Smartphone Inference of Alcohol Consumption Levels from Gait

Zachary Arnold, Danielle LaRose and Emmanuel Agu
Computer Science Department, Worcester Polytechnic Institute, Worcester, MA 01609
{zparnold, dmlarose}@wpi.edu, emmanuel@cs.wpi.edu

Abstract— **Excessive alcohol use is the third leading lifestyle-related cause of death in the United States. Smartphone sensing offers an opportunity to passively track alcohol usage and record associated drinking contexts. Drinkers can reflect on their drinking logs, detect patterns of abuse and self-correct or seek treatment. In this paper, we investigate whether a smartphone user's alcohol intoxication level (how many drinks) can be inferred from their gait. Accelerometer data was gathered from the smartphones of a group of drinkers. Time and frequency domain features were then extracted and used for classification in a machine learning framework. Various classifiers were compared for a task of classifying the number of drinks consumed by a user into ranges of 0-2 drinks (sober), 3-6 drinks (tipsy) or >6 drinks (drunk). Random Forest proved to be the most accurate classifier, yielding 56% accuracy on the training set, and 70% accuracy on the validation set. Using these results, AlcoGait, an Android smartphone application was developed and evaluated by real users. The results of user studies were encouraging.**

Keywords-alcohol consumption, inference, smartphone, gait

I. INTRODUCTION

Excessive alcohol use is the third leading lifestyle-related cause of death in the United States [4]. In 2012, 5.9 percent of all global deaths (3.3 million deaths), were attributable to alcohol consumption [2]. Alcohol also contributes to over 200 diseases and injury-related health conditions including alcohol dependence, liver cirrhosis, cancers, and injuries. Alcohol is abused for several reasons including coping with negative emotions, enhancing positive emotions and gaining social acceptance [5].

In this paper, we explore whether smartphone sensing can detect and track alcohol consumption. Specifically, we investigate whether a smartphone can passively infer how many drinks its owner has consumed while the user walks around as usual with their phone in their pockets, hand or bag. Aside from direct BAC or BrAC testing using breathalyzers, neuromotor testing including analysis of gait is the most reliable way to determine intoxication in humans [7]. Our general approach is to infer alcohol consumption by running machine learning classifiers on the smartphone (an app), which analyzes data gathered from its accelerometer. We wanted to determine whether alcohol-induced variances in the smartphone user's gait could be detected even in the presence of confounding factors that are typical of free-living situations. Such factors include diverse phone placements, orientation, gender, weight, activities and walking behaviors.

Prior methods to track alcohol consumption include questionnaires such as Alcohol Use Disorder Identification Test (AUDIT) questionnaires. Questionnaires are well known to suffer from recall bias and may be have inaccuracies of up to 20%. Existing alcohol apps mostly allow the user to manually record drinks consumed. To the best of our knowledge, there is no smartphone app that currently detects the drinking levels of users in real time [13]. Kao *et al* proposed a smartphone-based method to detect *whether (Yes/No)* a smartphone user has consumed alcohol but does not infer the quantity (*how many drinks*).

Smartphone inference of alcohol consumption levels can be used in multiple ways to either treat hard drinkers or to mitigate alcohol mishaps. *Usage in prevention:* First, the user can receive just-in-time notifications of [excessive] alcohol consumption. A drinker who is unsure if they are too drunk to drive at a party can walk a few yards and check their phone's alcohol inference. *Use in treatment:* A smartphone can log a frequent drinker's drinking patterns and associated contexts. Drinkers can reflect on their drinking logs, detect recurrent patterns of abuse and self-correct or seek treatment. Counselors can use such logs as evidence to prescribe treatment. In cases where the user loses consciousness, emergency room physicians would have an accurate record of a patient's consumption history.

A. Challenges

While the potential benefits are substantial, gait inference is a challenging research problem that is still actively researched in computer vision, biometric recognition and health assessment. Specific challenges are now summarized.

- *Alcohol is a controlled substance:* Supervised machine learning requires accurate labeled data (gait data of subjects labeled with the number of drinks consumed). To obtain such labeled data, ideally we would have conducted controlled experiments in which subjects were served various amounts of alcohol and then their corresponding gait data was collected. However, only licensed individuals and establishments can serve alcohol. As we did not have an alcohol license, we were not able to run controlled data gathering experiments.

- *Gathering reliable data is difficult:* Since we could not serve users alcohol, we had to rely on subject self-reports of alcohol consumption. Self-reports are good for generating rough estimates but for the high-precision labeled data required for machine learning, data quality issues arise. Users may forget to record entries, may not know how many drinks (e.g. cocktails served at a bar), or forget when they consumed alcohol.

978-1-4673-9548-9/15 $31.00 © 2015 IEEE
DOI 10.1109/ICHI.2015.59

- *Noise:* User behaviors in the wild are unpredictable causing noisy measurements. A user may place their phone in a wide range of pockets, coat pockets or bags. Many users (about 50%) also leave their phones on the table [26] during their day. They may also lend their phone to friends or their batteries may also die during data collection. Figure 1 shows some factors that may affect gait measurements using a smartphone.

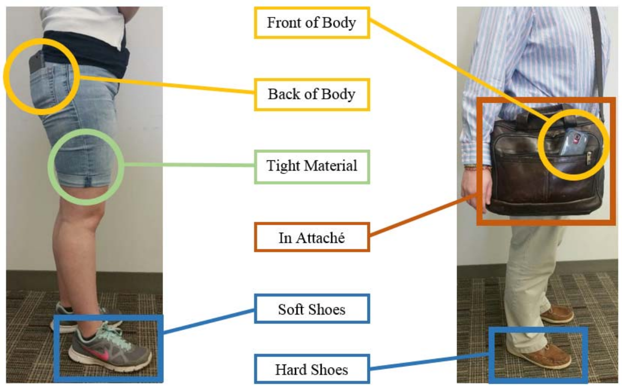


Figure 1 - Experimental Factors that could affect Gait inference

- *Confounding factors:* Apart from alcohol, several other factors alter user gait including fatigue [25] and their mood [24]. Accurately inferring alcohol consumption with these confounding factors present is challenging.
- *Generalizability:* Each person's walk is unique and gait patterns vary depending on gender, weight and many other factors. Classifiers trained on a given population may not generalize to other populations.

B. Our Contributions

This work takes a first step in an important new direction. Our key contributions, expounded in later sections include:

- *Novel IRB-approved data collection method:* Since we were not licensed to serve alcohol in controlled experiments, we devised a data collection method. A smartphone app continuously gathered subjects' gait data whenever they walked. Subjects then self-reported any alcohol consumption the next day (Section III).
- *Gait feature exploration:* We explored how sensitive various time and frequency domain gait features were to alcohol consumption (Section IV).
- *Trained machine learning classifiers:* Using frequency and time domain gait features, we trained machine learning classifiers to infer alcohol consumption from gait. Random Forest was the most accurate classifier, achieving 56% accuracy on the training set, and 70% accuracy on the validation set when classifying user alcohol consumption into ranges of 0-2 drinks (sober), 3-6 drinks (tipsy) or >6 drinks (drunk) (Section V). We were significantly more accurate than a random selection of the 3 bins, which would have yielded 33.3%.
- *Developed a smartphone application for alcohol detection:* Leveraging our machine learning classifiers, we developed AlcoGait, an Android app that can passively infer how many drinks its user has consumed in free-living situations (Section VI).
- *AlcoGait evaluation:* The results of a user study to evaluate AlcoGait were encouraging (Section VII).
- *Explored personalization:* We explored whether per-person machine learning classifiers that learned users' individual gait characteristics could improve AlcoGait's accuracy. Inference accuracy was improved for 66% of users. (Section VII).

II. Background

A. Measures of Alcohol Consumption

When a person drinks alcohol, it either goes into their blood or is released through their breath, urine, or sweat [10]. The standard measures of alcohol are Blood Alcohol Concentration (BAC) or Breath Alcohol Concentration (BrAC), which are the amounts of alcohol in a person's blood or breath respectively. BAC and BrAC can be measured by breath, blood, or urine tests [9].

B. Effects of Alcohol on Human Gait

Approximately ten minutes after initial alcohol consumption, the drinker's heart rate begins to increase in order to filter out the toxins from the bloodstream through the kidneys. After about twenty minutes, the alcohol penetrates the blood-brain barrier noticeably impacting cognitive and neuromotor functions such as human gait [7]. Human gait is a coordinated effort by the brain and muscles to produce mobility or walking [8]. Alcohol impairment can dramatically impact the ability to walk, jog, or run.

C. Gait Analysis Techniques

Gait analysis assesses human gait in order to determine any abnormalities [18]. In this paper, we explore gait changes due to alcohol consumed. Human locomotion produces a signal in the tri-axial accelerometer sensor of smartphones (See figure 2), which can be processed to infer various user activities.

D. Prior studies about the effects of alcohol on human gait

Prior studies have found anomalies in human gait following alcohol consumption. While they do not use a smartphone for sensing gait, we review them here because their results inspired our work and ultimately informed our choices of gait parameters to investigate as features in our machine learning model. Ando *et al.* [7] conducted a study to determine the effects of alcohol ingestion on neuromotor functions, postural sway, hand tremors, and reaction time in thirteen healthy males. Participants were served either alcohol or juice. Sway area and transversal sway tended to increase after alcohol ingestion [7].

Nieschalk *et al*. [14] observed the effects of low or moderate amounts of alcohol on the body with respect to equilibrium. The BrACs of the participants were measured 30 minutes after the ingestion of alcohol. They determined that "sway area was the most sensitive parameter for detecting increased body sway after alcohol ingestion" [14].

Demura and Uchiyama [6] observed the gait of fifteen male adults at normal and controlled tempos before alcohol ingestion and at 10, 20, and 30 minutes after alcohol ingestion. Gait was measured using a gait analysis apparatus to record time and spatial information. Their gait cycle, stance phase, gait velocity, cadence, stride and one step width of participants' gait were analyzed. They discovered a decline in static balance ability, stride length, gait velocity, and cadence around 20 minutes after alcohol ingestion [6].

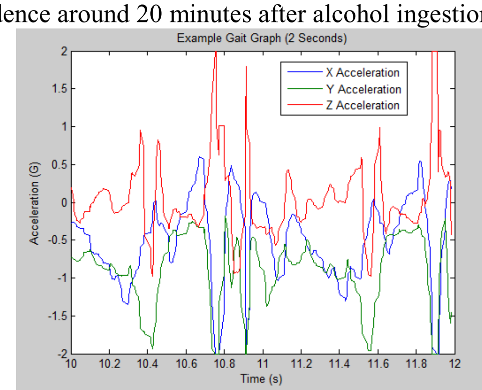


Figure 2 – Example gait data while user walks

E. *Existing alcohol applications on smartphones*

Several smartphone alcohol apps have emerged [16, 17]. Alcohol logging apps have menus through which a user can manually record and track the quantity, time and types of alcoholic drinks consumed. However, manual recording is tedious and the user's recollection of their alcohol consumption may be inaccurate.

BAC calculator smartphone applications calculate BAC based on information entered by a user such as the quantity and time of alcohol consumption. However, many BAC calculators are inaccurate because they use wrong formulas or the user inputs their alcohol consumption accurately. Weaver *et al.* [13] reviewed several alcohol-related Apple and Android smartphone applications to determine their accuracy and relevance in measuring a person's BAC. They found that most alcohol applications were for entertainment purposes and tended to encourage alcohol consumption, rather than for the anticipated health promotion purposes. They found that BAC levels calculated by 98 alcohol apps were highly unreliable and inaccurate [13].

Some smartphone applications to mitigate alcohol abuse have also been proposed. Kao *et al.* [15] created a smartphone sensing system that analyzed any anomalies in a user's gait in order to detect if they had consumed alcohol. Their system used the smartphone's tri-axis accelerometer and recorded the location and time whenever it detected that alcohol had been consumed. Gait anomalies due to alcohol consumption were detectable in 3 subjects, but it was concluded that more data from more participants was required [15]. The quantity of alcohol consumed was not inferred, which we investigate in this paper.

Wang *et al.* [3] designed SoberDiary, a smartphone application that patients recovering from alcohol dependence could use to track daily sobriety after completing alcohol withdrawal treatment. Patients recorded their alcohol levels using a Bluetooth Breathalyzer, from which SoberDiary retrieved their readings. SoberDiary also reduced relapse through artistic illustrations to educate patients on typical symptoms at each stage of the recovery process. SoberDiary allowed patients to perform the breath alcohol test, review personal progress, share their recovery, and input their current emotions [3]. Alcohol consumption in SoberDiary was tracked using a Bluetooth breathalyzer, but did not infer alcohol levels from human gait as investigated in this paper.

III. GAIT DATA COLLECTION

Since we were not permitted to serve alcohol, subjects just ran a smartphone app that continuously gathered gait data. Whenever they consumed alcohol, the following day, subjects would self-report how much and when, protocol for accurate alcohol studies suggested by Del Boca and Darkes [13]. Users were not to record their alcohol consumption at the same time that they were drinking. Instead, twice the next day, they were prompted to retroactively enter the amount and type of alcohol consumed on a smartphone application. In this way, users labeled their gait data with the corresponding number of drinks consumed. Subjects were not required or encouraged to consume alcohol during the study. The details of the study are summarized below:

- *Recruitment:* Once IRB approval was received, email was sent to WPI students and faculty inviting them to participate in the study. Friends and family were also recruited by word of mouth and emails.
- *Subject screening:* The AUDIT questionnaire was used to eliminate subjects who had prior history of alcohol issues. Eligible subjects had to be at least 21 years old, own an Android smartphone, and have received a score of 8 or less out of 40 on the AUDIT questionnaire.
- *Informed consent:* All qualifying subjects signed an informed consent form. Any questions they had about the study were answered at this time.
- *Installing data collection app on their smartphones:* Once subjects had signed the informed consent, our data collection app was installed on their smartphone.
- *Study protocol:* The participants were asked to continuously run the application over a two week period with an option to continue for an additional two weeks depending on the amount and quality of data collected. They were then asked to continue their daily routine as usual. The data recording app automatically detected whenever the user was walking and gathered accelerometer data till they stopped walking. Participants were given the ability to opt-out of the study at any point. If subjects opted out, they could

either allow any data collected thus far to remain in the study or could have their data removed from the study. For privacy reasons, each study participants was assigned a random ID number. We were not able to identify participants or them with associate their data. We only knew the randomized numbers and the data associated with those numbers.

A. Smartphone app to gather gait and alcohol data

Our gather accelerometer data, our data collection leveraged Funf, a third party library for the Android operating system. Funf is an open-sensing framework that enables smartphone sensor data to be recorded at a selected sampling rate and automatically transferred to a remote location for analysis [40]. The startup screen of the app is shown in Figure 3.

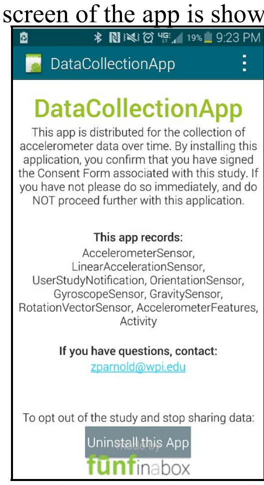


Figure 3 - Opening Screen of Data Collection App

IV. GAIT FEATURE EXPLORATION

In this section, we describe our explorations of gait features in both the time and frequency domain. Pre-processing steps, feature generation and exploration are described.

A. Pre-processing raw accelerometer readings:

The raw accelerometer data were pre-processed in the following steps before gait features could be generated.

a) *Normalization:* The accelerometer data was collected while the phone was oriented in different directions or while the phone was carried in different positions (pockets, hands or bags). To normalize the data, the gravity-corrected magnitude of groups of *n* accelerometer readings was calculated using equation 1:

$$b(t) = \sqrt{x_t^2 + y_t^2 + z_t^2} - \frac{\sum_{t=1}^{n} \sqrt{x_t^2 + y_t^2 + z_t^2}}{n} \quad (1)$$

b) *Smoothing to reduce noise effects:* Smartphone accelerometer data are sensitive to physical movements, and contain lots of noise. To derive a more stable signal, we calculated a moving average for a window size of 10 secs (~2000 accelerometer observations).

B. Gait Feature Generaton

A gait feature is a property of the signal representing gait that can be calculated from the phone's raw accelerometer data [19]. Since human walking is a periodic motion, both time and frequency domain features were useful [8].

1) Time domain features: We calculated time domain gait features that had been useful in other work on gait. Demura and Uchiyama [6] found that the number of steps taken in a given time interval was impacted by alcohol consumption. The number of steps taken per time window can be calculated from a time-series of accelerometer data by finding the number of local maxima of the gravity-corrected magnitude of the accelerometer signals that exceed one standard deviation from the mean of the signal [20]. Figure 4 shows example accelerometer data with the number of steps highlighted.

Kao *et al* [15] also found that the average step length and step time also change when alcohol is consumed. As more alcohol is consumed, the gait stretch (average step length) and step time both change (sober vs. intoxicated). Figure 5 from Kao *et al* illustrates these features.

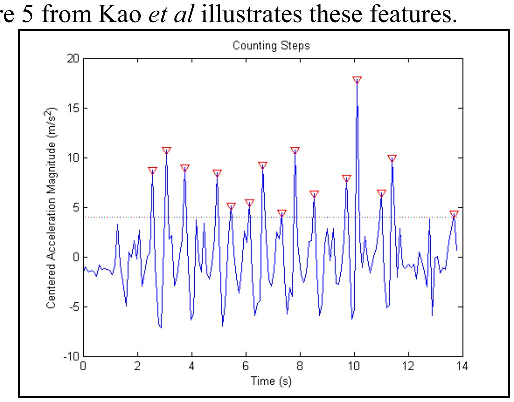


Figure 4 - Example Data with Number of Steps Highlighted

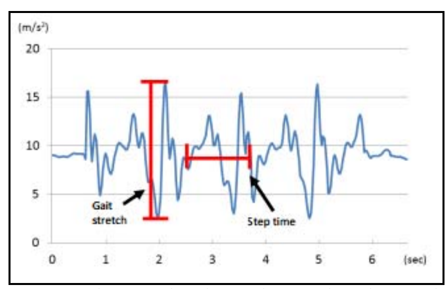


Figure 5 - Example Data showing Gait Stretch and Step Time [15]

Other time domain features explored include gait velocity and gait cadence [8], which were found to be affected by alcohol consumption levels [6]. Gait velocity is a ratio of the total distance covered divided by the total number of steps. Gait cadence is a ratio of the number of steps taken divided by the amount of time taken. Additionally, we added the skewness and kurtosis features of the marginal distribution of the signal to help further characterize the data for the classifier. Skewness is a measure of the lack of symmetry in a dataset. Similarly, kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. The time-domain features we explored are shown in table 1.

TABLE 1 – TIME DOMAIN GAIT FEATURES EXPLORED

Time Domain Feature	Definition
Number of Steps [8]	The number of steps taken in a given time interval
Average Step Length [15]	Average in the distance covered by each step
Average Step Time [15]	Average in the time covered by each step
Gait Velocity [8]	Ratio of the total distance covered by the total time
Cadence [8]	Ratio of the total number of steps by the total time
Skewness [8]	Asymmetry of the signal distribution
Kurtosis [8]	"Peakedness" of the distribution and the heaviness of its tail

2) Frequency Domain Features: To convert the raw accelerometer data to the frequency domain, we applied the Discrete Fourier Transform (DFT) as shown in equation 2.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k\frac{n}{N}}, k = 0, \ldots, N-1 \tag{2}$$

One helpful tool to provide insight on the presence of alcohol in gait was the one-sided power spectral density (PSD) of a signal. A periodogram is one way to estimate the PSD of the signal and is calculated from the DFT. Unlike the simple DFT, it describes how the variance of data in the time domain is distributed over the frequency components into which the signal can be decomposed [33]. Frequently in signal analysis, there is a considerable amount of noise in the periodogram of the PSD. To reduce the effects of noise, we calculated the one-sided PSD of the data using Welch's overlapped segment averaging estimator algorithm (Fig 6). From the PSD, the frequencies where the energy of the signal is distributed can be discovered. The highest peak was at the fundamental frequency of the signal. During the activity of walking, the fundamental frequency of the signal is typically between 1-5Hz [8].

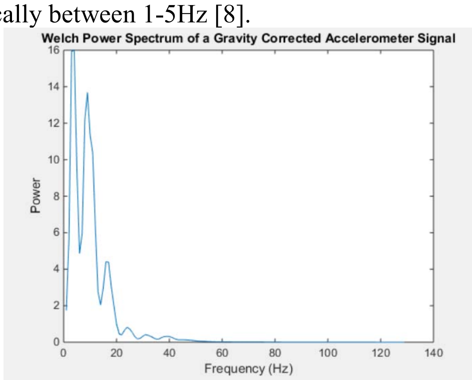


Figure 6 - PSD of Example Data Using Welch Algorithm

The frequency domain features of the signal that we extracted for gait classification were 1) the average power of the signal, 2) the ratio of high to low energy peaks of the PSD, 3) the signal to noise ratio of the signal, and 4) the total harmonic distortion of the signal. These features were selected based on their effectiveness in prior research involving passive gait verification [15] [8]. The average power of a signal is the mean of the total power underneath the curve of the PSD estimate for a signal [1].

Another useful frequency domain feature in gait classification is the ratio of the energy in high frequency peaks to the energy in low frequency peaks in the power spectral density estimate [8]. A value of 1 implies that there is an equal amount of energy in the high and low frequency peaks that are discovered by looking at the PSD estimate of the signal. Signal to Noise ratio (SNR) is a value typically calculated in decibels relative to the carrier (dBc) of a real-valued input signal. The SNR is used in radio broadcast transmission to express how much of the signal is distorted by noise in the airwaves. For an accelerometer, any outside force on the phone that is not directly attributable to walking is considered as noise. For instance, if the sidewalk the user is walking on is being excavated using a jackhammer. Mathematically, the SNR is determined using the periodogram of the signal. The periodogram was smoothed using a Kaiser window with β = 38. The Kaiser window is a windowing technique that is commonly used in digital signal processing, especially when smoothing periodograms. We considered the energy contained within the peaks of the first six harmonics (including the fundamental frequency) to be the signal. Thus, the power outside of these harmonics is considered to be the noise of the signal for the ratio value [1].

Another frequency domain feature that we selected was the Total Harmonic Distortion (THD). The THD is used in audio analysis to determine the amount of distortion the signal undergoes when being played through another source, such as a stereo speaker. On accelerometer data, the THD expressed how much the fundamental frequency of walking is distorted by other external factors such as hand movements and phone repositioning. The THD is calculated in dBc of a real-valued signal. It is determined from the fundamental frequency and the first five harmonics using a periodogram of the input signal from the phone [1]. It is most commonly defined as the ratio of the RMS amplitude of a set of higher harmonic frequencies to the RMS amplitude of the first harmonic, or fundamental, frequency [23]. Assuming V_1 is energy contained within the peak of the PSD at the fundamental frequency and V_i are the energy contained within the harmonics, THD can be calculated in general form using Equation 3. A lower THD (and higher SNR) indicate that that the signal is relatively clean with a small amount of contamination. Table 2 summarizes the frequency domain features we investigated .

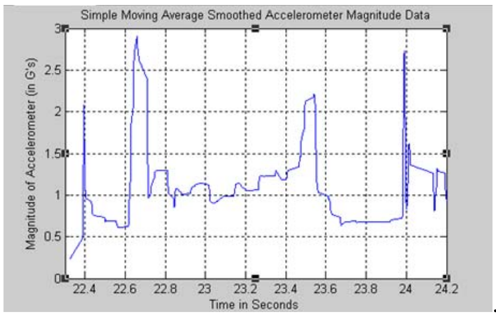
TABLE 2 – FREQUENCY DOMAIN GAIT FEATURES EXPLORED

Frequency Domain Feature	Definition
Average Power [8]	The variance per unit time
Ratio of Spectral Peaks [8]	Ratio of the energies of low and high frequency bands
SNR	Power of whole signal / power of its computed noise
THD	Distortion of the whole signal compared to its harmonics

$$THD_F = \frac{\sqrt{V_2^2 + V_3^2 + V_4^2 + V_5^2 + V_6^2}}{V_1} \quad (3)$$

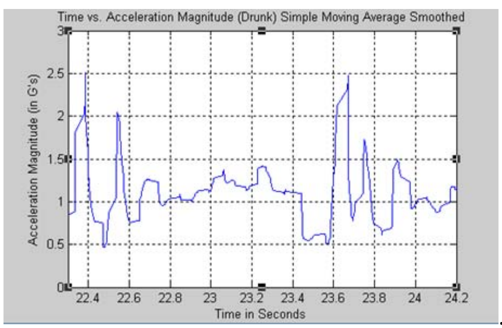
C. Results of Feature Extraction

Below are plots of pre-processed accelerometer data for sober gait (Figure 7) and the gait associated with an estimated BAC of 0.117, or > 6 drinks (Figure 8). Clear differences can be observed between the sober and intoxicated gait signals. The sober gait has more sharply defined "steps", while the intoxicated gait exhibits noise between steps. Table 3 lists the features that were extracted from the signals in figures 7 and 8



.

Figure 7 - MATLAB Plot of Sober Gait in Time Domain



.

Figure 8- MATLAB Plot of Intoxicated Gait in Time Domain

.TABLE 3 – SOBER VS INTOXICATED FEATURE COMPARISON

	Sober	Intoxicated
numSteps	12	12
cadence	1.1638	1.3327
skewness	1.6739	0.81458
kurtosis	6.1112	3.6834
gaitVelocity	0.096984	0.11106
stepLength	-1.9231	-1.9231
ratio	0.47392	0.79152
stepTime	3.6547	6.9889
avgPower	32307	13379
SNR	-2.9788	-5.1409
THD	-2.0745	-14.41
numDrinks	0	12

The number of steps and step length for this recording window (10 sec.) are the same. Interestingly though, the cadence and velocity were quicker when this particular study subject was walking. There is a noticeable difference in all of the features. The SNR measurements suggest that intoxicated gait produces a less noisy signal, and the THD measurements suggest that there is less distortion in the signal as the number of drinks increase. Figure 9 is a plot of all features investigated for varying levels of intoxication.

V. CLASSIFIERS OF ALCOHOL CONSUMPTION

To generate machine learning classifiers, we used our time and frequency domain features as inputs to the Weka machine learning engine. Of the 209 samples that were analyzed, the minimum number of drinks was 0 and the maximum number reported was 12. The mean of this data set was 4.643 drinks per subject, and the standard deviation was 3.896 drinks. Because of the mean and standard deviation of the data set, we decide to put the data into 3 bins in order to improve the accuracy of the model. These bins were 0-2 drinks, 3-6 drinks, and above 6 drinks consumed. After removing obviously wrong entries we were left with 61 instances of features for the 0-2 drink bin, 31 instances of the 3-6 drink bin and 48 instances of the >6 drink bin.

We compared the accuracy of the Naïve Bayes Net, the J48 Decision Tree, the Support Vector Machine, and the Random Forest machine learning classifiers. For each test run, five-fold cross validation was used to ensure that the model was not over-fitting or memorizing the data [21]. K-fold cross validation runs the classifier K times, but each time 1/K's worth of the dataset is withheld and used to test (or validate) the model [21]. By doing this, the model can be tested to ensure that it is actually predicting and not memorizing data.

The Random Forest classifier produced the most accurate classification of all of the investigated methods. A Random Forest is a collection of trees each considering a random number of features, a random first feature, and a random depth to find the best possible classification [21]. It had an accuracy of 56%, an F-score of 0.629, and an AROC of 0.658. Next, we analyzed other classifiers which could be selected for the model. The Naiive Bayes net performed worst of the four classifiers considered. Despite its simplicity, it was unable to draw useful inferences from the data about whether or not the presence of alcohol existed in gait. It had 42.1429% accuracy, a weighted F-score of 0.393, and a ROC area of 0.564.

We also explored the Support Vector Machine (SVM) classifier, which is implemented in Weka by using John Platt's Sequential Minimal Optimization (SMO) algorithm. Typically SVM's perform better in binary classification when not much is known about the problem domain. However, this data are not linearly separable. To overcome this, a technique called "kerneling" (shown in figure 1) is used to create the maximal separating hyperplane.

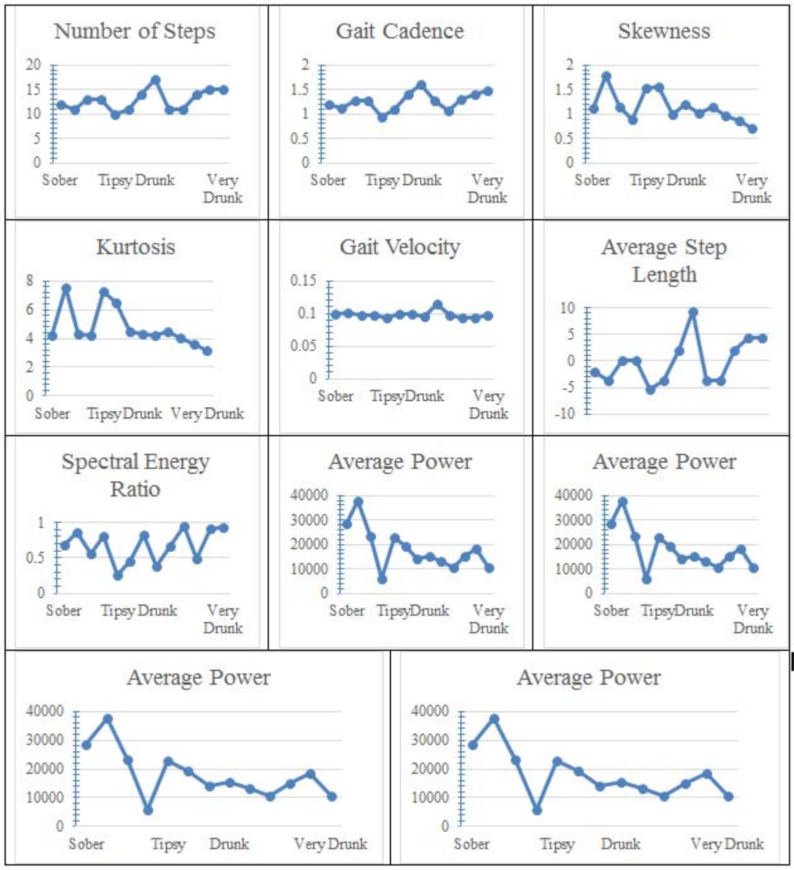


Figure 9 – Feature Values for Varying Levels of Intoxication

Kerneling is achieved using a similarity function that helps to separate instances of data in the input space and transform them into a linearly separable set in the feature space. By doing this, a maximal hyperplane distance can be calculated for non-linearly separable data. To handle multiple classes, the SVM in Weka uses a technique called "one-against-one." That is, for each class (sober, tipsy, and drunk) an SVM is trained for each pair of classes. For the alcohol dataset in our study, the SVM performed only slightly better than the Naïve Bayesian network.

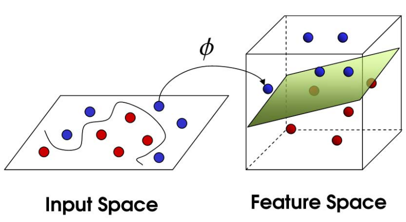


Figure 10 – A visual illustration of kerneling

The final classifier we considered was the J48 decision tree. The algorithm builds a decision tree from the training data that it is presented using the concept of information entropy. Information entropy is a quantitative measure of the amount of information contained in a single message (or single classified instance) [21]. For each feature of gait, the decision tree builder considers the feature that maximizes the information gain at that step. Information gain is the difference between the entropy of the tree at level $i+1$ and i. The decision tree seeks to minimize this value (thereby maximizing information gain) [21]. For the training data, the J48 tree performed second best of all of the classifiers, with an accuracy of 53.5714%, an F-score of 0.510, and an ROC Area of 0.646. Table 4 summarizes the results of all classifiers considered.

TABLE 4 – COMPARISON OF CLASSIFIER PERFORMANCE

Classifier	Accuracy %	F-Score	ROC Area
Naïve Bayes	42.1429	0.393	0.564
J48 Decision Tree	53.5714	0.510	0.646
SVM	47.1429	0.427	0.562
Random Forest	56.0000	0.629	0.658

Based on the previous results, the Random Forest was the best performing classifier and was selected to be implemented in AlcoGait, our real time alcohol inference app. After choosing this classifier, we investigated whether techniques such pruning, bagging, and boosting would improve its accuracy. This process is known as "ensembling." Bagging is a technique whereby the training data are sampled uniformly with replacement and trained. The output is the average of all classifications by the decision tree model being bagged. This technique is used to improve unstable dataset classification. From all of the ensembling techniques considered, a bagged Random Forest of 10,000 trees created the best model. It had an accuracy of 56.00%, an F-score of 0.629, and AROC of 0.658. Its training data confusion matrix is shown in Table 5. This model was loaded onto the phone for use in the algorithm that will classify gait.

TABLE 5 – RANDOM FOREST CONFUSION MATRIX

Ensemble Classifier	A (0-2 Drinks)	B (3-6 Drinks)	C (> 6 Drinks)
A (0-2 Drinks)	85	34	26
B (3-6 Drinks)	7	5	5
C (> 6 Drinks)	15	6	26

Overall, our results show promise but accuracy suffered slightly because of insufficient data. Since the majority of data was in the "sober" category, it makes sense that the classifier was able to accurately classify sober gait more than heavily intoxicated gait. It seems also that each classifier struggled to classify the 3-6 drink range. This could be due to both the lack of data for this bin and the fact that the feature response values are not well separated at the 3 and 6 drink mark.

After the classifier was selected, we verified its accuracy on a validation dataset previously unseen by the classifier. We used a sample of 30 instances to test the pre-trained model. It had an accuracy of 70.00%, an F-score of 0.786, and AROC of 0.825. Its validation data confusion matrix is shown in Table 6. This indicates the model is performing better than the cross-validated Random Forest.

TABLE 6 – VALIDATION SET CONFUSION MATRIX

Validation Set	A (0-2 Drinks)	B (3-6 Drinks)	C (> 6 Drinks)
A (0-2 Drinks)	1	0	9
B (3-6 Drinks)	0	6	0
C (> 6 Drinks)	0	0	14

The model performs well on predicting class C and appears to have a reduced number of errors on predicting class B when compared to the initial training confusion matrix. This could be due to the different proportions of instances in each bin between the training and validation sets. The hope is that it will improve its accuracy when more data is available.

VI. AlcoGait: Smartphone App to Infer Alcohol Consumption

Using the alcohol inference classifiers we generated above, a real time Android application was designed and developed. This app was a background service that continuously received updates from the Android Activity Recognition API. When walking was detected, the raw accelerometer data was fed into the background service's main thread in 5 second increments. If walking terminates, the data are discarded and the service waits for the next walking event. The implementation leveraged the GaitLib library's `GaitAnalysis()` method [22]. A maximum of three 5-second samples are gathered per hour to prevent overcrowding of data and to manage storage space on the device. Once a 5 second sample is captured and checked for continuity between readings, it is fed into the feature generator.

The feature generator is a JAR file exported from MATLAB that contains all of the functions written to generate features from the previous experiment and study. The JAR is called with raw accelerometer data and returns the list of features calculated from them in order. If the JAR does not return values for some of the features, the data are discarded and the service waits for a new dataset. Once the features are generated from a set, the accelerometer readings are discarded, and the generated features are saved as a new row in a database table locally on the phone.

These features are labeled on the following day by the user using the in-app survey which included when they began drinking, when they finished drinking, and how many drinks they had. The data are labeled with the first sample inside the window being the baseline "0" drink mark. The number of drinks is spaced out over the interval of time spent drinking and samples are labeled accordingly. In addition to labeling the window during which drinking occurred, the application also labels the period after drinking occurs based on the average rate of alcohol metabolization of 1 drink per hour. After this labeling process occurs, the model is retrained using 10-fold cross validation on the entire data set to date. Figure 11 illustrates the logic of our app.

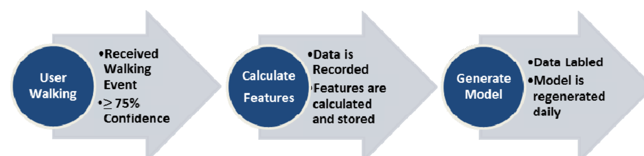


Figure 11 – Algorithm Model Flow

Once this training has been done, this model is loaded back into the application to be used to make inferences for the next 24 hours when it is updated again.

VII. Evaluation AlcoGait in User Studies

We conducted a usability study to evaluate how well the AlcoGait application inferred intoxication levels of real users and to gather user impressions. The application was installed on subjects' phones for a few days. They were then asked to complete a short usability survey. The volunteers were students of the WPI computer science department as well as students of a WPI sorority. The only restriction for participation was that the volunteer must have been over 21 years of age. The survey consisted of a brief description of the intended future use of the application once it has been fully implemented and is consumer ready. It asked if they thought the goal of this future application was meaningful and useful to society as well as if there were any improvements they thought would be beneficial for the future application.

A. Results of AlcoGait User Study

Approximately 86% of respondents found the app's vision useful when the application is complete. One of the users responded that they "found it to be very insightful." Another found it to be "functional and easy to use." All of users said they would recommend this application to a friend but only 57% would use the application themselves. When users were asked to rate the classifier's accuracy on their intoxication levels on a scale of 1-10 (with 10 being 100% accurate), the mean value was 8.86 with a standard deviation of 1.86. Table 7 includes responses to primary questions and the number of individuals who responded yes.

TABLE 7 – SUMMARY OF USER RESPONSES

Question	No. (percent) Responding 'Yes'
Do you see the envisioned app being useful when complete?	6 (86%)
Would you recommend the envisioned app to others when it is complete?	7 (100%)
Would you actually used this envisioned application?	4 (57%)
Did you like the current application?	5 (71%)
Do you like the appearance of the application?	6 (86%)

B. Personalization: Classification Improvements as users corrected wrong inferences

One of the most useful properties of supervised machine learning is the fact that classifiers tend to get more accurate when given more data. We speculated that after the initial loading of the model at application install time, classification would become more accurate if users corrected wrong inferences and as more data are available to train and test the model. The lower bound of classification is the 57% accuracy which comes with the application on installation. After only a few days' worth of use, the range of descriptive statistics retrieved from users in the usability study are expressed in Table 8.

TABLE 8 – RESULTS OF CLASSIFIER IMPROVEMENTS DUE TO PERSONALIZATION

User	Accuracy %	F-Score	ROC Area
User 1	58.435	0.597	0.612
User 2	52.784	0.513	0.538
User 3	62.947	0.640	0.631
User 4	67.249	0.651	0.683
User 5	73.854	0.754	0.749
User 6	33.333	0.410	0.551

Classification accuracy did improve for 66% of users but got worse for 33% of them. This reduction in accuracy could be due to abnormal walking conditions or accidental triggers of the recording mechanism. For example, shaking the phone rapidly in a rhythmic motion for 5 seconds engages the step detector which triggers our application to begin recording.

VIII. Discussion

The application on the phone produced an average of 57% accuracy in classification. This result is encouraging for an initial effort in the wild given that there is a large set of factors that influence gait along with alcohol. However, the result is not without flaw. The biggest problem is insufficient data. We gathered data for relatively few instances and not all bins were adequately covered. We believe that our app's accuracy will increase as we gather data for more people and over more time per user.

There are many ways in which this classification is not able to handle all edge cases. For starters, we assume that all drinking occurs in one sitting. In reality, drinks could be spaced out in many ways which all intoxicate the user differently. Different individuals also process alcohol differently. Other factors include whether food was consumed prior to or during drinking, and the type of alcohol consumed.

Specific to the classifiers, we noted that each technique was better at classifying the extremes (the 0-2 drink bin, and the > 6 drink bin) than classifying the 3-6 drink bin. We also noted that it was interesting that the validation set of data accuracy percentage (70%) was better than the overall model accuracy percentage (57 %). Over time, it seems that this result improved for some users of the application, and got worse for others. This could be due to gait conditions specific to some users, a mistake in the feature calculation, or an error in data collection. We noted that not all classifications improved over time, but after only a few days, most showed promise of improvements.

Additionally, a user's walk may change over time due to any number of circumstances including weather, phone placement, ground conditions or personal injury. A user's gait may also change if they are fatigued or in a bad mood. Ultimately, gathering more data to cover majority of cases and labeling any contributing factors should improve inference accuracy and robustness.

IX. Related Work

A. Alcohol Detection Devices

SCRAM: SCRAM Continuous Alcohol Monitoring [11] is a commercial alcohol detection device that is worn continuously around the ankle. It is mainly used for high-risk, DUI (Driving Under the Influence) alcohol offenders who have been ordered by a court to avoid consuming alcohol. It samples the user's perspiration every 30 minutes in order to measure their BAC levels. The user's data is sent to secure servers and can be accessed by court officials. The data is typically used to confirm abstinence from alcohol consumption and as evidence of good or bad behavior in court trials[11].

Kisai Intoxicated LCD Watch: The Kisai Intoxicated LCD Watch by TokyoFlash Japan [12] is a breathalyzer watch. In addition to being a normal watch, it has a built-in breathalyzer on its side that the user can use at any point. By simply breathing into the Breathalyzer, the watch determines and displays graphs of the user's BAC level.

X. Conclusion

In this paper, we investigated whether smartphones could infer the alcohol intoxication levels (how many drinks) of their users based on anomalies in their gait. Time and frequency domain features were extracted from accelerometer data of drinkers and used for classification in a machine learning framework. For a task of classifying the number of drinks consumed by a user into ranges of 0-2 drinks (sober), 3-6 drinks (tipsy) or >6 drinks (drunk), Random Forest yielded 56% accuracy on the training set, and 70% accuracy on the validation set. Using these results, AlcoGait an Android smartphone application was developed and deployed to real users. The results of user studies were encouraging.

In future, we would like to gather data from additional sensors including the gyroscope, GPS, bluetooth, the compass and other inertial sensors. Gathering more data from more users over longer periods will probably improve the accuracy of our models. This application could eventually be integrated into the healthcare system and used by alcoholics who are in therapy to generate accurate drinking records and associated contexts. Frequent partiers could also use it to check if they are too drunk to drive. Social networking could also be integrated so that drinkers could find each other, discuss their progress and support each other.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions of Joseph Petrucelli of the WPI department of mathematical sciences, who provided valuable advice and guidance in the statistical analysis of the data in this study.

REFERENCES

[1] A. V Oppenheim, and R W Schafer, Discrete-Time Signal Processing, 3rd Edition, Prentice Hall.

[2] World Health Organization, "Global status report on alcohol and health," WHO, 2014. URL: http://www.who.int/substance_abuse/publications/global_alcohol_report/msb_gsr_2014_1.pdf?ua=1.

[3] K.-C. Wang, Y.-H. Hsieh, C.-H. Yen, C.-W. You, M.-C. Huang, C.-H. Lee, S.-Y. Lau, H.-L. Kao, H.-H. Chu and M.-S. Chen, "SoberDiary: A Phone-based Support System for Assisting Recovery from Alcohol Dependence," in Proc Ubicomp 2013.

[4] J. Rehm, T. K. Greenfield and J. D. Rogers, "Average Volume of Alcohol Consumption, Patterns of Drinking, and All-Cause Mortality: Results from the US National Alcohol Survey," *American Journal of Epidemiology,* vol. 153, no. 1, pp. 64-71, 2001.

[5] J. Westmaas, S. Moeller and P. B. Woicik, "Validation of a Measure of College Students' Intoxicated Behaviors: Associations With Alcohol Outcome Expectancies, Drinking Motives, and Personality," *Journal American College Health,* vol. 55, no. 4, pp. 227-237, 2007.

[6] S. Demura and M. Uchiyama, "Influence of moderate alcohol ingestion on gait," *Sport Sci Health,* no. 4, pp. 21-26, 2008.

[7] S. Ando, T. Iwata, H. Ishikawa, M. Dakeishi and K. Murata, "Effects of acute alcohol ingestion on neuromotor functions," *NeuroToxicology,* vol. 29, pp. 735-739, 2008.

[8] H. Lu, J. Huang, T. Saha and L. Nachman, "Unobtrusive Gait Verification for Mobile Phones," in *ISWC*, Seattle, WA, USA, 2014.

[9] Centers for Disease Control and Prevention, "Alcohol and Public Health," Centers for Disease Control and Prevention, 19 August 2014. http://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm.

[10] M. Izzi, "SCRAM Bracelet Laws," LegalMatch, 3 Sept 2014. http://www.legalmatch.com/law-library/article/scram-bracelet-laws.html.

[11] SCRAM Systems, "SCRAM Continuous Alcohol Monitoring," Alcohol Monitoring Systems, Inc., 2014. http://www. scramsystems.com/index/scram/continuous-alcohol-monitoring.

[12] Tokyoflash Japan, "Kisai Intoxicated LCD Watch," Tokyoflash Japan, 2014. : http://www. tokyoflash.com/en/watches/kisai/intoxicated/.

[13] E. R. Weaver, D. R. Horyniak, R. Jenkinson, P. Dietze and M. S. Lim, ""Let's get Wasted!" and Other Apps: Characteristics, Acceptability, and Use of Alcohol-Related Smartphone Applications," *JMIR MHEALTH AND UHEALTH,* vol. 1, no. 1, pp. 1-11, 2013.

[14] M. Nieschalk, C. Ortmann, A. West, F. Schmäl, W. Stoll and G. Fechner, "Effects of Alcohol on body-sway patterns in human subjects," *Int'l Journal of Legal Medicine,* vol. 112, pp. 253-60, 1999.

[15] H.-L. Kao, B.-J. Ho, A. C. Lin and H.-H. Chu, "Phone-based Gait Analysis to Detect Alcohol Usage," in Proc ACM Ubicomp 2012.

[16] K. R. Hirsch, "The 19 Best Alcoholism iPhone & Android Apps of 2014," Healthline Networks, 21 May 2014. http://www.healthline.com/health-slideshow/top-alcoholism-iphone-android-apps#1.

[17] N. Pell, "The 9 Best Alcohol-Related Apps," Made Man, 30 December 2011. [Online]. Available: http://www.mademan.com/the-9-best-alcohol-related-apps/.

[18] "Sports Injury Clinic," 2014. http://www.sportsinjuryclinic.net/sports-specific/running-injuries/gait-analysis. [Accessed 20 March 2014].

[19] E. Sejdić, K. A. Lowry, . J. Bellanca, M. S. Redfern and J. S. Brach, "A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains.," *IEEE Trans Neural Systems and Rehabilitation Engineering,* vol. 22, no. 3, pp. 603-612, 2014.

[20] Mathworks, "Counting Steps by Capturing Acceleration Data from Your Android™ Device," Mathworks, Mar 1 2015. http://www.mathworks.com/help/supportpkg/mobilesensor/examples/counting-steps-by-capturing-acceleration-data-from-your-android-device.html?prodcode=ML. [Accessed 22 March 2015].

[21] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Upper Saddle River: Prentice Hall/Pearson.

M. M. A. Wu, O. S. Schneider, I. Karuei, L. A. Leong and K. E. MacLean, "Introducing GaitLib: A Library for Real-time Gait Analysis in Smartphones," 2014. https://circle.ubc.ca/bitstream/handle/2429/46848/gaitlib-paper.pdf?sequence=1.

[22] I. V. Blagouchine, "Analytic Method for the Computation of theTotal Harmonic Distortion by theCauchy Method of Residues," *IEEE Trams Communication,* vol. 59, no. 9, pp. 2478-2491, 2011.

[23] M Karg, K Kuhnlenz, M Buss, "Recognition of Affect Based on Gait Patterns," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.40, no.4, pp.1050,1061, Aug. 2010

[24] M Karg, G Venture, J Hoey, D Kulic, "Human Movement Analysis as a Measure for Fatigue: A Hidden Markov-Based Approach," Neural Systems and Rehabilitation Engineering, IEEE Trans, vol.22, no.3, pp.470,481, May 2014

[25] A K Dey, K Wac, D Ferreira, K Tassini, J Hong, J Ramos, Getting closer: an empirical investigation of the proximity of user to their smart phones, in Proc Ubicomp 2011.