensor Data,” in 2019 the IEEE/ACM 4th International Conference on Connected Health: Applications, Systems and Engineering Technologies: Demo and Poster, in press, 2019.

[21]. Schmidt P, Reiss A, Duerichen R, Marberger C and Van Laerhoven K. ”Introducing WESAD, a multimodal dataset for Wearable Stress and Affect Detection,” In 2018 International Conference on Multimodal Interaction (ICMI 18). October 16–20, 2018, Boulder, CO, USA. ACM, New York, NY, USA, 9 pages.

[22]. Trakulsrichai S, Sathirakul K, Auparakkitanon S, Krongvorakul J, Sueajai J, Noumjad N, Sukasem C, and Wananukul W, ”Pharmacokinetics of mitragynine in man, Drug Des Devel Ther, vol. 9, pp. 2421–9, 2015.

[23]. Ulbricht C, Costa D, Dao J, Isaac R, LeBlanc YC, Rhoades J, and Windsor RC, ”An evidence-based systematic review of kratom (Mitragyna speciosa) by the Natural Standard Research Collaboration,” J Diet Suppl, vol. 10, no. 2, pp. 152–70, 6, 2013. [PubMed: 23725528]

[24]. Wang CJ, Fang H, Carreiro S, Wang H, and Boyer E, ”A New Mining Method to Detect Real Time Substance Use Events from Wearable Biosensor Data Stream,” In Proceedings of International Conference on Computing, Networking and Communications: Social Computing and Semantic Data Mining, pp. 465470. ICNC2017. NIHMSID: 907527

[25]. Zhang M, Gao L, Huang J, and Honig ML, ”Hybrid pricing for mobile collaborative Internet access,” IEEE/ACM Transactions on Networking, 2019.

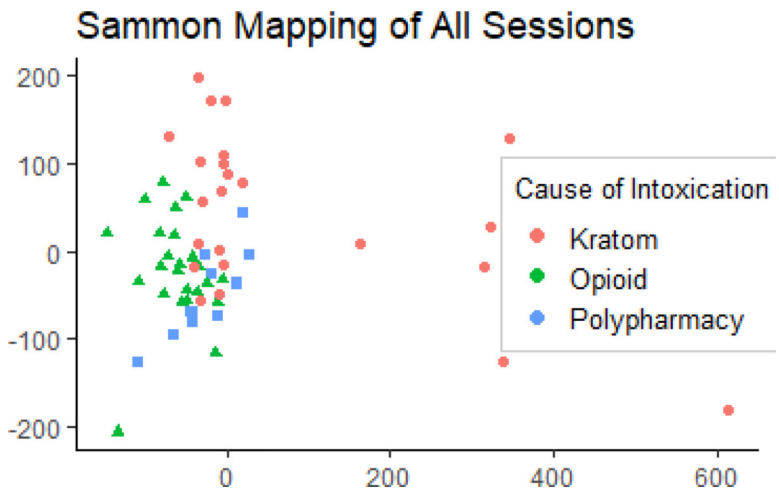

Fig. 1.
All values projected into a 2-dimensional space via Sammon mapping.

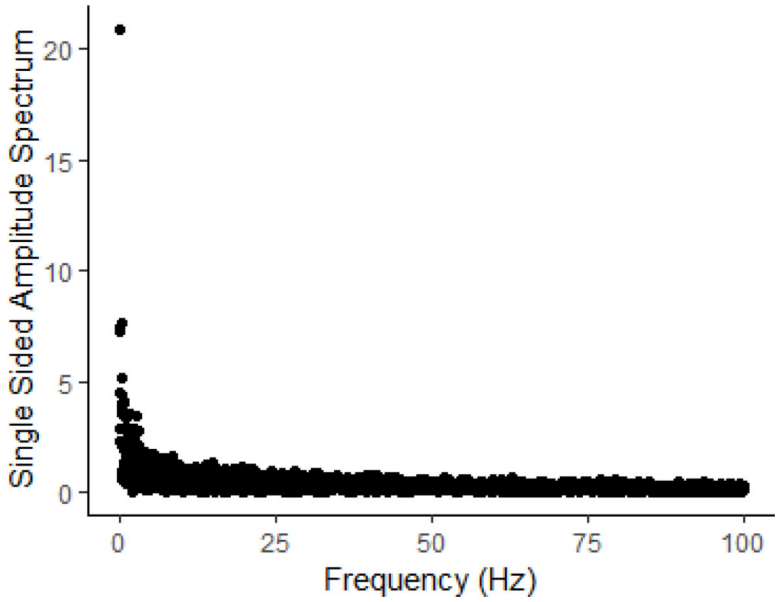

Fig. 2.
Frequency domain representation of accelerometer data.

TABLE I

Summary of Combined Dataset

Data Source	Features Extracted	Description
Accelerometry	Mean	Mean value of signal
	Median	Median value of signal
	Max	Max value of signal
	Min	Min value of signal
EDA	First Peak	Amplitude of first peak after DC
	Second Peak	Amplitude of second peak after DC
Heart Rate	First Peak Ord	Frequency of first peak after DC
	FreqInt	Integral of the spectrum estimated by trapezoid method
Temperature	First Peak Ord	Frequency of first peak after DC
	FreqInt	Integral of the spectrum estimated by trapezoid method
Drug Consumed	Count	Opioid: 24[a] Polypharmacy: 10 Kratom: 23

[a]22 non-recurrent and 2 recurrent

TABLE II

Model Results

Algorithm	Accuracy (95% CI)	Sensitivity (Kratom, Opioid, Poly)	Specificity (Kratom, Opioid, Poly)
KNN	0.7 (0.35, 0.93)	0.75, 1, 0	1, 0.5, 1
Naive Bayes	0.5 (0.18, 0.81)	0.25, 1, 0	1, 0.5, 0.75
Random For.	0.90 (0.56, 0.998)	1, 1, 0.5	1, 0.8, 1
GBM	0.90 (0.56, 0.998)	1, 1, 0.5	1, 0.8, 1

TABLE III

Predicted Versus Actual Classes in Test Set

	Reference		
Predicted	**Kratom**	**Opioid**	**Polypharmacy**
Kratom	4	0	0
Opioid	0	4	1
Polypharmacy	0	0	1

TABLE IV

Variable Importance

Random Forest		GBM	
Name	**Score**	**Name**	**Score**
Acc. Mean	100	Temp. Min.	100
Temp. Min.	97.8	Acc. Mean	51.6
Acc. FreqInt	84.1	Acc. FreqInt	39.1
Acc. SD	75.1	HR Min.	32.8
Acc. Median	56.3	Temp. Mean	25.8
Acc. Second Peak	42.6	Acc. SD	24.6
Acc. Max	39.4	HR First Peak	21.6
Temp. Mean	37.5	Acc. Min.	19.9
EDA Median	37.1	Acc. Second Peak	17.4
HR First Peak	36.3	EDA Median	15.9