A Machine Learning-based Approach for Collaborative Non-Adherence Detection during Opioid Abuse Surveillance using a Wearable Biosensor

Rohitpal Singh[1], **Brittany Lewis**[1], **Brittany Chapman**[2], **Stephanie Carreiro**[2], and **Krishna Venkatasubramanian**[1]

[1]Worcester Polytechnic Institute, Worcester, MA, U.S.A.

[2]University of Massachusetts Medical School, Worcester, MA, U.S.A.

Abstract

Wearable biosensors can be used to monitor opioid use, a problem of dire societal consequence given the current opioid epidemic in the US. Such surveillance can prompt interventions that promote behavioral change. The effectiveness of biosensor-based monitoring is threatened by the potential of a patient's ***collaborative non-adherence (CNA)*** to the monitoring. We define CNA as the process of giving one's biosensor to someone else when surveillance is ongoing. The principal aim of this paper is to leverage accelerometer and blood volume pulse (BVP) measurements from a wearable biosensor and use machine-learning for the novel problem of CNA detection in opioid surveillance. We use accelerometer and BVP data collected from 11 patients who were brought to a hospital Emergency Department while undergoing naloxone treatment following an opioid overdose. We then used the data collected to build a personalized classifier for individual patients that capture the uniqueness of their blood volume pulse and triaxial accelerometer readings. In order to evaluate our detection approach, we simulate the presence (and absence) of CNA by replacing (or not replacing) snippets of the biosensor readings of one patient with another. Overall, we achieved an average detection accuracy of 90.96% when the collaborator was one of the other 10 patients in our dataset, and 86.78% when the collaborator was from a set of 14 users whose data had never been seen by our classifiers before.

Keywords

Opioid Epidemic; Wearable Technology; Biosensor; Adherence; Machine Learning

1. INTRODUCTION

The Center for Disease Control (CDC) in the US reports that on average 192 Americans die every day from an opioid (e.g., heroin, oxycontin, morphine) overdose (CDC, 2017). This is a multi-faceted problem of immense societal consequence that needs to be addressed in many ways, such as improving data quality and the monitoring of opioid use and relapse,

developing opioid use prevention strategies, providing support health providers, equipping first responders with the antidote naloxone to minimize overdose related deaths, and encouraging the public to take safe precautions with opioid use (CDC, 2017).

In this work, we focus on the issue of data quality and monitoring opioid use for people with opioid use disorders (OUD). Effective monitoring is crucial to the proper treatment of patients with OUD. Opioid use monitoring is typically done through surveys and patient evaluations, and blood/urine tests during office visits (Weiss, 2004). These approaches are inherently reactive in nature and can only detect drug use after the fact. One way of addressing this problem close to real-time is to use wearable biosensors.

Wearable biosensors are wearable devices that can measure a variety of physiological (e.g., blood volume pulse, and electrodennal activity) and movement (e.g., triaxial accelerometer) streams from the patient they are deployed on. These biosensors are of tremendous interest in the drug abuse treatment space for their potential to detect drug use in near real time (Carreiro et al., 2015)(Carreiro et al., 2016). They provide a significant advantage for behavioral interventions aimed at harm reduction or abstinence as they can alert medical practitioners of changes in physiological markers, practitioners can then intervene. Biosensors can not only be useful in tracking long-term reactions to drug withdrawal or therapeutic measures, but can also detect immediate harm to the patient such as an overdose.

Success in this type of clinical application for biosensors relies upon *patient's adherence*, that is the patient's wearing of the biosensor given to them. A major concern in this context, therefore, is that patients under surveillance can simply give the biosensor to another person to wear during the period of substance use to avoid opioid use (or relapse) detection[1]. In this paper, we present a machine learning-based detection approach for **collaborative non-adherence (CNA)**. We define CNA as the process of giving one's biosensor to someone else when surveillance is ongoing. The reason we focus on CNA is because patients with OUD are usually stigmatized in the society. Consequently, they have considerable incentive to keep any drug use or relapse from others, especially in the presence of any systematic surveillance of their opioid use (Olsen and Sharfstein, 2014). It is therefore imperative that as we work on developing new monitoring tools for the opioid epidemic, we also, at the same time, investigate ways to deal with deliberate non-adherence to the monitoring regime (Figure 1).

Note that currently there is no data on CNA in people with OUD to our knowledge. This is because the field of wearable biosensor-based interventions in this population itself is in early stages. However, given the propensity for individuals to try to "cheat" other wearable tracking devices such as step counters (Alshurafa et al., 2014) and the stigma associated with drug use, we view this as more of an eventuality than a theoretical concern. Further, since such behaviors would dramatically lower the success rates of biosensor-based interventions, we are proactively seeking solutions that can be build into interventions from the start.

---

[1]They could also simply take off the device, in which case their data-stream would flatline and would lead to an intervention by the caregiver, which we assume the patients want to avoid.

To develop and evaluate our detection approach, we rely on the *simulation of CNA* using biosensor data collected from overdosing patients in the University of Massachusetts medical center emergency department (ED). We used an Empatica E4 wrist-mounted biosensor (Empatica, Milan, Italy) for our data collection. Data was collected from 11 patients who presented to the ED for medical care following an opioid overdose. The patients were in various states of recovery subsequent to an administration of naloxone[2]. Our patients were real medical patients suffering from OUD, and all data was gathered in a way that prioritized patient care and wellness over research goals with approval from our Institutional Review Board (IRB) (ethics board).

In order to detect CNA, we use machine learning techniques to develop personalized classifiers for individual patients that capture the uniqueness of their cardiac physiology (blood volume pulse) and movement (accelerometer). This personalized classifier for a given patient in our dataset is trained using biosensor data from that patient when they are not intoxicated (also referred to as neutral state) and data belonging to the rest of the 10 patients in our dataset. This data from the other 10 patients can be any of three physiological dates: neutral, withdrawal, or intoxicated/overdosed. This way of building the classifier simulates the case where a patient gives their biosensor to someone else (a *collaborator*) of whom they have no control. This collaborator can therefore themselves decide to abuse drugs or abstain. The personalized classifier is able to detect the presence or absence of CNA because the patterns in the physiological and movement data obtained from a patient are unique to that patient (i.e., act as a biometric), irrespective of their health state. Note that, our chosen physiological signals have been used before as biometrics to identify a person uniquely (Sarkar et al., 2016)(Yang et al., 2015). However, none of these studies involved participants in various stages of drug use. To the best of our knowledge, this is the first work on developing an automated approach for CNA detection in the context of drug abuse. An analysis of our detection approach demonstrates its viability. Overall, the results show the efficacy of our personalized CNA detection approach. We achieved an average detection accuracy of 90.96% when the collaborator was one of the other 10 patients in our dataset, and 86.78% when the collaborator was from a set of 14 individuals, whose data has never been seen by our classifiers during training.

2 SYSTEM MODEL AND PROBLEM STATEMENT

In this paper we make several assumptions based on our eventual goal of identifying CNA. First, since the patient is compromising the source of the biosensor data, we assume that they have the ability to hand off the biosensor for any length of time. We also assume that the collaborator has unfettered access to the biosensor after it has been handed off. Second, we assume that the patient does not have the ability to mount other forms of non-adherence against the surveillance, such as replaying their own historical biosensor data. Finally, we assume that the user has no control over the collaborator once the device is handed off. This means that the collaborator could choose to use opioids themselves.

---

[2]Naloxone is an antidote that is given to someone who is overdosing on opioids. It immediately reverses the effect of opioids by competitively binding to the opioid receptors in the body.

In this context, the principal problem we address in this paper is to detect if the data received from a wearable biosensor assigned to person X is truly coming from person X (and not from person Y, where X ≠ Y); and that person X is not exhibiting opioid toxicity.

3 RELATED WORK

CNA During Physical Activity Monitoring:

In (Alshurafa et al., 2014) the authors present an approach to detect CNA in the use of fitness trackers to fool insurance companies who might want to give premium reductions for people who exercise. Their approach was specifically focused on the use of accelerometer data, collected from a wrist-worn fitness tracker and machine-learning, for detecting CNA. Further, their paper only focused on providing cross validation results for their models and did not test their models using data from previously unseen user data. Their accuracy in this work for cross validation was around 90%. In this work, we apply CNA detection to the domain of opioid abuse. Since the intake of opioid affects the physiology of the patient as much as their movement, we use both accelerometer and blood volume pulse from the patients as a way to identify CNA. Further, we evaluate our model using unseen test data, which shows the efficacy of our detection approach.

Wearable Biosensor Tracking of Opioid Use:

Some work has been done in identifying opioid use using a wearable biosensor (Carreiro et al., 2016)(Carreiro et al., 2015)(Smelson et al., 2014)(Chintha et al., 2018). However, these efforts have focused on detecting opioid ingestion using physiologic measurements made using wearable biosensors. They have not explored issues with patient adherence. In our work, we focus around the novel problem of collaborative non-adherence.

4 CNA DETECTION APPROACH

In this paper we are primarily interested in determining the feasibility of our non-adherence detection using the movement and cardiac signals in people who are drug users. Consequently, we decided to initially build and test our CNA detection approach by simulating CNA behavior using biosensor data collected from actual patients admitted to the University of Massachusetts medical center emergency department from drug overdose. Before we describe the details of the simulation (in Sections 5 and 6), we present the details of our data collection and the CNA detection approach.

4.1 Data Collection and Cleaning

The data used for building and evaluating our detection approach was collected from ED patients who received the opioid antagonist naloxone for known or suspected diagnosis of opioid toxicity. These patients were chosen as they are part of our target demographic. Once study staff obtained informed consent, the E4 was placed on the participant's non-dominant wrist and continuous biometric data was obtained until a predefined endpoint was reached (either discharge from or admission to the hospital).

Patients (i.e., study participants) were assessed approximately every few hours while enrolled in the study by a research staff member. Based on physical exam findings, they were noted to be in one of three states: neutral, opioid intoxication, or opioid withdrawal. The *neutral state* is defined as the patient being sober and awake. In the *intoxication state* the patient has exhibits signs and symptoms of intoxication or overdose. Finally, in the *withdrawal state*, the patient exhibits signs and symptoms of opioid withdrawal due to the administration of the opioid antagonist naloxone. The data from the patients in the intoxication and withdrawal states were collected while they were primarily lying still in a hospital bed. In the neutral state, the patients were more likely to stand/walk ad lib. This would probably include brief bouts of activity (like walking to the bathroom), as opposed to any prolonged or brisk activity. They also may be sitting up, having conversations with clinicians, or eating and drinking lightly.

We used the Empatica E4 (Empatica, Milan, Italy) wrist-worn biosensor for our data collection. The E4 is capable of monitoring a variety of physiological (e.g., blood volume pulse, heart rate variability, skin conductance) and movement (e.g., triaxial accelerometer) information. For this work we only use the triaxial accelerometer (sampled at 32 Hz) and blood volume pulse (sampled at 64 Hz) data from the E4 device. We collected accelerometer data because it has been used before in classification and authentication tasks similar to ours, and we believed it would be able to reasonably accurately identify our patients (Zheng et al., 2014) and (Alshurafa et al., 2014). Further, we used blood volume pulse (BVP) data because opioid use directly affects a person's breathing (Santiago and Edelman, 1985), which is reflected in the BVP signal.

Once the data is collected, the next step is to clean it and extract features from its data streams which can be used to identify the source of data. In order to account for noise in the BVP data, a low pass filter with finite-duration impulse response was utilized on the data. The filter utilized a pass-band frequency of 0.6 Hz. This was chosen because it aligns with a heartbeat of approximately 40 beats per minute (bpm), which is less than the range of 50–80 bpm expected for a heart at rest (Spodick, 1993). Similarly, the stop-band frequency was chosen at 3.33 Hz (aligning with approximately 200 bpm) based on estimates by Tanaka et. al (Tanaka et al., 2001). The decision was made to only apply the low pass filter to the BVP data and not to the accelerometer data to avoid accidentally removing variations in the accelerometer data that might be the results of physiological reactions such as shaking from withdrawal. Finally, we re-sampled the triaxial accelerometer data to 64 Hz to match the sampling rate of the BVP measurement.

Feature Extraction: Inspired by Alshurafa et al. (Alshurafa et al., 2014), our feature extraction process relies on statistical features extracted from the accelerometer and BVP data collected using the wearable biosensor. The features taken for accelerometer are: Mean, Median, Skewness, Variance, Standard Deviation, Mean Crossing Rate, Maximum Value, Minimum Value, Mean Derivatives, Inter-Quartile Range, Zero Crossing Rate, and Kurtosis. Those taken for BVP data are: number of peaks, mean of the peaks, standard deviation of the peaks, root mean square, distance from the maximum to the minimum of the peaks, mean peak-to-peak distance, standard deviation of peak-to-peak distance, and power spectral

density of the signal. Since there are 12 statistical features gathered from each of accelerometer axes and 8 statistical features from BVP measurement, there are a total of 44 features that are extracted. Identifying features to improve the accuracy of CNA detection is important future work.

4.2   Training and Detection

Once we have the dataset and know which features to extract, the next step is to build the CNA detection approach. Our detection approach uses a machine learning-based classifier to address our *principal question*: data received from a wearable biosensor assigned to patient X are actually coming from patient X and that patient X is not opioid intoxicated.

Our classifier learns the essential uniqueness of the accelerometer and BVP data of a patient's neutral state. We only consider the neutral state of the patient for two reasons: (1) it may be impossible or impractical to gather data from the patient during training for intoxication or withdrawal state in the real world, and (2) since the purpose of our system is to help introduce monitoring of the patient for their physical well-being, we think that showing that they are in the neutral state is necessary to meet that goal. As long as any newly received accelerometer and BVP snippet match the understanding of a patient's neutral state, the classifier can be sure there is no CNA. Our detection approach has to two phases: the training phase and the detection phase (see Figure 2).

Training Phase: In the training phase, we build a classifier to identify the uniqueness of a patient in their neutral state so that it can later be used to identify whether the received accelerometer and BVP measurements are also representations of the same patient's neutral state. We build a personalized (binary) classifier for each patient in our dataset, whose two class points are obtained as follows. **(1) Positive class points:** During training, for a given patient, we take that *patient's neutral state data*, and divide it up into windows of length w time-units. We then extract our 44 features from each of these windows. That is, we produce a 44-dimensional feature point per window. We call the feature points as the positive class points and these signify no CNA. **(2) Negative class points:** We then divide the *neutral, intoxicated, and withdrawal state data from all remaining 10 patients* in our dataset into w-size windows, extract the 44 features from each of these windows, and call them the negative class points. Since we have no real instances of CNA, these negative class points are used to simulate CNA in our training set. Once we have the positive and negative class points, we take a subset of these feature points to train a classifier which forms our CNA detection approach and use the rest for validation purposes.

Detection Phase: Once the classifier is trained for a patient, it is now able to classify whether an unseen feature point derived from a *w-sized* snippet of accelerometer and BVP measurements came from that patient, or whether it is an instance of CNA. When an unseen data point is evaluated on our classifier, it returns a confidence value from 0 to 1 with 1 indicating that the model has full confidence that the unseen snippet belongs to the patient in the neutral state, and with 0 indicating full confidence that the point does not belong to that patient's neutral state and hence signifies CNA in our setup. We are then able to decide

whether to accept or reject that data point depending on whether its confidence value meets a chosen threshold.

5 EVALUATION METHODOLOGY

Given that patients in our dataset were examined by clinicians intermittently (every hour), we do not have the ground-truth about the patient's health state at all times. Consequently, we *curate* the biosensor data we collected from the 11 patients, to extract a dataset that captures the blood volume pulse and the accelerometer readings from the patients when we are reasonably confident of their health states (i.e., neutral, intoxicated or withdrawal). This *curated dataset* can then be used for training our detection classifiers and evaluating their efficacy. In this section, we describe our dataset curation process, and its use for training and evaluating any CNA detection, followed by a brief description of our evaluation metrics.

5.1 Dataset Curation

Overall, we have data collected using the wearable biosensor from 11 patients who were brought to the ED as a result of opioid overdose for several hours. Patient data was gathered in accordance with hospital policy and with approval through our institution's Institutional Review Board (IRB) approval process (ethics committee review). During this time, patients were experiencing different health states (i.e., neutral, intoxicated or withdrawal) depending upon when they were administered naloxone. We extracted wearable biosensor data for each patient from the various health states. This was done by extracting 20 minutes of biosensor data immediately surrounding a patient evaluation by the research/clinical staff, which resulted in the recording of their health state at the time. This translates to 10 minutes before and 10 minutes after the evaluation. Since the patient is in a hospital setting, we can have reasonable assurance that during these 20-minutes there would not be a significant change in this health state (see Figure 3).

Note that, depending upon when they were brought into the ED after their overdose, when naloxone was administered, and when evaluations were conducted, some patients may only have data collected in their neutral state, others may have intoxication and neutral state data, or withdrawal and neutral state data. All patients in our dataset have some neutral health state data. Once we have extracted the pertinent data from the patients, we have the necessary dataset for answering our principal question of whether CNA can be detected.

Data Used for Training: To be able to train the classifier to detect CNA, we have to compensate for the idiosyncrasies in our curated dataset that originated from our data collection protocol. For instance, during the data collection, we were not able to carefully control the patients' actions (especially in the neutral state). Further, given the way we extracted data for different health states for a patient, we do not have enough contiguous data to train our classifier in any one health state. Therefore, in order to train our classifier, we first generate the positive and negative class points (as described in Section 4.2) and then shuffle them. We then use the first 8 minutes' worth of feature points for training. Since a feature point is produced from w-sized windows, 8 minutes of data will have $8/w$ number of feature points. The rest of the shuffled feature points are used as testing (i.e., CNA

detection). We chose the value of 8 minutes because every patient had at least 10 minutes of neutral state data. This allowed us to train our classifier for every patient and still have some data left over to test for non-CNA scenarios. After the data is shuffled and training data is chosen, all of the resulting negative and positive samples are fed into a classifier (more details on the classifier is given in Section 6) to train it. The bottom part of Figure 3 shows the data curation for training. As mentioned before, this way of building the classifier simulates scenarios where a patient gives their biosensor to someone else (a *collaborator)* of whom they have no control. This collaborator can themselves decide to abuse drugs or abstain.

Data Used for Evaluating CNA Detection: Once our classifier is chosen and trained, the next step is to see how well it performs in detecting the presence and absence of CNA. We do this by feeding the patient-specific classifiers in our CNA detection approach: *w*-second snippets of completely unseen triaxial accelerometer and BVP data. These simulate the CNA detection approach determining whether the patient data collected every *w*-seconds (3 seconds) is coming from the patient in their neutral state or not. Each previously unseen *w*-second snippet produces a feature point which is evaluated by the classifier. These unseen test feature points come from: (1) **the validation set**, portions of the data of the 11 patients in our dataset that were not used in training, and (2) **the external set** of 14 healthy/non-opioid using individuals, whose data has never been seen before in any form by any of our classifiers. Out of those 14 individuals, 9 were at rest, and 5 were moving around in a busy conference setting (see Table 1). The bottom right of Figure 3 shows the data curation for performing evaluation.

5.2. Metrics

In order to evaluate the efficacy of our system, we use the following five core metrics. (1) *True Reject Rate (TRR):* the rate at which true negative feature points (data points from patients other than the one that the system is trained for) is rejected by the detection approach. This indicates the presence of CNA. (2) *False Accept Rate (FAR):* the rate at which true negative feature points are accepted by the detection approach. This indicates a false alarm. (3) *True Accept Rate (TAR):* the rate at which positive feature points (the patient's own unseen data) is accepted by the detection approach. This indicates the absence of CNA. (4) *False Reject Rate (FRR):* the rate at which an unseen positive feature points (the patient's own data) is rejected from the detection approach. This indicates when CNA was missed. (5) *Equal Error Rate (EER):* The rate at which the FAR and the FRR are equal. This is the point at which our detection approach balances the accuracy of its detection with usability.

6 CLASSIFIER TRAINING

In this section, we describe how we identify the classifier of choice for our detection approach along with how we choose the window size (*w*), given the classifier of choice.

Choosing the Classifier:

In order to choose a machine learning classifier that would be effective, we used 5-fold stratified cross validation on our training data. Since there are far more negative class points than positive class points, we use weighted classifiers in order to avoid the model favoring the negative state. The weighted classifiers consider the correct minority (positive) class decision much more valuable than a correct majority (negative) class decision. We used a value of 1:10 for the ratio of the weighting as there are 11 patients in our data set, and so for every 1 positive point from a particular patient, there are approximately 10 points belonging to other patients. The weighted algorithms used were weighted Random Forest, weighted Adaboost, weighted Support Vector Machine (SVM) with a radial-bias function (rbf) kernel, and weighted logistic regression. Our cross validation was performed with a 5-second window, which was initially chosen arbitrarily, as we tune our window size after the algorithm is chosen.

The summary of our evaluation in the form of receiver operating characteristic (ROC) curves is seen in Figure 4. ROC curves plot TAR against FAR in order to form a curve. The closer the area under the curve (AUC) is to 1, the better the algorithm performed in cross validation. These curves are averaged overall 11 patients. In addition, for evaluation we looked at the equal error rate (EER) which is the point at which FAR and FRR are equal (in other words the place with the lowest combined error). Algorithms with a lower EER were considered preferable to those with a higher EER if the AUC was the same. As can be seen in Figure 4, the best performing algorithm is weighted Random Forest, which we chose as our machine learning classifier.

Choosing Window Size (*w*):

Once the machine learning algorithm is chosen, the next step is to find the appropriate window size. Once again, we run with 5-fold cross validation on our training data where the data was split by windows of varying size. As can be seen in Figure 5, which shows the average ROC curves for various window sizes, *w*=3 seconds had the highest average AUC of 0.94 and the lowest EER at 12.0%. Therefore, we chose 3 seconds as the window size. Contrary to expectation, here, a smaller window size gives a better result. We hypothesize that the smaller window is more reliable at capturing the sudden variability in the blood volume pulse that appear from opioid ingestion and naloxone use. We plan to investigate this further in the future.

7 CNA DETECTION EVALUATION

In this section we present the results of our classifiers using the validation set and external set to show the efficacy of our CNA detection approach.

Validation Set Results:

For the validation set, we achieved promising performance from our CNA detection approach. Figure 6 shows the ROC curve of the 11 patient-specific models. The average AUC on the validation set comes to 0.94. The EER is of course different for each patient-

specific detection approach. The average EER for all patients is 9.04% (i.e. average accuracy of 90.96%). The performance variability for different patients can be largely accounted for by the difference in level of activity between different patients. The ROC results show how the classifier's performance varies for individual patients, but overall is quite accurate for the validation set.

External Set Results:

In addition to testing the performance of our classifiers on our validation set, we performed additional testing on an *external set.* The data in the external set has never been seen by our detection approach classifiers in any form before. These results show the generalizability of our data-driven CNA detection approach. This data in the external set was collected from healthy individuals while they were attending a conference. This *simulates* the case, where a patient gives their watch to another person (i.e., engages in CNA) who is healthy, does not abuse opioids, and is completely unknown to the surveillance system. Since all the data in the external set comes from someone unknown, ideally, all the feature points generated from this data should be rejected (i.e., classified as negative). Figure 7 shows the ROC curve for the external set. Here, we use the patient's own test data to form the TAR for our approach. As expected, the average AUC is slightly lower than the validation set at 0.92, with a slightly higher average EER at 13.22% (i.e. average accuracy of 86.78%). These results show that we are overfitting our detection models.

8 LIMITATIONS

In this preliminary study, we showed the promise of our system for detecting CNA in an opioid surveillance program that relies on a wearable biosensor. Even though our system focuses on collaborative non-adherence during opioid abuse surveillance, it can also detect accidental non-adherence as well. This means that if someone else puts on a patient's biosensor by mistake or if the biosensor fell off or was recording incorrect data, it would be detected as well because the received data would not match the data expected by our classifier for the patient.

However, there are *two* major limitations to our current work that we plan to address in the future. First, in this work we view the neutral state as a monolithic state. We were able to do it largely because the data was collected in the ED, where the patient was not liable to be doing many different things in the neutral state. However, that is not true outside the ED as the patient may be performing a variety of actions when they are in the neutral state. Our classifiers therefore by necessity have to be aware of the various activities the patient is engaged in and differentiate these activities from one patient to another. This will result in detection models that are much more complicated. Further, this will reduce false alarms.

Second, we have only simulation of CNA in this work. However, to be truly effective these classifiers have to detect CNA where patients actively try put the sensor on a different person. To be able to achieve this we need to create one or more clinical studies that allow patients to intermittently put the sensor on another person and then record (e.g., using a survey app) the fact that there was collaborative non-adherence. This will allow us to train

classifiers and check to see if they are able to detect CNA in the real world. Further, this non-adherence study has to include a larger participant pool than what we used in this paper which is somewhat limited.

9 CONCLUSION

In this paper we presented an approach for CNA detection for a patient who is being surveilled, using a wrist-worn wearable medical device, for opioid abuse. In the future, we plan to immediately expand on this work in several directions: (1) undertake clinical studies that create actual CNA scenarios to fine-tune our models including activities outside of a hospital setting, (2) identify improved features and machine learning classifiers to increase the detection accuracy, and (3) build a cloud-based application that can detect the presence of CNA at scale. In addition, we have the long-term goal of exploring applications of our CNA detection approach to other non-adherence scenarios such as pre-exposure HIV prophylaxis medications, using wearable biosensors.

REFERENCES

Alshurafa N, Eastwood J-A, Pourhomayoun M, Nyamathi S, Bao L, Mortazavi B, and Sarrafzadeh M (2014). Anti-cheating: Detecting self-inflicted and impersonator cheaters for remote health monitoring systems with wearable sensors. In BSN, volume 14, pages 1–6.

Carreiro S, Fang H, Zhang J, Wittbold K, Weng S, Mullins R, Smelson D, and Boyer EW (2015). imstrong: deployment of a biosensor system to detect cocaine use. Journal of medical systems, 39(12): 186. [PubMed: 26490144]

Carreiro S, Wittbold K, Indie R, Fang H, Zhang J, and Boyer EW (2016). Wearable biosensors to detect physiologic change during opioid use. Journal of medical toxicology, 12(3):255–262. [PubMed: 27334894]

CDC (2017). Understanding the epidemic. CDC website

Chintha KK, Indie P, Chapman B, Boyer EW, and Carreiro S (2018). Wearable biosensors to evaluate recurrent opioid toxicity after naloxone administration: a hilbert transform approach

Olsen Y and Sharfstein JM (2014). Confronting the stigma of opioid use disorder—and its treatment. Jama, 311(14): 1393–1394. [PubMed: 24577059]

Santiago TV and Edelman N (1985). Opioids and breathing. Journal of Applied Physiology, 59(6): 1675–1685. [PubMed: 2934364]

Sarkar A, Abbott AL, and Doerzaph Z (2016). Biometric authentication using photoplethysmography signals. In Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on, pages 1–7. IEEE.

Smelson DA, Ranney ML, Boudreaux ED, and Boyer EW (2014). Real-time mobile detection of drug use with wearable biosensors: A pilot study

Spodick DH (1993). Survey of selected cardiologists for an operational definition of normal sinus heart rate. The American journal of cardiology, 72(5):487–488. [PubMed: 8352202]

Tanaka H, Monahan KD, and Seals DR (2001). Age-predicted maximal heart rate revisited. Journal of the American College of Cardiology, 37(1):153–156. [PubMed: 11153730]

Weiss RD (2004). Adherence to pharmacotherapy in patients with alcohol and opioid dependence. Addiction, 99(11):1382–1392. [PubMed: 15500591]

Yang J, Li Y, and Xie M (2015). Motionauth: Motion-based authentication for wrist worn smart devices. In Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on, pages 550–555. IEEE.

Zheng Y-L, Ding X-R, Poon CCY, Lo BPL, Zhang H, Zhou X-L, Yang G-Z, Zhao N, and Zhang Y-T (2014). Unobtrusive sensing and wearable devices for health informatics. IEEE Transactions on Biomedical Engineering, 61(5):1538–1554. [PubMed: 24759283]
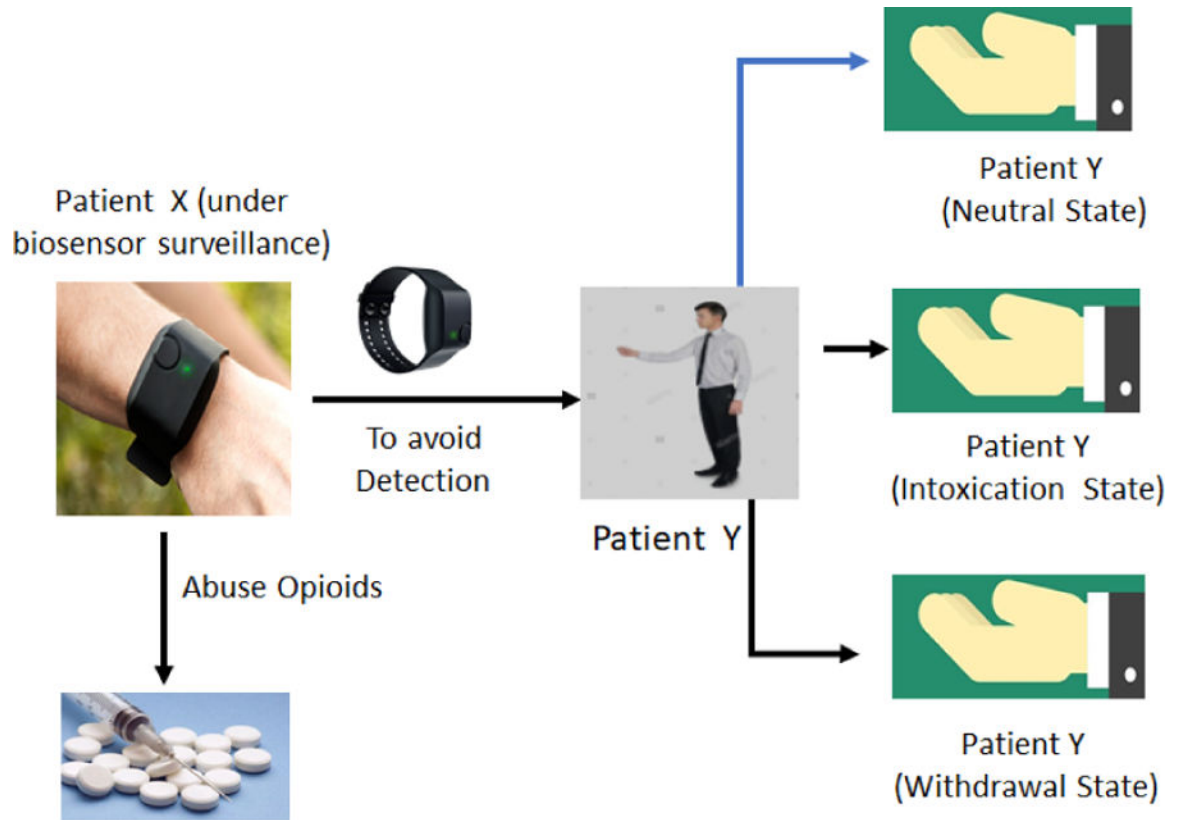
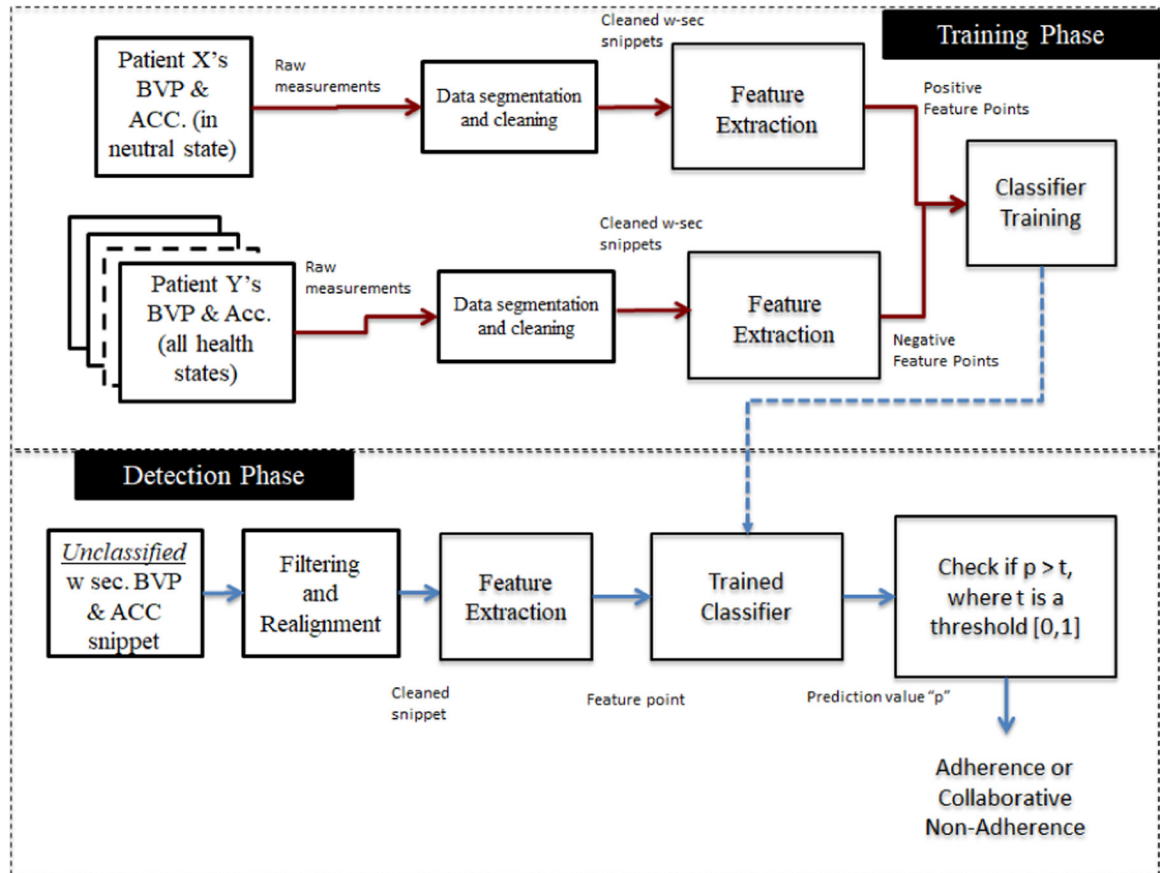**Figure 1:**
Collaborative Non-Adherence (CNA).

Figure 2:
Overview of collaborative non-adherence (CAN) detection approach.

Figure 3:
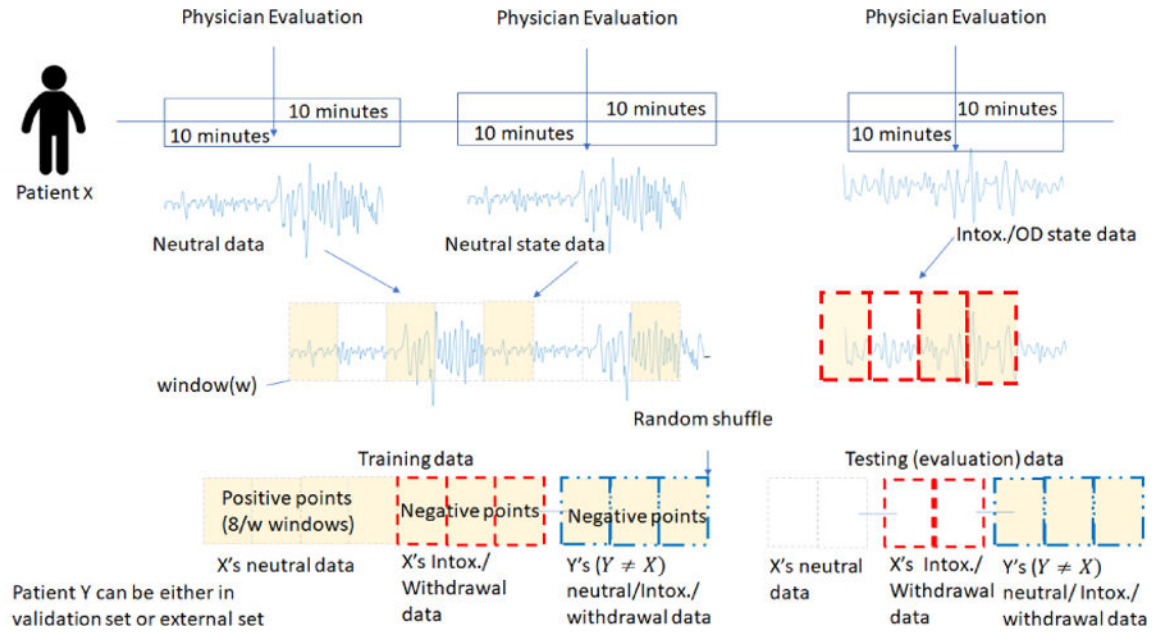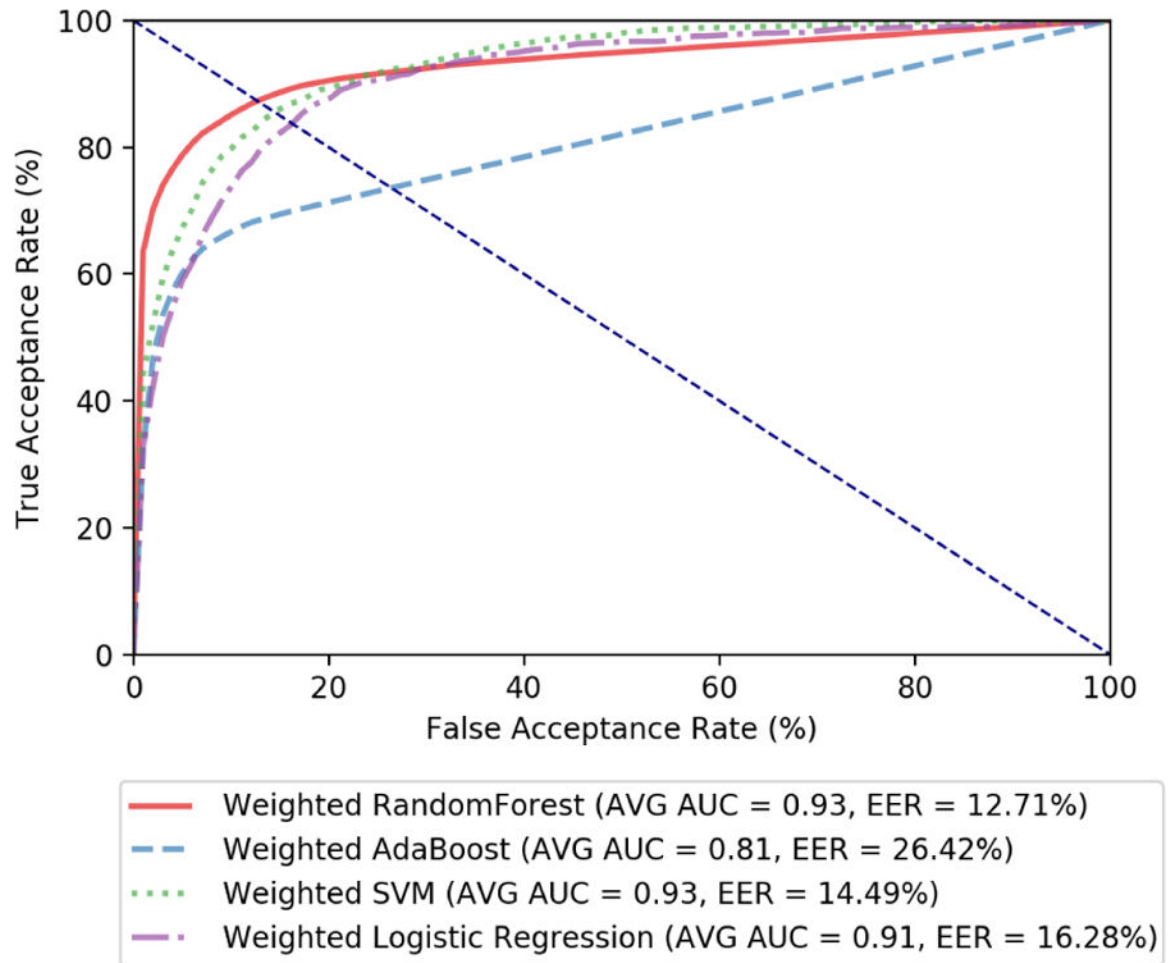Overview of our data curation process.

Figure 4:
ROC curves for 5-fold cross validation on various machine learning algorithms. Here, we set the window size (*w*) to 5 seconds, provisionally.
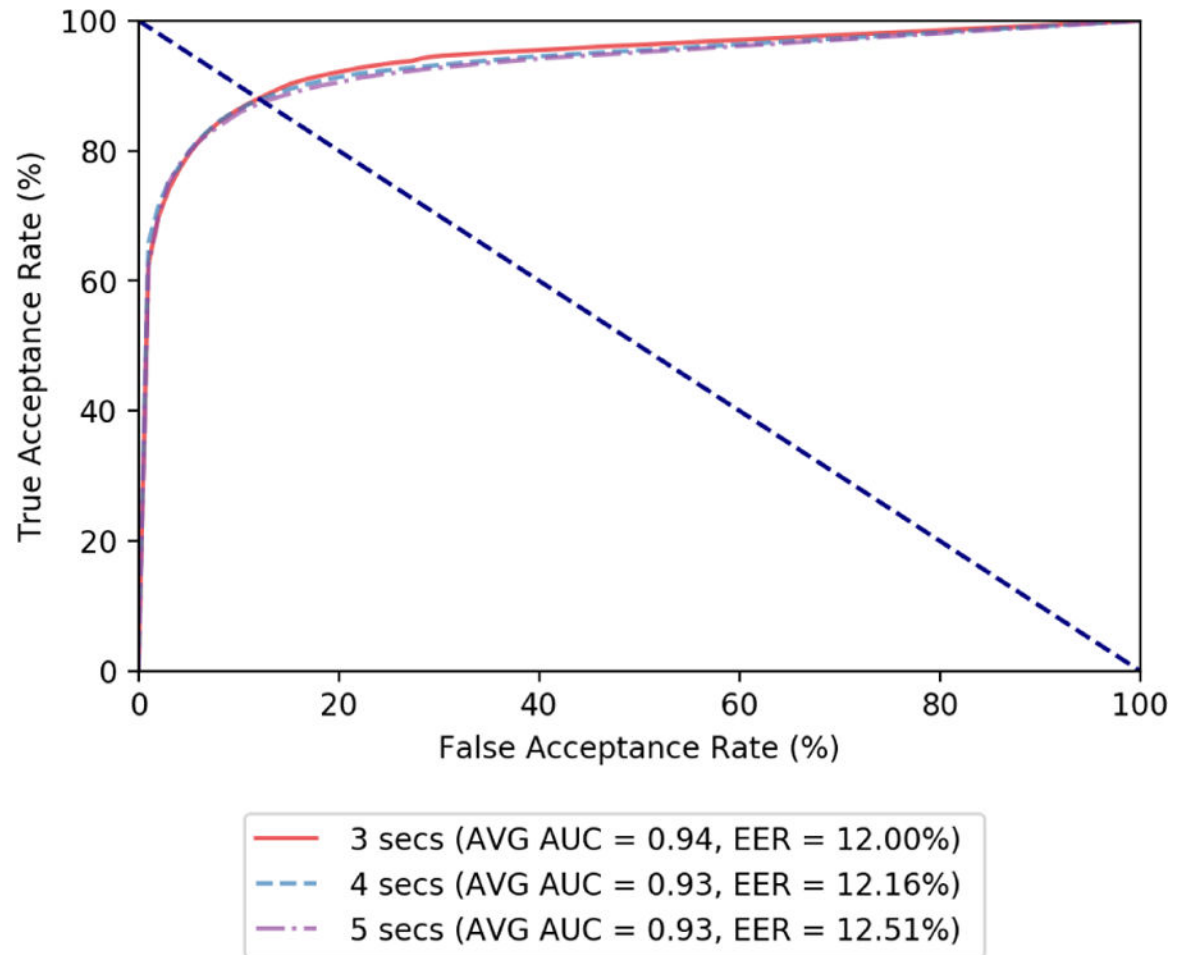
Figure 5:
ROC curves for 5-fold cross validation on penalized Random Forest with various window sizes. A window size (w) of 3 seconds was chosen based on these results.
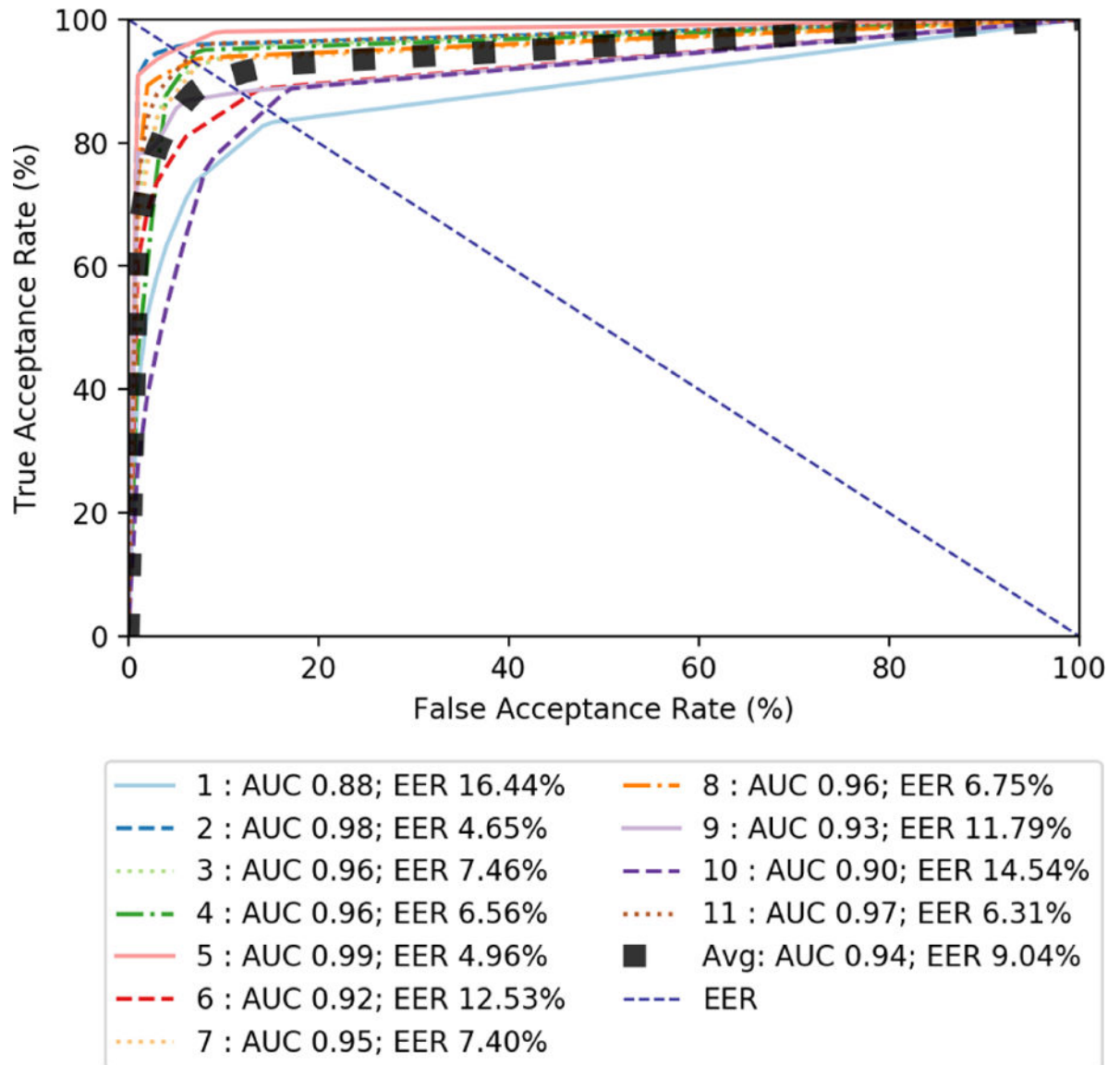
Figure 6:
Graph of ROC curves for validation set test features using weighted random forest with a window of 3 seconds.
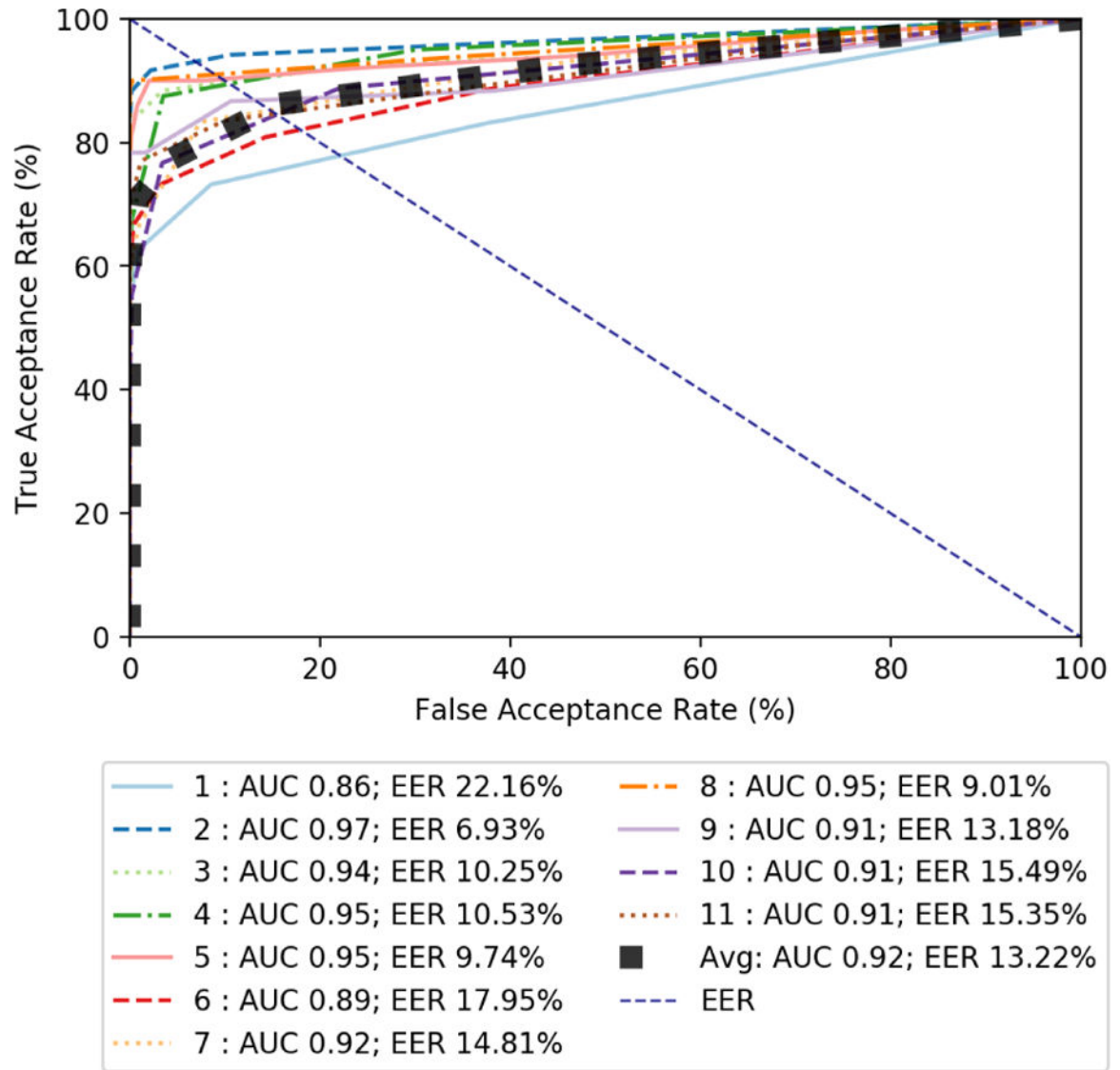
Figure 7:
Graph of ROC curves for validation set test features using weighted random forest with a window of 3 seconds.

Table 1:

Demographics from both data sets used for our evaluation.

Set	Count	Avg. Age (std.)	♂	♀
Validation	11	37.09 ± 10	9	2
External	14	32.8 ± 5.3	10	4