An Approach for Deliberate Non-Compliance Detection during Opioid Abuse Surveillance by a Wearable Biosensor

by

Rohitpal Singh

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Data Science

by

July 2018

APPROVED:

Assistant Professor Krishna Kumar Venkatasubramanian, Thesis Advisor

---

Assistant Professor Randy Clinton Paffenroth, Thesis Reader

Professor Elke A. Rundensteiner, Director of Data Science

Abstract

Wearable sensors can be used to monitor opioid use and other key behaviors of interest, and to prompt interventions that promote behavioral change. The effectiveness of such systems is threatened by the potential of a subject's deliberate non-compliance (DNC) to the monitoring. We define deliberate non-compliance as the process of giving one's device to someone else when surveillance is on-going. The principal aim of this thesis is to develop an approach to leverage movement and cardiac features from a wearable sensor to detect such deliberate non-compliance by individuals under surveillance for opioid use. Data from 11 participants who presented to the Emergency Department following an opioid overdose was analyzed. Using a personalized machine learning classifier (model), we evaluated if a snippet of blood volume pulse (BVP) and accelerometer data received is coming from the expected participant or an alternate person. Analysis of our classifier shows the viability of this approach, as we were able to detect DNC (or compliance) with over 90% accuracy within 3 seconds of its occurrence.

Acknowledgements

Thank you, Professor Krishna Kumar Venkatasubramanian, for your guidance throughout the research and documentation of my thesis work. I came to WPI with industrial experience in software engineering and basics of machine learning concepts. It was his direction that leads to the successful execution of my research in the data science field. He taught me how to showcase results to end users at each phase of the research. Thank you for your part in getting me to where I am now.

Thank you, Professor Randy Clinton Paffenroth, for providing alternatives to solve a problem associated with my thesis work. Thank you, Matthew Liam, for taking out time from your busy schedule to help me and for invaluable discussions and suggestions to deal with the noisy data. Thank you, Brittany Faith Gradel, for providing a helping hand during the last phase of my thesis. I appreciate her effort for getting my thesis work done early.

Thank you, Abhilash Hegde, Sayan Saha, Vinit Tougani, and Siddhi Khamitkar for listening crazy technical terms and for motivating me throughout the thesis work. I thank for their willingness to participate in intellectual discussions and to give the right perception to deal with problems pertaining not only to academia but to outside world as well.

Thank you, parents, for supporting me morally and financially. Thank you, for keeping faith in me and believing in my decisions. Thank you, sister, Deepika Singh, for making me smile even when nothing seems perfect. Thank you, WPI, for giving me the opportunity to pursue masters in data science.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Smart wearable devices can monitor patients' vital signs and other indicators based on embedded sensors to facilitate diagnosis, low-cost medical treatment, and more effective disease/patient management [1] [2]. The use of wearable devices allows continuous monitoring of patients' health which can help to differentiate their health patterns in normal and abnormal conditions [3]. Wearable sensors include physiological, biochemical and motion sensors. Physiological sensors could help in both diagnosis and ongoing treatment of neurological, cardiovascular and pulmonary diseases. Motion sensors are the compliment to physiological data that can be used to identify subject's specific motion activity.

Clinical studies data that relies on remote monitoring system helps in predicting, diagnosis and ongoing treatment of critical diseases. However, subjects can deliberately hand over the dedicated device to others to avoid drug abuse detection or to gain financial incentives; e.g., Fitbit wearers log extra steps by attaching the device to a dog or spoke of a car and want to gain payouts for false claims to which they are not entitled. Data acquisition from impostors leads to inappropriate medications and decisions.

Wearable sensors are of tremendous interest in the drug abuse treatment space for their potential to detect drug use in real time [4] [5]. This would provide a significant advantage for behavioral interventions aimed at harm reduction or abstinence. Success in this type of clinical application for wearables would rely on subject's compliance, and a major concern is that individuals under surveillance can simply put the sensor on another person during the period of substance use to avoid detection. On the other hand, other persons who are not the legitimate subjects could be in one of 3 possible states called intoxication, withdrawal or neutral as shown in Figure 1.1. Holding this assumption, we present a machine learning-based approach to detect such deliberate non-compliance (DNC) to the monitoring with a wearable sensor.

In our work, we used blood volume pulse (BVP) and triaxial accelerometer data to capture cardiac physiology and movement information respectively. Movement information is helpful to distinguish an individual from others under different motion activities as each subject exhibits a unique pattern. However, during opioid abuse

subjects have irregular movements and therefore, movement information is less informative. Furthermore, opioid abuse directly affects one's breathing and that can be captured by BVP. Therefore, BVP and triaxial accelerometer both are useful for detecting DNC in case of opioid abuse.

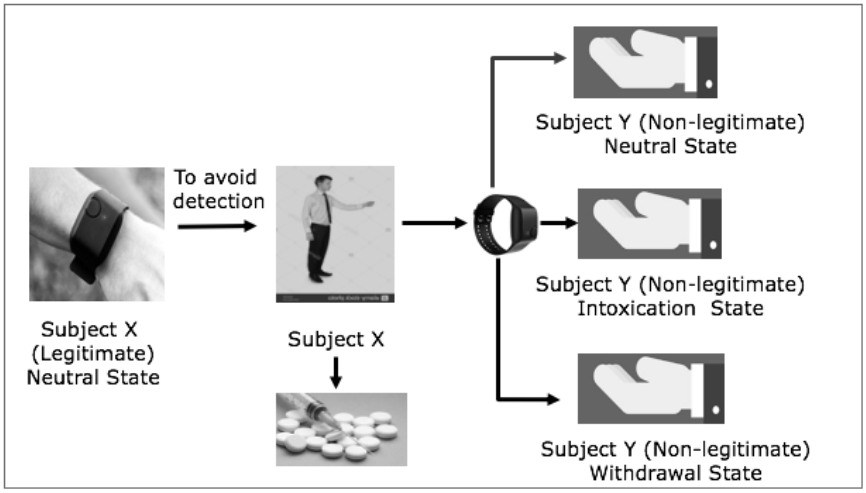


Figure 1.1: Overview of Deliberate Non-Compliance. Images from [6] [7] [8].

We define deliberate non-compliance as the process of giving one's device to someone else when surveillance is on-going. A subject can deliberately hand over the smart device to others during the opioid abuse and thus don't comply with the monitoring system. Our detection approach is based on pretrained personalized machine learning model that can detect continuously whether data is coming from the same subject or not. The personalized machine learning model consists of samples from two classes of positive (genuine subject) and negative (other subjects). The positive class consists of samples only from the neutral state of a subject as expected by a monitoring system. But, subjects falling in the negative class could be in any one of three states intoxication, withdrawal or neutral. Each sample has 36-time domain features of an accelerometer and 8-time domain features of BVP based on a fixed window of size w. Window size and machine learning algorithm are chosen by experimentation as described in the experimental section.

In this work, we build personalized models for individual participants that capture the uniqueness of their cardiac physiology (through blood volume pulse (BVP)) and movement (through triaxial accelerometry). This is done by first using a wearable biosensor to monitor a person when they are in a neutral (non-intoxicated) state of health. This information is then used to build a classifier (model) for the individual, which forms the basis of DNC detection. If the subject decides to not comply and have someone else wear the device, physiological and movement signal patterns that are different from what the model expects will be identified, implying that someone other than the intended subject is wearing the device.

Personalized modeling leads to imbalanced class distributions. There are many approaches to deal with imbalanced distributions. Few of them are Synthetic Minority Oversampling Technique (SMOTE) [9] and Adaptive Synthetic (ADASYN) [10] sampling method that oversamples minority class. In our case, SMOTE didn't perform well therefore, we chose weighted machine learning algorithms. Weighted machine learning algorithms assign more weight to minority class samples for being misclassified by the model and this is how they train from imbalanced distributed data.

This thesis presents preliminary results of our approach based on a dataset collected using the E4 wrist-mounted biosensor (Empatica, Milan, Italy) [11]. Data were collected from 11 subjects who presented to a single Emergency Department (ED) for medical care following an opioid overdose. We simulated DNC scenarios based on this dataset by presenting data collected from a participant other than whom the personalized model was created for. Analysis of our classifier demonstrates the viability of this approach, as we were able to detect cheating and non-cheating with over 90% accuracy within 3 seconds of its occurrence.

1.1 Related Work

To the best of our knowledge, there are two other works in the literature related to DNC detection. First is [12] where authors present an approach to detect DNC 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, only for detecting DNC. However, the main drawback of their approach is that accelerometer data is helpful when subjects are moving but less informative in case of opioid abuse and subject's static. Other drawback is that their paper only focused on the providing cross-validation results for their models and did not test their models using data from previously unseen subjects. Cross-validation results are representative of test accuracy, but the true unknown value of test accuracy can differ significantly if models are overfitted.

Other work in this field identifies DNC in an ongoing manner. The author in [13] proposed an activity-aware ECG based patient authentication for remote health monitoring in an ongoing manner. DNC detection in their work is continuous. There are two drawbacks of their work. First, they implemented a mobile prototype portable device using wearable ECG sensors to record accelerometer and ECG data. It shows the lack of a smart device having an ECG sensor. Second, only two features mean, and standard deviation were used for each axis of accelerometer that might not provide the true characteristic of subjects in response to activities.

In our work, we apply DNC detection to the domain of opioid abuse. Since the intake of opioid affects the physiology of the subject as much as their movement, we use both accelerometer and blood volume pulse (BVP) from the subjects to identify DNC. Further, we evaluate our model using unseen test data, which shows the efficacy of our detection system. Our work includes 12 statistical features for each axis of an accelerometer and 8-time domain features of BVP.

1.2 Problem Statement

In this section, we present the principal assumptions about our work. We assume that surveillance is being done using a wearable biosensor which the individual wears always during the treatment period. In this context, we attempt to address the problem of detecting deliberate non-compliance, where a person under surveillance for opioid use gives their wearable biosensor to someone else. More precisely, 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 person X is not exhibiting opioid toxicity).

1.3 Background

The validity of using physiological signals in biometric recognition has been demonstrated in various related fields. Prior research work [14] is focused on identification and authorization of an individual from remote place using ECG and EEG sensors. Emotion detection paper [15] used BVP, EDA, EEG and ECG biometrics traits.

Physiological signals are highly correlated with activities of the subjects. Different activities, which are performed by subjects, induce variations in their physiological signals. Further, we use blood volume pulse (BVP) data because opioid use directly affects a person's breathing [16], which is reflected in the BVP signal. Furthermore, BVP has not been used to best of our knowledge in studying opioid abused subjects.

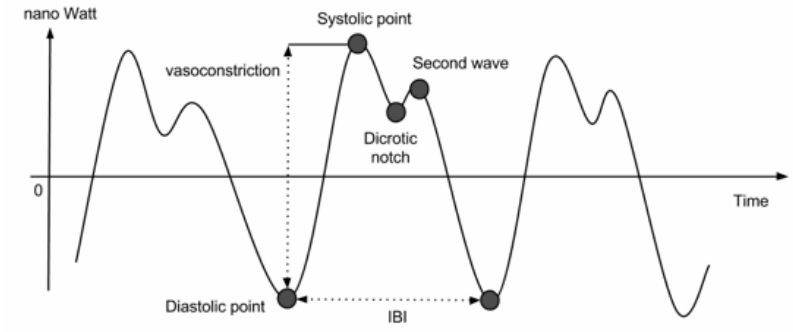


Figure 1.2: Snippet of BVP data.

BVP uses infra-red light to measure the change in the amount of blood present in the skin. This amount is maximum at each heartbeat and decreases between two successive heartbeats as shown in Figure 1.2 from [11]. BVP can also derive heart rate and heart rate variability. BVP efficiently distinguishes an individual under various conditions. Figure 1.2 is a typical condition but it may vary depending on subjects and conditions. The BVP signal is a relative measure so it does not have a standard unit.

BVP signal is characterized by four terminologies (1) Diastolic points, (2) Systolic points, (3) Dicrotic notch and (4) Dicrotic.

Diastolic points: Diastolic points are local minima's of the BVP signal and these are used to compute the Inter Beat Interval (IBI).

Systolic points: Systolic points are local maxima's of the BVP signal and could be used in conjunction with the Diastolic point to estimate the vasoconstriction of the subject.

Dicrotic notch: Dicrotic notch under some conditions (i.e. when subject is static) we observe the presence of this point which can be used to study different types of cardiac diseases.

Dicrotic: Dicrotic is an effect of Dicrotic notch.

An accelerometer sensor is a motion sensor that measures linear acceleration on vibration. Research in human activities recognition uses triaxial accelerometer sensors [17] [18] [19] [20]. An accelerometer sensor is helpful to detect patterns associated with each subject's motion activities. In our work, we used accelerometer along with BVP data by capitalizing on the importance of BVP during opioid abuse.

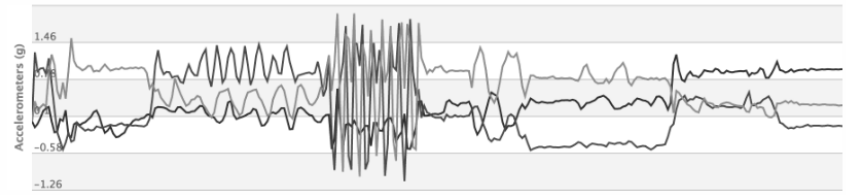


Figure 1.3: Snippet of triaxial accelerometer data.

A snippet of triaxial accelerometer data from Empatica device E4 is shown in Figure 1.3. The accelerometer is configured to measure acceleration in the range [-2g, 2g] where g is the gravitational force applied to each of the three axis of an accelerometer.

In our work, instead of considering BVP and triaxial accelerometer signals as time series for modelling, we considered observations as fixed set of features extracted from a fixed window size. Given these observations, we created personalized model (classifier) for each subject.

1.4 Thesis Organization

The rest of the thesis work is divided into different sections. Chapter 2 describes methodologies used that include data collection, data cleaning, features extraction, and overview of DNC architecture. Chapter 3 describes experiment setup used for data curation and for choosing metrices, best window size, and machine learning algorithm. Chapter 4 describes the result on training, validation and external data sets. Chapter 5 is conclusion and Chapter 6 covers all future work that we will focus on to scale DNC detection approach.

Chapter 2

Methodology

Our approach is to detect deliberate non-compliance to biosensor-based surveillance by developing a machine learning-based classifier utilizing triaxial accelerometer and BVP data gathered from the wearable biosensor. This involves five steps: (1) collection of subject data, (2) data cleaning, (3) feature extraction, (4) training, and (5) detection.

2.1 Data Collection

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. Once study staff obtained informed consent, the E4 was placed on the participants non-dominant wrist and continuous biometric data was obtained until a predefined endpoint was reached (either discharge from or admission to the hospital).

Subjects (i.e., study participants) were assessed approximately every hour 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 subject being sober and awake. In the intoxication state the subject has exhibits signs and symptoms of opioid toxicity or overdose. Finally, in the withdrawal state, the subject exhibits signs and symptoms of withdrawal due to the administration of the opioid antagonist Naloxone. The data from the subjects in the intoxication and withdrawal state were collected while they were primarily lying still in a hospital bed. In the neutral state, the subjects were more likely to stand/walk. 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 [11] wrist-worn biosensor for our data collection. The E4 can monitor a variety of physiological (e.g., blood volume pulse, heart rate variable, 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 collect accelerometer data

because it has been used before in classification and authentication tasks like ours, and we believed it would be able to reasonably accurately identify our subjects [2] [12]. Further, we use blood volume pulse (BVP) data because opioid use directly affects a persons breathing [16], which is reflected in the BVP signal.

2.2 Data Cleaning

Once the data is collected using the medical wearable biosensor, the next step is to clean it and extract features from its data streams which can be used to identify the source of data.

There are many motion artifact removal techniques that have been used by researchers. Every technique has some pros and cons and is applicable in limited settings. We did experiments using the approach mentioned in [21]. The approach in [21] can effectively estimate heart rate from a foot worn photoplethysmogram (PPG) sensor (PPG measures changes in volume of blood at each heart beat similarly like BVP) during fast bike exercise. We assumed that the subjects are doing different range of activities in neutral state and thus employed the approach in [21].

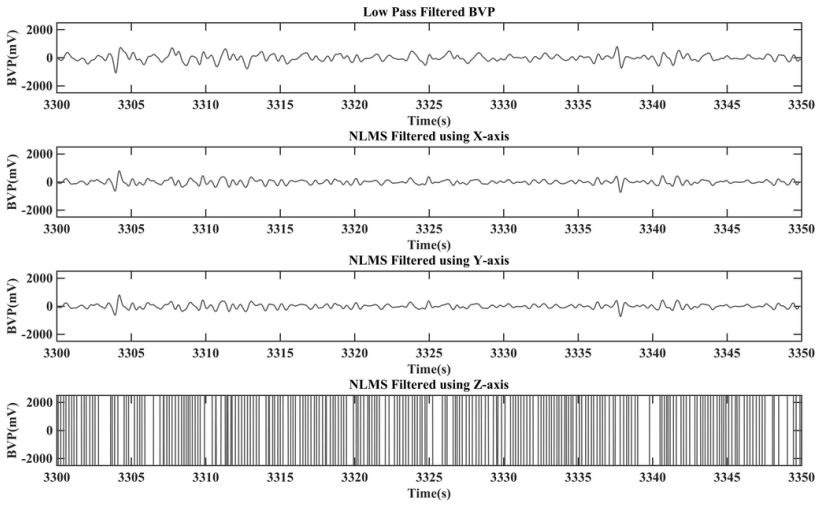


Figure 2.1: A snippet of BVP data for a particular subject. Comparison of Low pass filtered BVP with NLMS adaptive filtered output for each axis of 3d accelerometer.

According to the work in [21], the accelerometer data is re-sampled at 64Hz to align with BVP sampling rate. The BVP and re-sampled accelerometer signals are first low pass filtered with finite-duration impulse response (FIR) to remove high frequency noise. Adapted FIR filter causes delay in output response. Therefore, to compensate

for this delay we shifted the filtered signal to line up the data. It works by removing the first delay samples from filtered data and then by removing the last delay samples of the original and of the time vector.

Once BVP and accelerometer data are filtered, adjusted, and passed to 3 Normalized least mean square (NLMS) adaptive filters, each represents one axis of triaxial accelerometer. The adaptive filter takes two inputs: desired signal (BVP) and input signal. In our case, input signal is each axis of accelerometer and thus 3 NLMS filters to capture motion information in all 3 directions separately. The objective of NLMS filters is to find the filter coefficients adaptively so that it can produce the least mean square error of error signal (difference between desired and actual signal). This is done by suppressing motion artifacts from BVP signal. Thereafter, short-time Fourier transformation (STFT) is used to obtain frequency spectrum from each NLMS adaptive filter output. Resulting coefficients of 3 individual motion-reduced BVP spectrum are multiplied and then cube root is taken to obtain a single dominant frequency spectrum of BVP signal. There is no metric as such to evaluate the performance of employed NLMS and STFT based approach that could be used to evaluate the motion reduced signal. Therefore, visual inspection of motion reduced signal is the only viable option for our work.

Figure 2.1 contains four subplots. Top plot represents low pass filtered BVP data. Three bottom subplots represent outputs (motion reduced BVP) of three NLMS adaptive filters. Each NLMS filter takes into consideration the low pass filtered BVP and one axis of accelerometer data. Last subplot is an exception to others and indicates that the corresponding subject has high motion in z direction. Thereby the resulting NLMS filtered BVP has a high order of amplitudes in z direction.

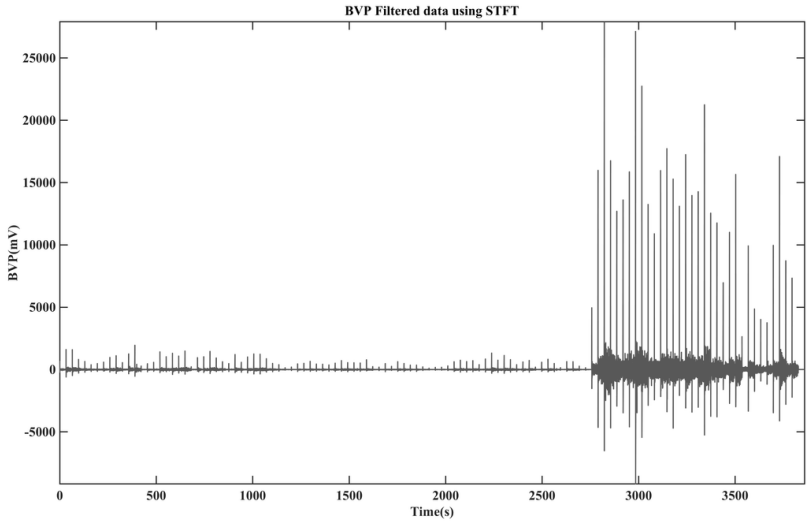


Figure 2.2: Motion reduced entire BVP signal of the same subject mentioned in Figure 2.1 after applying NLMS and STFT. Resulting spectrum is transformed back to time-domain signal.

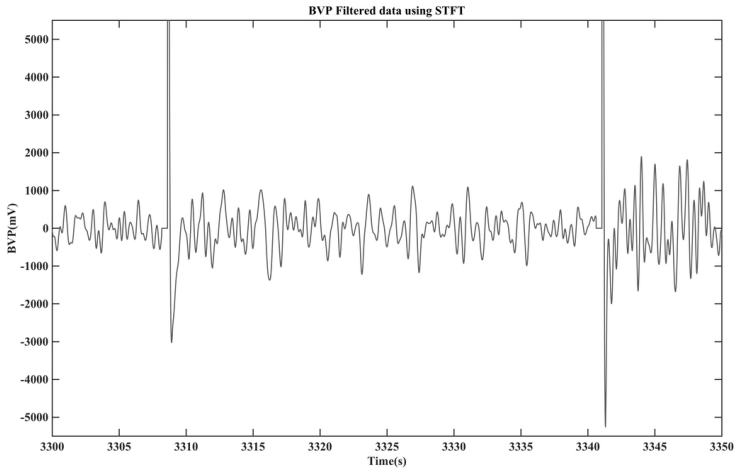


Figure 2.3: Motion reduced BVP signal's snippet corresponding to same region as in Figure 2.1.

The final motion reduced BVP signal as shown in Figure 2.2 has high order outlier peaks occurring at regular intervals of time. Figure 2.1 clearly shows that subject has high motion in z direction but minimal movement in x and y direction that leads to distortion of BVP signal as shown in Figure 2.3.

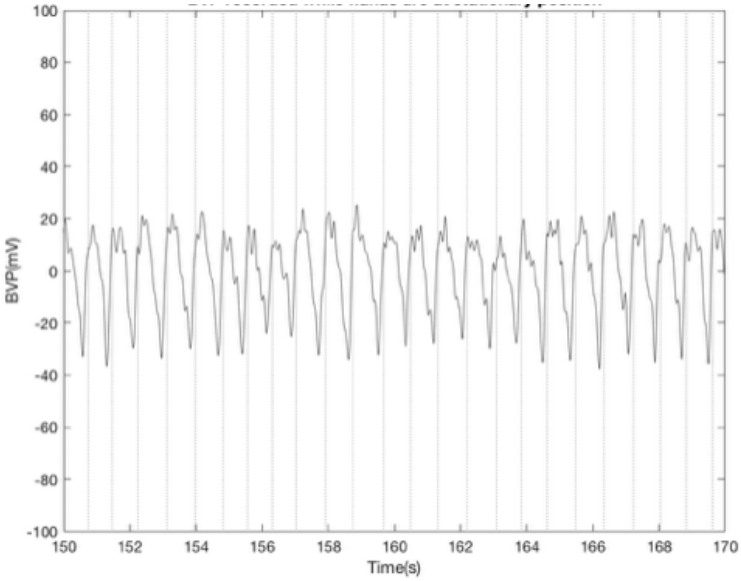


Figure 2.4: Raw BVP data with spurious peaks.

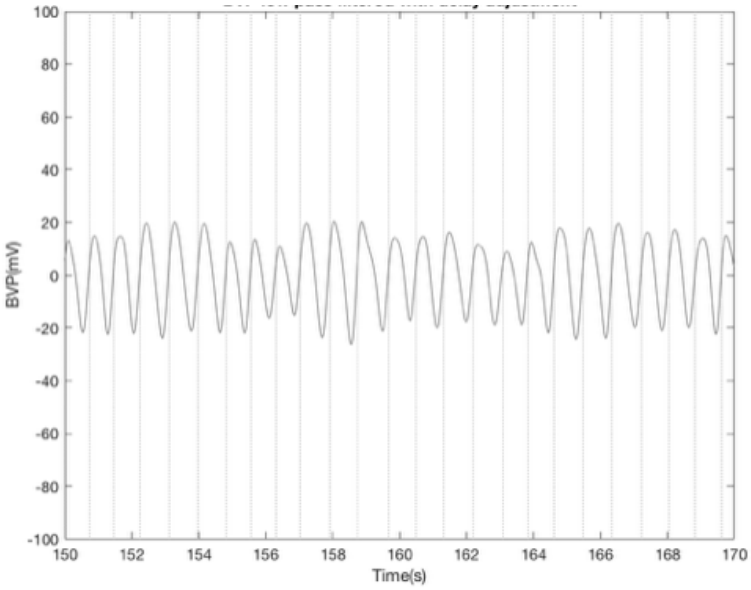


Figure 2.5: Low Pass Filtered BVP.

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 passband frequency of 0.6 Hz. This was chosen because it aligns with a heartbeat of approximately 40 beats per minute, which is less than the range of 50-80 bpm expected for a heart at rest [22]. Similarly, the stop band frequency was chosen at 3.33 Hz (aligning with approximately 200 bpm) based on estimates by Tanaka et. all [23]. 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 resampled the triaxial accelerometer data to 64Hz to match the sampling rate of the BVP measurement.

Figure 2.4 and 2.5 are examples of raw and filtered BVP signal respectively. Figure 2.4 is a raw BVP signal. It consists of dicrotic notch, dicrotic, and spurious peaks (at each local maxima). Spurious peaks are the components of noise that are available in BVP signal and they occur either because of the subject's motion or device's sensors. Dicrotic notches are useful for studying various types of cardiac diseases but only local maxima and minima peaks can be used to distinguish one subject from other and thus to detect non-compliance. Therefore, we ignored dicrotic information in our analysis and considered only local maxima and minima peaks for extracting time domain features. We achieved this using FIR low pass filter that performs smoothening of BVP signal, removes unwanted spurious peaks from it and produces clean BVP signal as shown in Figure 2.5.

2.3 Feature Extraction

Inspired by [12], our feature extraction process relies on statistical features extracted from accelerometer and BVP data collected using the wearable biosensor. The features taken for accelerometer are shown in Table 2.1. Those taken for BVP data are shown in Table 2.2.

Table 2.1: Statistical time-domain features extracted from window = w of each axis of accelerometer.

1	Mean	7	Mean Derivatives
2	Median	8	Inter-Quartile Range
3	Variance	9	Maximum Value
4	Skewness	10	Minimum Value
5	Standard Deviation	11	Zero Crossing Rate
6	Mean Crossing Rate	12	Kurtosis

Table 2.2: Statistical time-domain features extracted from window = w of BVP.

1	number of peaks
2	distance between maximum and minimum
3	mean of the peaks
4	mean peak-peak distance
5	std of the peaks
6	std of peak-peak distance
7	root mean square
8	power spectral density

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.

2.4 Overview of Deliberate Non-Compliance Detection Approach

Once we have the dataset and know which features to extract then the next step is to build the DNC detector. Our detector uses a machine learning-based classifier to address our principal question: if the wearable biosensor data being received from a wearable biosensor assigned to subject X is coming from subject X and that subject X is not opioid intoxicated.

Our classifier learns unique behavioural changes of a subject in neutral state using the accelerometer and BVP data. We only consider the neutral state of the subject for two reasons: (1) it may be impossible or impractical to gather data from the subject during training for intoxication or withdrawal state and (2) since the purpose of our approach is to help introduce monitoring of the subject for their physical well-being, we think that showing that they are in the neutral state is necessary to meet that goal.

If any newly received accelerometer and BVP snippet match with the data of a subject in neutral state used in training, the classifier can be sure there is no DNC. Our detection approach has two phases, the training phase and the detection phase. A diagram of our DNC detection approach is shown in Figure 2.6.

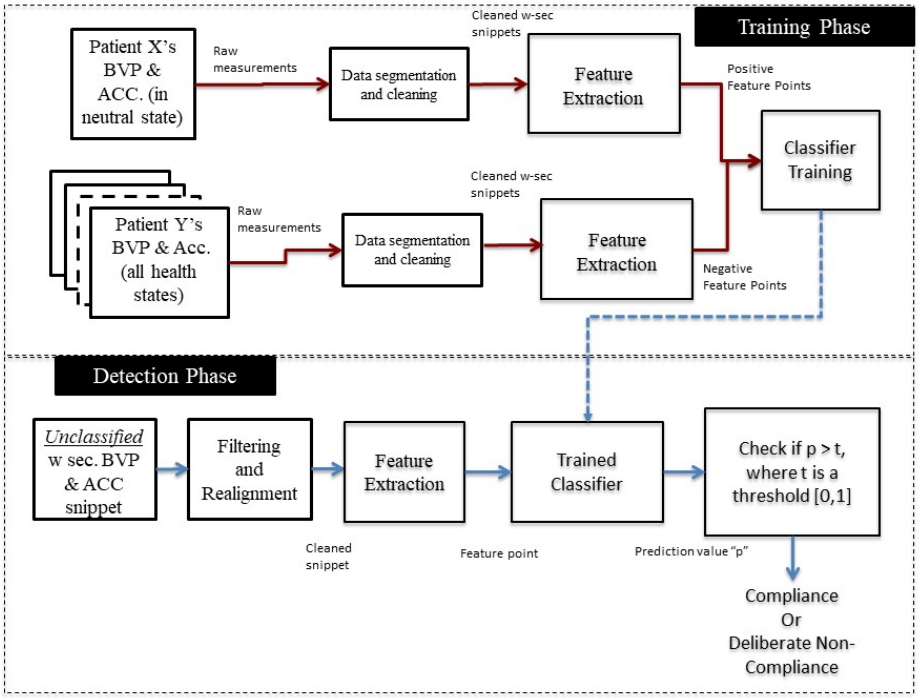


Figure 2.6: Overview of deliberate non-compliance detection approach.

In the training phase, we build a classifier to identify the uniqueness of a subject 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 subjects neutral state. We build a personalized classifier (model) for each subject in our dataset.

During training, for a given subject, we take that subjects neutral state data, and divide it up 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 DNC. We

then divide the neutral, intoxicated, and withdrawal state data from all remaining 10 subjects in our dataset into w size windows, extract the 44 features from each of these windows and call them the negative class points. The negative class points are scenarios there is DNC. Once we have the positive and negative class points, we take a subset of these feature points to train a classifier which forms our DNC detection approach and use the rest for validation purposes.

Once the classifier is trained for a subject, it is now able to classify whether an unseen feature point derived from a w-sized snippet of accelerometer and BVP measurements came to that subject, or whether it is an instance of DNC. 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 subject in the neutral state, and with 0 indicating full confidence that the point does not belong to that subjects neutral state and hence signifies DNC in our setup. We are then able to decide whether to accept or reject that value depending on whether it meets a chosen threshold between 0 and 1 or not.

Chapter 3

Experimental Setup

In this section we briefly discuss our experimental methodology to evaluate the efficacy of our DNC detection approach. In this regard, we provide a short overview of our dataset, the training setup, the metrics of choice, the machine learning classifier chosen, and length of the window w.

3.1 Dataset Curation

Overall, we collected data using the wearable biosensor from 11 subjects who were brought to the ER because of opioid overdose for several hours. During this time, subjects were experiencing different health states. We extracted wearable biosensor data for each subject from the various health states as shown in Figure 3.1.

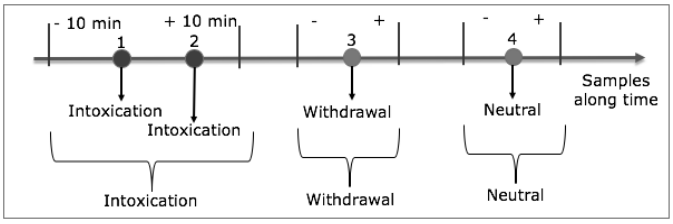


Figure 3.1: Data collection process for subjects from neutral, intoxication, and withdrawal state.

Data collection was done by extracting 20 minutes of device data immediately surrounding a subject evaluation by the research staff, which resulted in the recording of their health state (intoxication, withdrawal, or neutral) at the time. This translates to 10 minutes before and 10 minutes after the evaluation. Since the subject is in a controlled hospital setting, we can have reasonable assurance during this 20-minute

time frame that there would not be a significant change in the health state. In some instances, two consecutive evaluations produce the same health state (e.g., intoxicated), in which case all the data collected from the wearable biosensor between the two evaluations is considered to belong to that state (e.g., intoxicated in the above example). In this manner, we get data from the wearable biosensor for different health states of the subjects. We do not have data for all three health states for all subjects. Depending upon when they were brought into the ED after their overdose, when naloxone was administrated, and when evaluations were conducted, some subjects may only have data collected in their neutral state, others may have intoxication and neutral state data or withdrawal and neutral state data. All subjects in our dataset have neutral health state data. Once we have extracted the pertinent data from the subjects, we have the necessary dataset for answering our principal question on detecting DNC. As we did not have actual DNC in our data collection process, we simulate the presence of DNC in this work. This is done by sending snippets of accelerometer and BVP data from a subject Y as belonging to subject X based on the definition of DNC as described in Figure 1.1. This replacement of device data of subject Xs data with subject Y is akin to DNC because if subject X wants to cheat, we assume they will give their wearable biosensor to subject Y, producing the same result. The health state of subject Y, who is pretending to be subject X, can be in any state (and not necessarily in the neutral state). This is because when a non-complying subject X gives their wearable biosensor to someone else, there is no guarantee that the other person may not abuse opioids themselves. Note that, presumably subject X gives their device to subject Y so that they can use illicit substances, if subject Y would not abuse opioids at the same time. But this is of course not necessarily the case, and our detection systems should be able to identify DNC when subject Y is not neutral as well.

3.2 Curation of Training Data

To be able to train the classifiers, we must compensate for the idiosyncrasies of our data collection protocol. For instance, during the data collection, we were not able to carefully control the subjects actions (especially in the neutral state). Further, given the way we extracted data for different health states, we do not have enough contiguous data to create models at this time that can identify the subjects activities in a given health state, particularly in the neutral state. Therefore, to train our classifier, we first generate the positive and negative class points and then shuffle them.

According to the Figure 3.2, if we want to create a personalized model for a subject X (positive) then we consider data only from subject X's neutral state despite presence of subject X in all the three states as we mentioned in the Figure 1.1. Once subject X's neutral data is extracted then we perform shuffling and took initial 8 minutes for training and rest for validation. Further, for subject Y (negative) data from all the health states are used for training the model. Then we shuffle each health state data separately and consider initial 8 minutes data from each state as training and rest as validation. Figure 3.3 shows that we have total 11 subjects and all are in neutral state. Therefore, we create model for each subject keeping all remaining (10) subjects as negative. Furthermore, negative subjects can be in any possible combination of health states. Therefore, P, Q, and R denotes number of subjects in neutral, intoxication and withdrawal state respectively that distinctively sum up to 10.

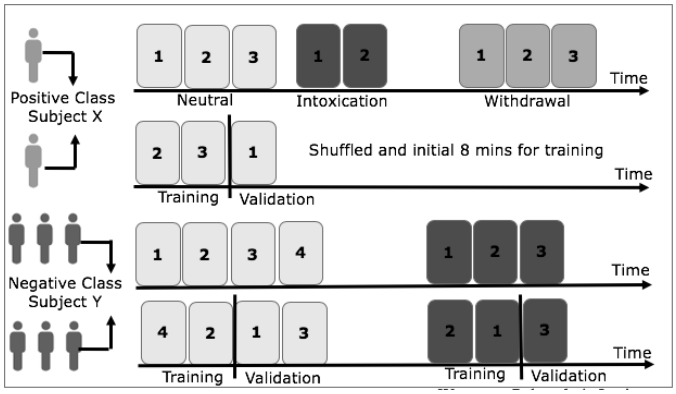


Figure 3.2: Overview of data curation for positive and negative class samples.

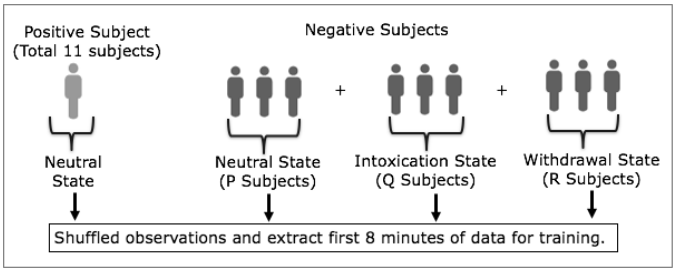


Figure 3.3: Overview of data curation for training and validation set.

We use the first 8-minutes' worth of feature points for training. Since a feature point is produced from w-sized window, 8 minutes of data will have 8/w number of feature points. The rest of the shuffled feature points are used as testing. We chose the value of 8 minutes because every subject had at least 10 minutes of neutral state data. After the data is shuffled and training data is chosen, all the negative and positive samples are feed into a classifier to train it. Since there is far more negative data than positive data, we use weighted classifiers to avoid the model favoring the negative state.

3.3 Metrics

To evaluate the efficacy of our approach, we use the following five core metrics. (1) True Reject Rate (TRR): the rate at which true negative feature points (data points from subjects other than the one that the system is trained for) is rejected by the detector. This indicates the presence of DNC. (2) False Accept Rate (FAR): the rate at which true negative feature points are accepted by the detector. This indicates a false alarm. (3) True Accept Rate (TAR) the rate at which positive feature points (the subjects own unseen data) is accepted by the detector. This indicates the absence of

DNC. (4) False Reject Rate (FRR): the rate at which an unseen positive feature points (the subjects own data) is rejected from the detector. This indicates when DNC 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 detector balances the accuracy of its detection with usability.

3.4 Choosing the classifier

In order to properly run our experiment there are several variables we can set on our model. These include the machine learning classifier and the window size. In order to choose a machine learning classifier that would be effective, we used 5-fold stratified cross validation on our training data. We compared several different algorithms all of which were weighted. That is a correct minority (positive) class decision was considered equally valuable as a correct majority (negative) class decision. We used a value of 1:10 for the ratio of the weighting as there are 11 subjects in our data set, and so for every 1 positive point from a particular subject, there are approximately 10 points belonging to other subjects. The weighted algorithms used were weighted Random Forest [24], weighted Adaboost [25], weighted Support Vector Machine (SVM) with a radial bias function (rbf) kernel [26] [27], and weighted logistic regression [28]. 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.

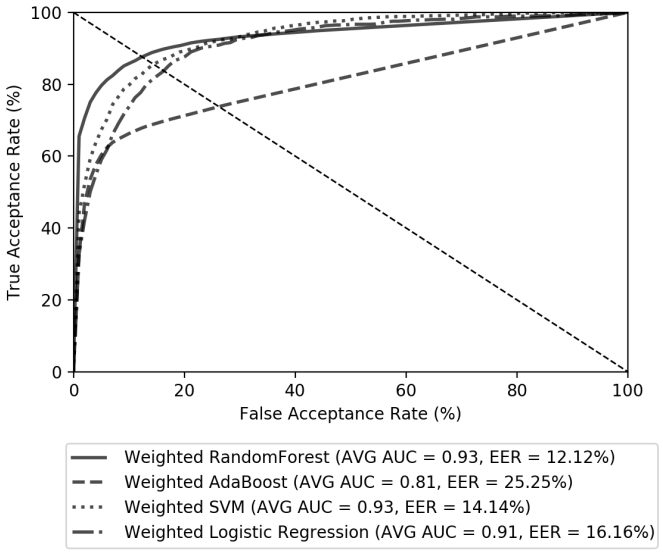


Figure 3.4: ROC curves for 5-fold cross validation on various machine learning algorithms. Here, we set the window size (w) to 5 seconds, provisionally.

The summary of our evaluation in the form of receiver operating characteristic (ROC) curves is seen in Figure 3.4. ROC curves plot TAR against FAR to form a curve. The closer the area under the curve (AUC) is to 1, the better the algorithm performed in cross validation. 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 3.4, the best performing algorithm is weighted Random Forest, which was chosen as our machine learning classifier.

3.5 Choosing window size (w)

Once the machine learning algorithm was 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 3.5, w=3 seconds had the highest average AUC of 0.94 and the lowest EER at 11%. Therefore, it was chosen as the window size for our evaluation.

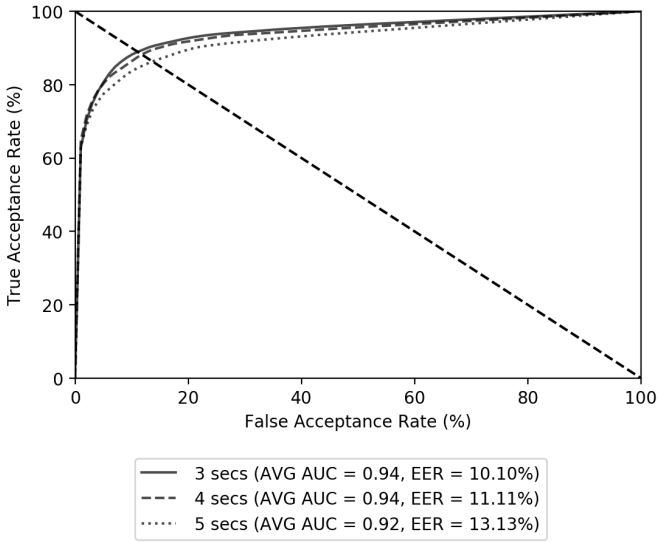


Figure 3.5: ROC curves for 5-fold cross validation on penalized Random Forest with various window sizes. A window size (w) of 3 seconds was chosen based on these results.

Chapter 4

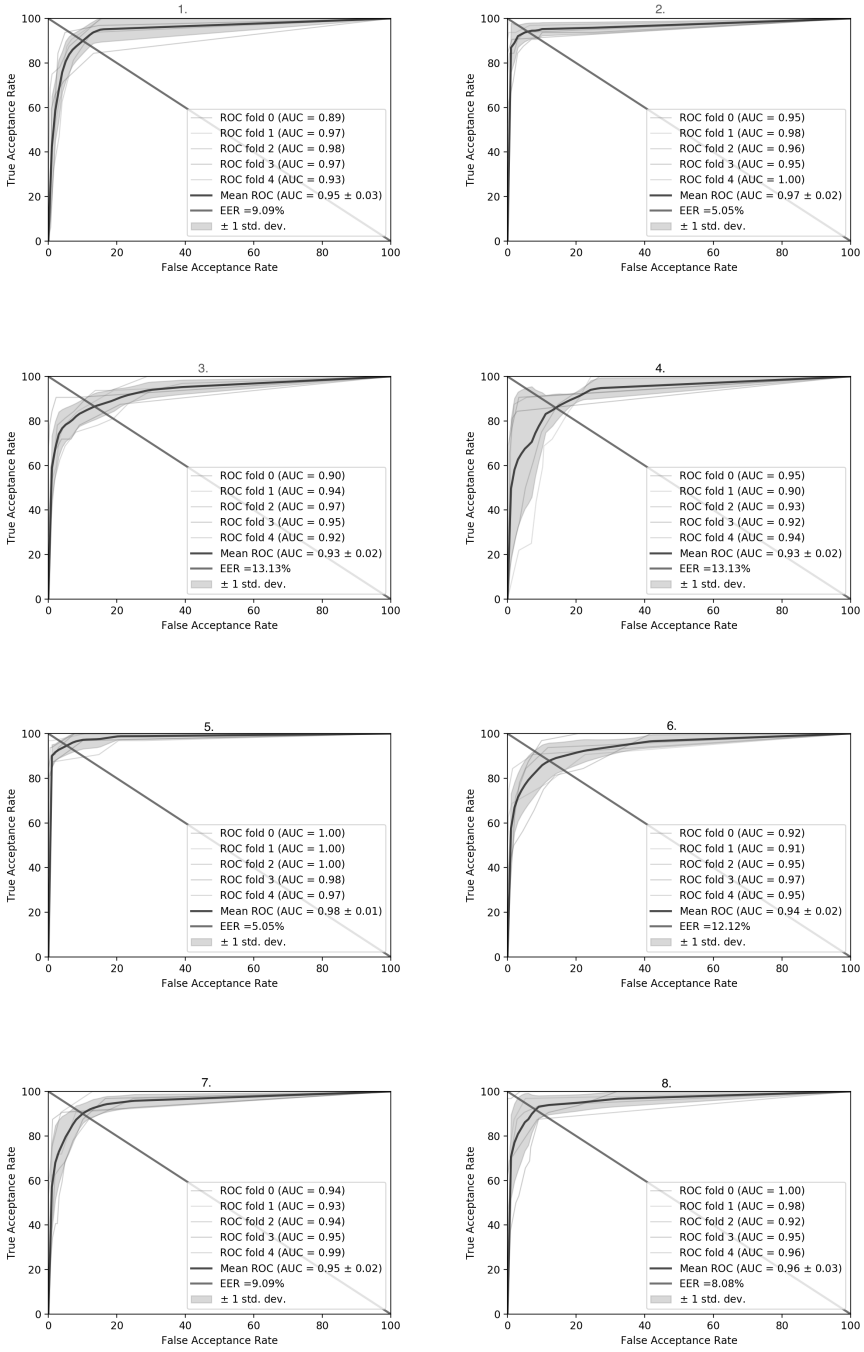
Results

Once our classifier is chosen and trained the next step is to see how well it performs in detecting the presence of DNC. We do this by feeding the subject-specific classifiers in our DNC detector, w-sec snippets of completely unseen triaxial accelerometer and BVP data. These essentially simulates the DNC detector deciding on subject data collected every w-seconds (3 seconds) to say if the data coming from the subject 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 subjects in our dataset that was not used in training, and (2) an external set of subjects, whose data has never been seen before in any form by our classifier. The test data in the validation set contains both positive and negative feature points. The external data set contains only negative feature points, all of which should be labeled negative (i.e., DNC) by the classifier in the ideal case. Our external set included 14 individuals, all of whom were in the neutral state. Out of those 14 individuals, 9 were at rest with limited data movement, and 5 were moving around in a busy conference setting.

4.1 Training Results

Graphs in Figure 4.1 represent stratified 5-fold cross validation results for 11 subjects. Each graph is a ROC curve of a particular subject that is computed based on average stratified 5-fold cross validation. In a graph, there is a thick black line running in the center of the grey region. It represents an average ROC curve. The grey region in the graph represents the deviation of ROC associated with different folds or training set of stratified 5-fold cross validation. More specifically, each fold of stratified 5-fold cross validation results in one ROC and region computed based on minimum and maximum ROC results the grey region. Wider the grey region, more will be the variance in ROC of a subject if tested over different but similar dataset. Therefore, all those subjects which have a large grey region indicate the presence of high variance in their model, leading to a phenomenon called overfitting. However, all above graphs have very less

variance thus indicating the absence of overfitting. Furthermore, the threshold at EER for each subject is used to make predictions on the same subject's test set.



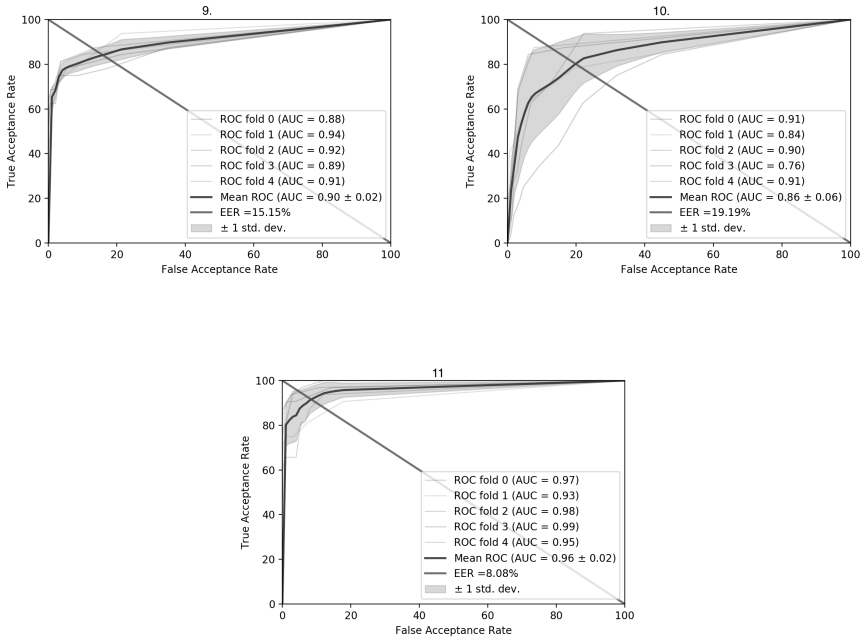


Figure 4.1: Graph of ROC curves for 5 stratified cross validation using weighted random forest with a window of 3 second for all subjects.

4.2 Validation Results at EER Threshold

For the validation set, we achieved promising performance from our DNC detection classifier. Figure 4.2 shows the ROC curve of 11 subject-specific models. The average AUC on the validation set comes to 0.91. The point where the ROC curve interests the diagonal of the graph is the point where EER occurs. This value is of course different for each subject-specific detector. Figure 4.3 lists the individual EER for all subjects in our dataset, the average EER for all subjects is 12.4%. The performance variability for different subjects, can be largely accounted for by the difference in level of activity for between different subjects.

The ROC results show how the classifiers performance varies for the validation set. However, during operation the individual classifier has to choose a specific instantiation of the classifier. Ideally, we want to set this instantiation to be the point where the classifier performed the best during training, i.e., cross-validation. For our classifier is of course the EER point in Figure 3.5 when the window size is 3 seconds.

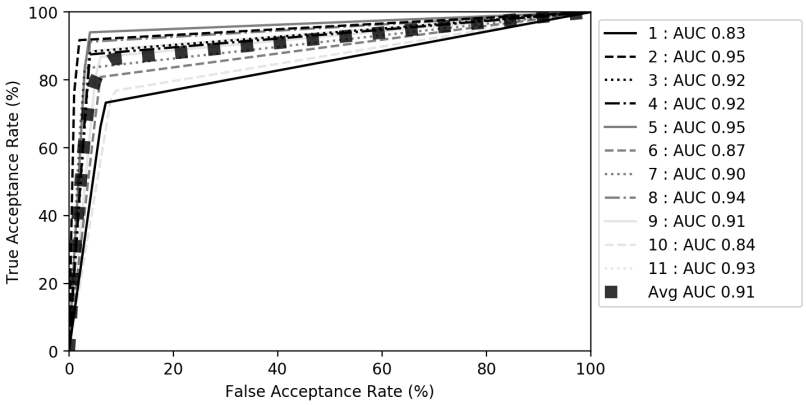


Figure 4.2: Graph of ROC curves for validation set test features using weighted random forest with a window of 3 second for all subjects.

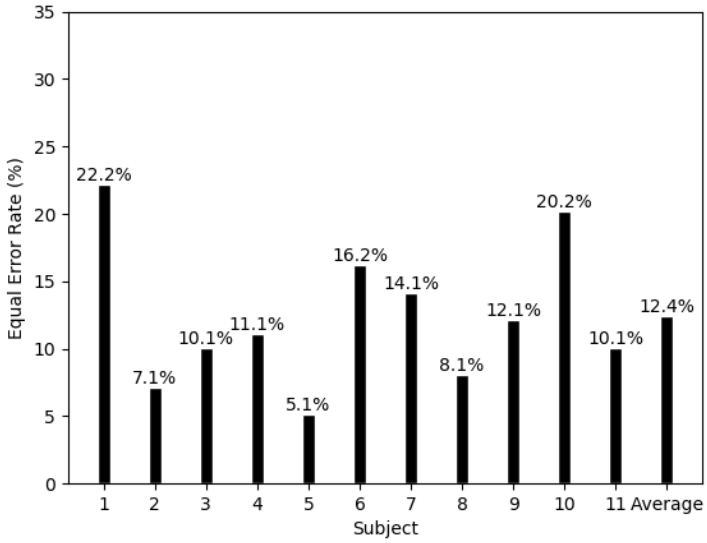


Figure 4.3: Graph of EER rates for all subjects for the test results of the validation set. EER is a measurement of error, so lower bars indicate better performance.

Figure 4.4 shows the performance of our weighted Random Forest classifiers, on the validation set, when it is instantiated such that they produce an EER of .11 (11%) during cross-validation. We see that the classifiers on average achieve a TAR (correctly detecting the absence of DNC) of 86.2%. Our TRR (detecting the presence of DNC)

was even higher with an average of 95.2%. FAR and FRR are correspondingly low. The cumulative accuracy of our DNC detection classifier comes to 90.7%. It is interesting to note that, that TRR spread is much lower than TAR, which shows that our classifiers are very good at detecting DNC, while allowing for some false alarms.

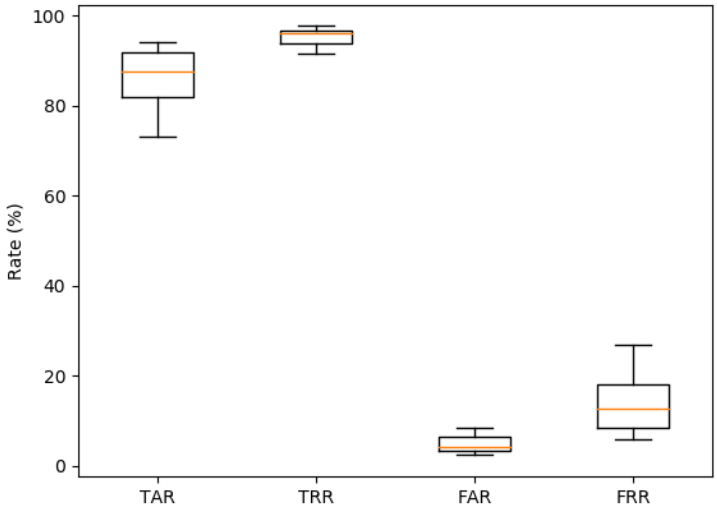


Figure 4.4: Boxplot of TAR, TRR, FAR, and FRR for the validation set.

Table 4.1: Validation set AUC, TRR, FAR, FRR and TAR metrics for each subject.

		Confusion Metrics			
Subjects	Validation set AUC	TRR	FAR	FRR	TAR
1	0.832	93.3	06.7	26.8	73.2
2	0.952	98.8	01.2	08.3	91.7
3	0.922	95.9	04.1	11.6	88.4
4	0.918	96.1	03.9	12.5	87.5
5	0.953	96.6	03.4	06.0	94.0
6	0.874	94.1	05.9	19.2	80.8
7	0.903	97.3	02.7	16.7	83.3
8	0.936	96.0	04.0	08.8	91.3
9	0.906	94.6	05.4	13.3	86.7
10	0.842	91.7	08.3	23.3	76.7
11	0.927	96.8	03.2	11.5	88.5

Table 4.2: Train EER, Validation EER and train EER threshold for each subject.

	EER		
Subjects	Train	Validation	EER Threshold
1	09.1	22.2	0.151
2	05.1	07.1	0.210
3	13.1	10.1	0.247
4	13.1	11.1	0.188
5	05.1	05.1	0.199
6	12.1	16.2	0.184
7	09.1	14.1	0.153
8	08.1	08.1	0.173
9	15.2	12.1	0.136
10	19.2	20.2	0.147
11	08.1	10.1	0.164

Table 4.1 shows the performance of each subject's machine learning model on validation set using AUC, TRR, FAR, FRR and TAR metrics. AUC is computed using ROC curve that depicts the relationship between TAR and FAR. Every subject has high TRR and hence low FAR. However, subjects 1 and 10 have 73.2% and 76.7% TAR respectively that is low as compared to other subjects. Low values of TAR for subjects 1 and 10 lead to low AUC. Training and validation EER for each subject along with their corresponding EER threshold is shown in Table 4.2. Each subject has personalized EER threshold different from default threshold (0.5) of classification model which is used to make predictions on subject's specific test data.

4.3 External Results at EER Threshold

In addition to the 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 detector classifiers in any form before. Further, this data was collected from healthy individuals while they were attending a conference. This simulates the case, where a subject gives their watch to another person (i.e., engages in DNC) who is healthy and always in the neutral state. The evaluation using the external set allows us to make sure that our classifiers are not over-fitting to the training data.

Since all the data in the external set comes from some unknown, ideally, all the feature points generated from this data should be rejected (i.e., classified as negative). Hence, we only report the TRR and FAR results for this portion of the evaluation. To evaluate the feature points from the external set, once again, we instantiate our classifiers at the point where it achieved EER in Figure 3.5 for window size of 3 seconds.

Figure 4.5 shows the cumulative performance results for our classifiers. We can see that our classifiers are able to correctly reject these unseen feature points effectively. Our average TRR overall classifiers is the 86.8%. We did notice some outliers, at least

one classifier performed poorly. We believe that the situation with our outliers could be more effectively dealt with if we had a more comprehensive training set to help train our models.

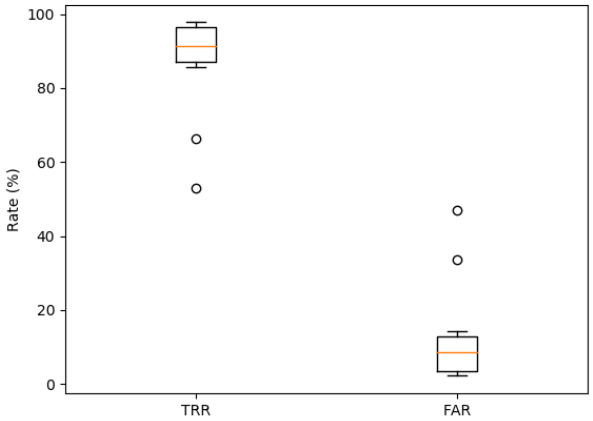


Figure 4.5: Boxplot of TAR, TRR, FAR, and FRR for the external set.

Table 4.3: TRR and FAR for each subject on external data set.

Subjects	TRR	FAR
1	91.5	08.5
2	97.8	02.2
3	96.5	03.5
4	96.4	03.6
5	52.9	47.1
6	85.7	14.3
7	92.5	07.5
8	88.7	11.3
9	89.4	10.6
10	96.6	03.4
11	66.4	33.6

Table 4.3 has TRR and FAR for each subject based on external data set that has not seen by training models. Subject 5 and 11 has TRR = 52.9% and 66.4% respectively. Misclassification of samples from external data set to positive class in these scenarios could be because of two reasons. First, when there are similar variations of accelerometer and BVP signals for these subjects and external data set. Second, subjects in external data set are doing intensive motion activities and their corresponding BVP signal has motion artifacts.

4.4 Validation Results at Default Threshold

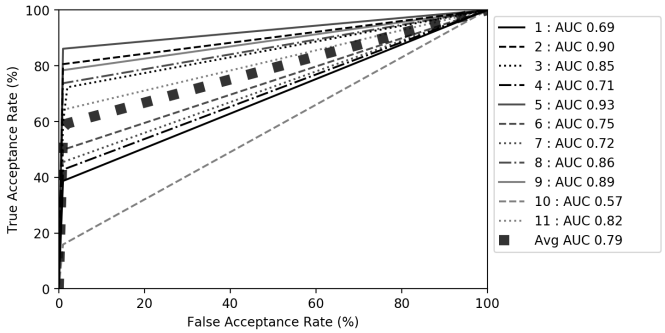


Figure 4.6: Graph of ROC curves for each subject at default threshold.

Figure 4.6 is ROC curve for each subject on validation set when EER threshold is chosen as default (0.5). Machine learning algorithms has default decision value = 0.5 and all predictions that are high than 0.5 are classified to one of the classes in binary classification problems. We compared our validation results in both the cases when this decision value is default and other at personalized decision value. We found that average AUC value drops to 0.79 when decision value is chosen as default. Furthermore, few subjects performed very poorly and has AUC = 0.57 and 0.69. Choosing the threshold value is always critical and depends upon the application. We opted personalized EER threshold value to make final predictions as identifying true negative samples with high accuracy and minimizing false alarm both are critical in our application.

4.5 Discussion

Overall these results show the efficacy of our personalized DNC detection classifiers. We achieved an average DNC detection accuracy (TRR) of 95.2% for the validation set and 86.8% for the external set. Further, our approach can detect the presence of DNC continuously every 3 seconds, which also makes it extremely responsive in detecting the presence of DNC.

However, there are three major limitations to our current work that we plan to address in the future. First, BVP data is sensitive to motion artifacts associated with high motion activities of subjects. In our work, we performed experiments using different approaches to deal with motion artifacts and other noises. However, we used only low pass filter to smoothen the BVP data. But, highly intensive motion activities of a subject can corrupt whole BVP recording and cannot be used for any further analysis. Also, motion artifacts are not linear. They follow nonlinear patterns. Therefore, an appropriate nonlinear technique will be useful to surpass these external noises as mentioned in [29]. In our work, we used only time domain statistic features for BVP.

Additionally, advanced frequency and time-frequency related features could be used for increasing the performance of machine learning models.

Second, 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 subject was not liable to do be doing many different things in the neutral state. However, that is not true outside the ER as the subject may be performing a variety of actions when they are in the neutral state. Our detection models therefore by necessity must be aware of the various activities the subject is engaged in and different these activities from one subject to another. This will result in detection models that are much more complicated. Further, this will improve the TAR spread in figure 4.4 and reduce false alarms.

Third, we have only simulation of DNC in this work. However, to be truly effective these classifiers must detect DNC where subjects actively try put the sensor on a different person. To be able to achieve this we need to be able to we need to create clinical studies that allow subjects to intermittently put the sensor on another person and then record the fact that there was deliberate non-compliance. This will allow us to then train models and check to see if these models are able to detect actual DNC in the real-world setting. An important metric to evaluate in this regard is how quickly is the DNC detected and what the consequences on the subjects and the caregivers are for wrong detections at scale.

Chapter 5

Conclusion

The principal aim of this thesis is to develop an approach for deliberate non-compliance detection by a subject. This subject is being surveilled, using a wrist-worn wearable medical device, for opioid abuse. We define deliberate non-compliance as the process of giving one's device to someone else when the surveillance is on-going. An analysis of our approach shows that it can identify instance of deliberate non-compliance and compliance with over 90% accuracy for most of the subjects in our dataset. We used BVP and triaxial accelerometer data in our work because of their importance in capturing physiological and movement information.

Our approach works using a personalized machine learning classifier (model). This model can identify if a snippet of physiological and movement received is coming from the expected subject or someone else. Each personalized machine learning classifier is based on three factors. First, the classifier will assign more weight to samples of minority class (legitimate) subject for being misclassified as compared to samples of majority class (non-legitimate) subjects. These models are called weighted classifiers and they are used to avoid overfitting when distribution of samples for legitimate and non-legitimate subjects are highly imbalanced. Second, the classifier will make predictions on test set based on each subject's EER threshold chosen during training. Each subject's EER threshold is a tradeoff between TAR and FAR and can be chosen based on the type of application or clinical requirements. Third, legitimate subjects are always in neutral state, but non-legitimate subjects can be in any one of the three states intoxication, withdrawal or neutral.

In this preliminary study we showed the promise of our approach as a method for detecting DNC in an opioid surveillance program that relies on a wearable biosensor. Even though our approach focuses on deliberate non-compliance during opioid abuse surveillance, it can also detect accidental non-compliance as well. This means, if someone else puts on a subjects 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 subject.

Chapter 6

Future Work

In the future we plan to expand on this work in several directions. First, setup clinical studies that create actual DNC scenarios fine-tune our models. In our work, we simulated DNC by considering all remaining subjects but one as non-legitimate subjects. Also, all non-legitimate subjects are assumed to be in one of three states intoxication, withdrawal or neutral state. In addition to that neutral state of all subjects are considered as monolithic state. However, in real life activities can vary from walking to highly intensive running activity etc. Considering all possible types of activities for a subject, machine learning classifier (model) can be fine-tuned for everyone so that personalized model can be used with high accuracy.

Second, build a cloud-based application that can detect the presence of DNC in clinical studies, at scale, and other surveillance situations and thus allow for timely intervention in the event of opioid abuse among individuals. DNC can be used in wide area of surveillance; e.g., subjects can deliberately hand over the device to other subjects or can attach the device to a dog or spoke of a car for gaining financial incentives for false claims from insurance company. By scaling, it means that cloud-based application can process large amount of data from millions of patients and can detect DNC for each individual subject. Furthermore, cloud-based application can monitor other vital signs of patients in an ongoing treatment of neurological, cardiovascular and pulmonary diseases

Third, different noise cancellation approaches have been proposed to compensate motion artifacts and other noises in BVP signal. Signal analysis work in [30] consists of two stages. In first stage, they detect the corrupt signal using a Short-Term Fourier Transform (STFT). In second stage, Lomb-Scargle Periodogram (LSP) is used to approximate the Power Spectral Density (PSD) of the BVP signal. The algorithm has proven to be effective in removing artifacts which disrupt the signal for a short period of time. However, disruption of BVP signal for a long period of time can make this approximation lose auxiliary information.

Research work in [21] has attempted to estimate heart rate from foot worn PPG signal during fast bike activity. The raw PPG and co-located accelerometry signals are first band-pass filtered (3rd order Butterworth, 0.520 Hz) and down-sampled by a factor

D (default D = 10). They have presented a normalized least mean square adaptive filter and short-time Fourier transform based algorithm to remove motion artifacts from contaminated BVP signal. There is a lot of research as described above that can effectively remove motion artifacts from the BVP signal. In our work, we used low pass filter with finite-duration impulse response (FIR) to remove high frequency noise and spurious peaks from the BVP signal. The detection and removal of motion artifacts from BVP data create an obstacle in getting accuracy with higher confidence interval. In our future work, we will address this issue using non-linear motion artifact removal techniques. This is because artifacts correspond to non-linear patterns. Additionally, we used only time domain statistic features for BVP. Moreover, advanced frequency and time-frequency related features could be used for increasing the performance of machine learning models for detecting DNC.

Bibliography

[1] A. Pentland, "Healthwear: medical technology becomes wearable," *Computer*, vol. 37, no. 5, pp. 42–49, 2004.

[2] Y.-L. Zheng, X.-R. Ding, C. C. Y. Poon, B. P. L. Lo, H. Zhang, X.-L. Zhou, G.-Z. Yang, N. Zhao, and Y.-T. Zhang, "Unobtrusive sensing and wearable devices for health informatics," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1538–1554, 2014.

[3] P. F. Binkley, "Predicting the potential of wearable technology," *IEEE engineering in medicine and biology magazine*, vol. 22, no. 3, pp. 23–27, 2003.

[4] S. Carreiro, H. Fang, J. Zhang, K. Wittbold, S. Weng, R. Mullins, D. Smelson, and E. W. Boyer, "imstrong: deployment of a biosensor system to detect cocaine use," *Journal of medical systems*, vol. 39, no. 12, p. 186, 2015.

[5] S. Carreiro, K. Wittbold, P. Indic, H. Fang, J. Zhang, and E. W. Boyer, "Wearable biosensors to detect physiologic change during opioid use," *Journal of medical toxicology*, vol. 12, no. 3, pp. 255–262, 2016.

[6] TheNewStack, "Empatica e4 device," 2016.

[7] H. University, "The opioid crisis in america."

[8] FZB and A. S. Photo, "Isolated business man give something."

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[10] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pp. 1322–1328, IEEE, 2008.

[11] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica e3a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*, pp. 39–42, IEEE, 2014.

[12] N. Alshurafa, J.-A. Eastwood, M. Pourhomayoun, S. Nyamathi, L. Bao, B. Mortazavi, and M. Sarrafzadeh, "Anti-cheating: Detecting self-inflicted and impersonator cheaters for remote health monitoring systems with wearable sensors.," in *BSN*, vol. 14, pp. 1–6, 2014.

[13] J. C. Sriram, M. Shin, T. Choudhury, and D. Kotz, "Activity-aware ecg-based patient authentication for remote health monitoring," in *Proceedings of the 2009 international conference on Multimodal interfaces*, pp. 297–304, ACM, 2009.

[14] D. Gafurov, E. Snekkenes, and P. Bours, "Gait authentication and identification using wearable accelerometer sensor," in *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*, pp. 220–225, IEEE, 2007.

[15] F. Canento, H. Silva, and A. Fred, "Applicability of multi-modal electrophysiological data acquisition and processing to emotion recognition," *Computing Paradigms for Mental Health*, vol. 59, 2012.

[16] T. V. Santiago and N. Edelman, "Opioids and breathing," *Journal of Applied Physiology*, vol. 59, no. 6, pp. 1675–1685, 1985.

[17] D. De, P. Bharti, S. Das, and S. Chellappan, "Multi-modal wearable sensing for fine-grained activity recognition in healthcare," *IEEE Internet Computing*, no. 1, pp. 1–1, 2015.

[18] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv preprint arXiv:1604.08880*, 2016.

[19] V. Radu, N. D. Lane, S. Bhattacharya, C. Mascolo, M. K. Marina, and F. Kawsar, "Towards multimodal deep learning for activity recognition on mobile devices," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 185–188, ACM, 2016.

[20] M. Zhang and A. A. Sawchuk, "Usc-had: a daily activity dataset for ubiquitous activity recognition using wearable sensors," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 1036–1043, ACM, 2012.

[21] D. Jarchi and A. J. Casson, "Estimation of heart rate from foot worn photoplethysmography sensors during fast bike exercise.," in *EMBC*, pp. 3155–2158, 2016.

[22] D. H. Spodick, "Survey of selected cardiologists for an operational definition of normal sinus heart rate.," *The American journal of cardiology*, vol. 72, no. 5, pp. 487–488, 1993.

[23] H. Tanaka, K. D. Monahan, and D. R. Seals, "Age-predicted maximal heart rate revisited," *Journal of the American College of Cardiology*, vol. 37, no. 1, pp. 153–156, 2001.

[24] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[25] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and computation*, vol. 121, no. 2, pp. 256–285, 1995.

[26] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[27] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152, ACM, 1992.

[28] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 215–242, 1958.

[29] Y. Ye, Y. Cheng, W. He, M. Hou, and Z. Zhang, "Combining nonlinear adaptive filtering and signal decomposition for motion artifact removal in wearable photoplethysmography," *IEEE Sensors Journal*, vol. 16, no. 19, pp. 7133–7141, 2016.

[30] S. Luo, J. Zhou, H. B.-L. Duh, and F. Chen, "Bvp feature signal analysis for intelligent user interface," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1861–1868, ACM, 2017.