Investigating Postural Sway Features, Normalization and Personalization in Detecting Blood Alcohol Levels of Smartphone Users

Christina Aiello, Emmanuel Agu, *Member, IEEE*

***Abstract*—Alcohol abuse causes 88,000 deaths annually. In this paper, we investigate a machine learning method to detect a drinker's Blood Alcohol Content (BAC) by classifying accelerometer and gyroscope sensor data gathered from their smartphone. Using data gathered from 34 "intoxicated" subjects, we generated time and frequency domain features such as sway area (gyroscope) and cadence (accelerometer), which were classified using supervised machine learning. Our work is the first to classify sway features such as sway area and sway volume, which are extracted from the smartphone's gyroscope in addition to accelerometer features. Other novel contributions explored include feature normalization to account for differences in walking styles and automatic outlier elimination to reduce the effect of accidental falls. We found that the J48 classifier was the most accurate, classifying user gait patterns into BAC ranges of [0.00-0.08), [0.08-0.15), [0.15-0.25), [0.25+) with an accuracy of 89.45% (24.89% more accurate than using only accelerometer features as in prior work). Our classification model was used to build AlcoGait, an intelligent smartphone app that detects drinkers' intoxication levels in real time.**

I. INTRODUCTION

Alcohol abuse results in physical harm, mental malfunction [10] and is responsible for 1 in 10 deaths among adults aged 20-64 years in the United States annually [19]. Despite these effects, binge drinking (defined as 4 or more drinks for women on a single occasion, and 5 or more drinks for men on a single occasion [19]) has been on the rise. Between 2002 and 2005, over a third of college students aged 18-20 reported binge drinking [17].

Drunk driving endangers not only the intoxicated driver but also pedestrians and other sober drivers [20]. In 2010, 47.2% of pedestrian fatalities and 39.9% of vehicle occupant fatalities were caused by drunken driving [18]. However, in many Driving Under the Influence (DUI) cases, the drinker is unaware that they are over the legal driving limit. In this paper, we focus on a method for a drinker's smartphone to passively sense their intoxication level from their gait (walk). Passive methods to continuously detect and log a drinker's intoxication level can be used in multiple ways to either treat hard drinkers or to prevent alcohol-related mishaps. *Use in prevention:* Drinkers who are over the legal driving limit can receive just-in-time notifications of excessive alcohol consumption, preventing drunk driving. Preventive monitoring is timely since DUI offenses carry severe consequences including license suspension, fines, high insurance premiums and even jail time [37]. *Use in treatment:* a smartphone can log a frequent drinker's drinking patterns and associated contexts (e.g. time, place, who with). Drinkers can reflect on their drinking logs, detect patterns of abuse and either self-correct or seek treatment. Counselors can use such logs as evidence to prescribe treatment. In cases where the drinker loses consciousness, emergency room physicians would have an accurate record of a patient's drinking.

Alcohol consumption raises the Blood Alcohol Content (BAC) of the drinker's blood [24]. Alcohol penetrates the blood-brain barrier approximately 20 minutes after alcohol consumption, impacting neuromotor and cognitive functions [25]. Gait, or the manner in which a person walks is one of the neuromotor functions affected by alcohol consumption. In fact, aside from direct BAC or BrAC testing, neuromotor testing including analysis of gait is the most reliable way to determine intoxication in humans [8]. Leveraging this fact, the walk-and-turn field sobriety test used by the police is based on gait assessment.

In this paper, we investigate a method to detect the BAC level of a drinker from their gait by classifying accelerometer and gyroscope sensor features gathered from their smartphone using a machine learning approach. This paper significantly improves on our previously proposed AlcoGait intoxication inference smartphone app [15] in several ways, yielding an accuracy of 89.45% for classifying user gait patterns into BAC ranges of [0.00-0.08), [0.08-0.15), [0.15-0.25), [0.25+). Our initial AlcoGait app only had a classification accuracy of 57% for the task of classifying the number of drinks a user had consumed into 3 bins: 0-2 drinks (sober), 3-6 drinks (tipsy) and > 6 drinks (drunk). Specific limitations in our prior effort that are now addressed in this paper include:

1. *Explore Gyroscope features:* Time and frequency domain features were generated previously using only smartphone accelerometer data. This paper explores the use of smartphone gyroscope features in addition to accelerometer features. Ando *et al* [8] found that subjects' postural sway increased after they ingested alcohol. Nieschalk *et al* [3] determined that sway area (a gait attributes widely used in posturography) was the most sensitive attribute for detecting increased body sway after alcohol ingestion. Posturography is a clinically validated approach for assessing balance disorders from gait. We synthesized gyroscope sway features, which when combined with accelerometer features yielded 89.45% classification accuracy (24.89% more accurate than accelerometer features alone as in prior work).
2. *Including physical attributes of subjects as features:* Other alcohol-sensitive physical attributes such as subject weight, height and gender were included in our classification model.
3. *Gather data from more users:* Our previous AlcoGait work gathered smartphone accelerometer data from only 7 participants, which resulted in inaccurate machine

learning models. In this paper, data was gathered from 34 participants wearing sensor-impairment goggles that simulate the effects of intoxication.

4. *Exploring feature Normalization:* to reduce inaccuracies caused by different walking styles. Normalization prevents a person whose sober gait sways a lot from being misclassified as being drunk.

5. *Investigating personalization (per-user machine learning models) and ensembling:* to compare per-user classification models to models trained on large numbers of subjects.

6. *Alcohol measured using BAC Level rather than the number of drinks:* The intoxication of subjects was previously measured in terms of "number of standard drinks", which was imprecise. In this paper, we measure subject intoxication in terms of BAC (the same measure as a breathalyzer).

With over 2 billion individuals owning smartphones in 2016, smartphone apps for preventing alcohol abuse can become widely used and have great impact [26].

II. RELATED WORK

Intoxication detection devices: SCRAM: SCRAM Continuous Alcohol Monitoring [4] is an ankle-worn commercial alcohol detection device. It is mainly used for high-risk, Driving Under the Influence (DUI) alcohol offenders who have been ordered by a court not to consume alcohol. It samples the user's perspiration every 30 minutes in order to measure their BAC levels. *Kisai Intoxicated LCD Watch:* by TokyoFlash Japan [2] is a breathalyzer watch. In addition to being a normal watch, it has a built-in breathalyzer on its side. By breathing into its Breathalyzer, the watch determines and displays graphs of the user's BAC level. These are dedicated devices that must be purchased and used, unlike smartphones which users already own.

Intoxication detection using smartwatches: As smartwatches have emerged, attempts have been made to utilize them to detect alcohol consumption levels [18]. Gutierrez et al. [1] estimated if individuals wearing smart watches were intoxicated (BAC > 0.065) using Microsoft Band smartwatch sensor data. These investigators did not use accelerometer and gyroscope data while subjects were walking. Instead, they utilized the participants' heart rates and temperatures. Moreover, smartwatch based detection will likely never be as ubiquitous as smartphones since less than a quarter of people wear wristwatches [7].

Alcohol-related smartphone apps: Several alcohol-related apps exist on the iPhone and Android app markets. Existing smartphone applications targeting alcohol abuse allow users to manually record their alcohol consumption, estimate Blood Alcohol (BAC) levels using built-in formulas, and offer manual cognition tests to assess users' intoxication levels. Other applications attempt to encourage positive drinking habits. The smartphone application "Intoxicheck" can detect alcohol impairment in users [5]. Users take a "series of reaction, judgment and memory challenges before and after drinking, which are compared to estimate their intoxication level. However, Intoxicheck usage requires manual supervision, which may deter adoption and reduce its scalability.

Tjondronegoro et al. [21] describe a mobile social smartphone application designed to encourage positive drinking habits in users. The application's goal was to encourage drinkers to drink in groups and look out for each other. However, their app did not include any passive intoxication detection. Wang et al. [23] developed SoberDiary, a phone-based support system that lets individuals recovering from alcohol addiction self-manage and self-monitor their alcohol consumption over time. Users breathalyzed themselves using a portable breathalyzer that sent data to their smartphone. While SoberDiary was successful in reducing heavy drinking, it did not auto-detect intoxication from gait.

Intoxication-detection from gait: Kao et al. [9] designed a passive phone-based system that used the smartphone's accelerometer data to detect whether users had consumed alcohol or not (Yes/No), but did not try to estimate how much (BAC level or number of drinks) was consumed. Arnold *et al* [15] created an app to detect intoxication from gait data. However, unlike this work, they did not utilize postural sway features extracted from gyroscope data, or use normalization to account for different walking styles. Moreover, intoxication was estimated in terms of number of drinks consumed rather than BAC and their reported accuracy was 57% compared to 89% achieved in this paper.

III. METHODOLOGY

Our methodology followed a typical machine learning classification approach with a flow diagram illustrated in figure 1. Smartphone gyroscope and accelerometer features were gathered from 34 subjects as they walked while wearing special goggles designed to simulate intoxication. Extremal (outlier) data values were eliminated to reduce noise. Gyroscope and accelerometer features that are sensitive to alcohol consumption were extracted in the time and frequency domain. Feature normalization was also explored to account for differences in walking styles. The accuracy and performance of various classifier types (e.g. random forest, JVM, Naïve Bayes) were compared and a custom classification model was generated using Ensembling. This classification model was used to create an Android app that detected the BAC levels of smartphone users in real-time from their gait while they walked.

A. Data Gathering Study

Thirty four (34) participants (14 male and 20 female) were recruited via a pool of psychology students who receive academic credit for participating in user studies. Subjects were also recruited via email advertisements, social media advertisements, and word-of-mouth.

Sensor-Impairment Goggles: Subjects wore sensor-impairment goggles and walked while accelerometer and gyroscope sensor data was collected. These "Drunk Busters" goggles use vision distortion to simulate the effects of alcohol consumption on the body [6]. Goggles rated at various Blood Alcohol Concentration (BAC) levels simulated the corresponding impairment causing wearers to experience intoxication effects including reduced alertness, delayed reaction time, confusion, visual distortion, alteration of depth and distance perception, reduced peripheral vision, double vision, and lack of muscle coordination [6]. These goggles have been used to educate individuals regarding the effects of alcohol on one's motor skills. It is instructive to note that each alcohol goggle simulates an approximate BAC range (e.g. 0.08 – 0.15) and different people could move slightly differently while wearing the same goggles.

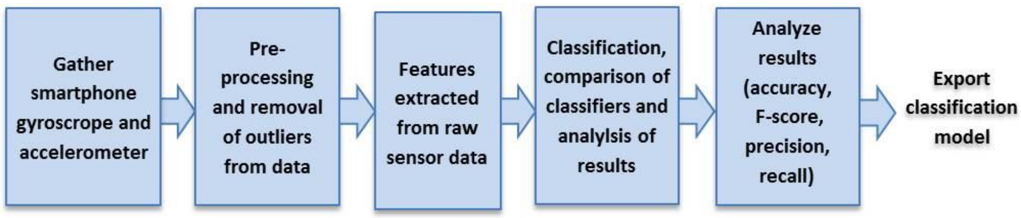

Figure 1 Flow Diagram for data collection, feature extraction and classification

Figure 2 Participants walking while wearing Drunk Busters Goggles

Study Procedure: In an IRB-approved study, participants placed an Android smartphone in either their front or back pants pocket. The participants then walked normally (no impairment) for 90 seconds at a comfortable speed, while a data gathering app running MATLAB mobile recorded the smartphone's gyroscope and accelerometer data (figure 2). Subjects then repeated the 90-second walk while wearing goggles rated at BAC of [0.04-0.06, 0.08 – 0.15, 0.15-0.25, 0.25 – 0.35]. To explore the effects of different walking speeds on the gathered data, a few subjects were made to walk at various speeds while sensor data was collected.

Reproducibility advantage of drunk busters goggles: Our study design using intoxication goggles is novel. In our prior work [15], accelerometer data was gathered from subjects after they drank (e.g. during an evening at a bar). The next day at noon, subjects then self-reported the quantity and timing of alcohol consumption. We found that the previously collected data collected was quite noisy with participants having difficulty in accurately estimating the number and timing of drinks consumed. These issues were mitigated by using drunk busters goggles to collect training data in a controlled study.

B. Pre-Processing (including outlier removal)

Gyroscope and accelerometer data was gathered from subjects and stored in segments of 5 seconds. As such no further segmentation was required. However, subjects may trip or fall while intoxicated, which would generate extremal gyroscope and accelerometer data values (outliers). We synthesized a simple outlier removal algorithm by sorting the accelerometer and gyroscope data and removing the top and bottom 1 percent of values on the x, y and z axes.

C. Feature extraction

Gyroscope features: The gyroscope's x, y, and z axes (figure 3 (left)) can be directly related to the three body axes, which are the mediolateral, anteroposterior, and superoinferior axes (Figure 3 (right)).

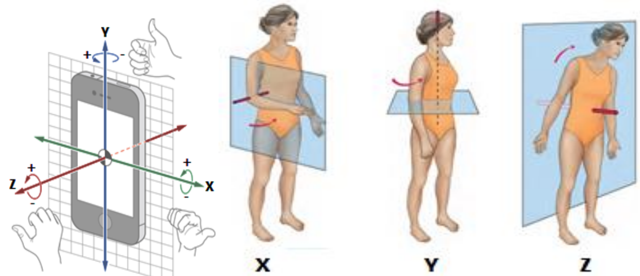

Figure 3 The three gyroscope rotation axes in Android devices and their relationship with the three axes of the body

In prior work, Nieschalk *et al* [3] found that a gait attribute called sway area, which was measured using a force plate, was the most sensitive attribute for detecting increased body sway after subjects ingested alcohol. While much of prior work on sway area and posturography has utilized accelerometers and force plate measurements in clinics, Kaewkannate [27] showed that gyroscope features can accurately capture posturography variables (including sway area). To the best of our knowledge, our work is the first to utilize sway area and postural sway features of a smartphone gyroscope as features for machine learning classification of BAC levels [3].

The smartphone gyroscope sensor returns the rate of rotation around the smartphone's X, Y and Z axes in radians per second. Sway area is calculated by plotting values from two of the gyroscope's axes (figure 3 (top)). For the XZ sway area, all observed gyroscope X and Z values in a segment were projected unto an X-Z plane (see figure 4(top)). The area of an ellipse that encloses the 95 percent confidence interval of all observed points was returned as the XZ sway area. This methodology is similar to that used for calculating sway area using force plate readings [3]. However, our use of the gyroscope to synthesize these sway areas is novel. As a further contribution, we synthesized gyroscope-based sway volume, the 3D analogue of sway area as a novel feature explored (See figure 4 (bottom)). Sway volume is the sphere that contains the 95 percent confidence interval of all X, Y, Z points in a segment. Accelerometer and gyroscope features were generated from the sensor data gathered from all study participants. Table 1 lists gyroscope features extracted and their formulas, while table 2 lists accelerometer features extracted and their formulas.

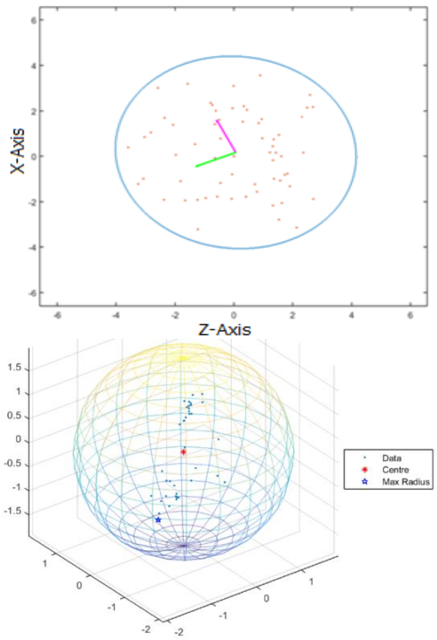

Figure 4 XZ Sway area (top) and Sway Volume (bottom) Plots

Feature Normalization: No two individuals walk exactly the same, especially while intoxicated. To minimize such inter-person differences in gait patterns, users' sway area and sway volume features were normalized per person. This task involves dividing each subject's intoxicated sway area by the average of sway area values calculated while they were sober. The formulas for average sway area and normalized sway area can be seen below (equations 1 & 2).

$$Average\ sway\ area = \frac{\Sigma\,(sober\ sway\ area\ calculations)}{\#\ sober\ readings} \quad (1)$$

$$Normalized\ sway\ area = \frac{(current\ sample's\ sway\ area)}{average\ sway\ area} \quad (2)$$

D. Machine Learning Classification

All gyroscope and accelerometer features were imported into the Weka machine learning library to explore classification. The accuracy of various classifiers were compared.

User-Specific Classification Model Using Ensembling: A general machine learning model trained from data gathered from all users was used as the default in our classification model. Per-user models (i.e. classification models trained for an individual using only their data). We combined per-user data with data for all users using ensembling. Ensembling methods are "learning algorithms that construct a set of classifiers and then classify new data points by taking a weighted vote of their predictions" [11]. Possible improvements using ensembling were also explored.

IV. RESULTS

We now present results of analyses and classification of gait data from our 34 participants (20 female and 14 male). Participants' heights ranged from 150cm to 200cm (mean = 172cm, s = 10.22cm), weights ranged from 100lbs to 250lbs (mean = 155lbs, s = 31.96lbs), and ages ranged from 18 to 22 (mean = 20 years, s = of 1.32 years).

We plotted and visually examined all features. Figure 5 shows that gyroscope sway areas (1 participant) generally increased as they became more intoxicated, as expected. Normalization generally compressed the ranges of sway area boxplots and separated them more (more statistically significant), which improved classification accuracy.

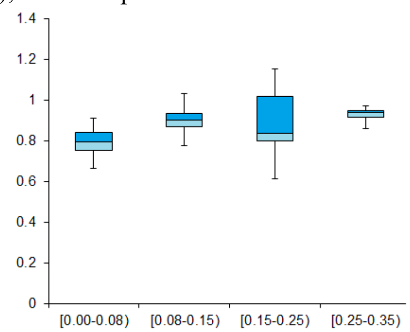

Figure 5 Boxplots showing gyroscope sway area as participants became more intoxicated (normalized)

A. Feature Exploration using Correlation-Based Feature Selection (CFS)

To quantify the predictive value of each extracted feature, we used Correlation-Based Feature Selection (CFS) [13] wherein each feature's correlation with the subject's BAC level and p-value are computed. The features that are most strongly correlated with BAC levels (p-value < 0.05) have the highest predictive value. Figure 6 is a table of p-values and correlation coefficients for all our features. Features that are statistically significant (p-value < 0.05) are highlighted in green.

B. Results of Classification

Using the Weka machine learning library, the extracted features (normalized vs non-normalized) were classified into the labeled ranges of [0.04 – 0.06, 0.08-0.15, 0.15-0.25, 0.25-0.35]. The accuracy of the J48, JRip, Bayes Net, Random Forest, Random Tree, and Bagging classification algorithms were compared. The classification accuracy of normalized vs. not normalized data were investigated, in addition to removing certain generated features (all gyroscope features, all accelerometer features, certain gyroscope features, and certain attributes describing the participant such as height, weight, and gender). To evaluate the performance of a classifier, its percent accurate, precision, recall, F-measure, and ROC area [15] were evaluated. Figure 7 below summarizes our main results. Figure 10 shows all our results. Our main results were:

i. *All features with normalization:* When classifying all generated features from both sensors (with sway areas and volume normalized), ID, height, weight, and gender, the J48 classifier using percentage split, 99% train and 1% test had the **highest accuracy of 89.45%** and an ROC area of 0.916.

ii. *All features, no normalization:* When classifying all generated features from both sensors (with sway areas and volume not normalized), ID, height, weight, and gender, the J48 classifier using percentage split, 99% train and 1% test had the **highest accuracy of 88.89%**.

iii. *Accelerometer features only:* When classifying all generated features from just the accelerometer, Random Forest using cross-validation, 10 folds had the **highest accuracy of 64.56%** (24.89% less accurate than all features+ normalization) and an ROC area of 0.851.

iv. *Gyroscope features only:* When classifying all generated features from just the gyroscope, Random Forest using cross-validation, 10 folds had **the highest accuracy of 75.79%** (13.66% less accurate than all features+ normalization) and an ROC area of 0.919.

P-Value	Correlation Coeff	Feature
0.21537	-0.02881	Steps
0.21441	-0.02887	Cadence
4.26E-12	-0.1601	Skew
3.71E-10	-0.14496	Kurt
0.002902	-0.06918	Gait Velocity
0.2366	-0.02752	Residual Step Length
NaN	NaN	Ratio
8.22E-06	0.10344	Residual Step Time
2.38E-18	-0.20114	Bandpower
NaN	NaN	SNR
NaN	NaN	THD
8.49E-18	-0.19788	XZ Sway
6.71E-13	-0.16596	XY Sway
2.87E-24	-0.23314	YZ Sway
3.59E-08	-0.12763	Sway Volume
0.95949	0.001181	Weight
0.71924	-0.00836	Gender
0.69053	0.00926	Height
0.29829	0.024188	Participant ID

Figure 6 P-values and correlation coefficients for all features (statistically significant values are shaded in green)

Including participant gender, weight and height as features: Since intoxication is affected by weight, height and gender we explored including them as features in our classification. Including weight, height and gender as features improved classification accuracy. For example, when classifying the normalized data using Random Forest with percentage split (66% train, 33% test), including gender, height, and weight was 5.3% more accurate than using features and weight.

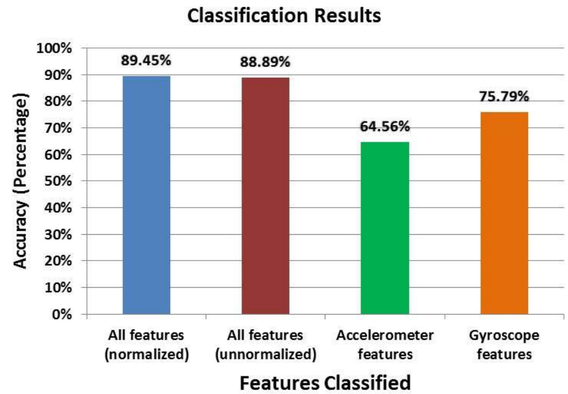

Figure 7 Classification Accuracy of features into [0.04-0.06, 0.08-0.15, 0.15-0.25, 0.25-0.35] ranges

Comparison of classifier types: The J48 and Random Forest classifiers were 21% to 39% more accurate in classification compared to the other classifiers explored (Random Tree, Bagging, JRip, and Bayes Net).

Investigating personalization: We investigated the idea of personalization, which involved training classifiers using only single user's gait data. Unfortunately our results were inconclusive: personalization improved classification accuracy for some subjects, but worsened it for others. Figure 9 shows the results of our personalization exploration, with rows highlighted in green if personalization was more accurate than our general model.

V. ALCOGAIT SYSTEM DESIGN AND IMPLEMENTATION

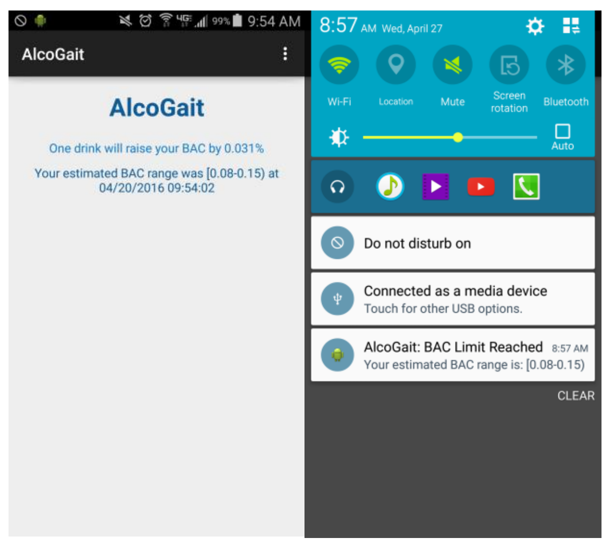

Figure 8 AlcoGait app Main screen (left) Notification sent to user when they have reached their pre-set BAC limit (right)

Using our classification model trained with data from our 34 subjects, we developed AlcoGait, an Android application that detected the BAC level of smartphone users in real-time based on their gait data. Figure 8 shows the main screen of our AlcoGait app (left) and a notification delivered to the smartphone owner when they have reached their pre-set BAC limit (right).

CONCLUSION AND FUTURE WORK

Alcohol abuse kills 88,000 people annually in the United States. In this paper, we investigate a machine learning method to detect the BAC level of a drinker by classifying accelerometer and gyroscope sensor features gathered from their smartphone. Our work is the first to classify body sway features such as sway area and sway volume, which are extracted from the smartphone's gyroscope in addition to accelerometer features. We found that the J48 classifier was the most accurate, classifying user gait patterns into BAC ranges of [0.00-0.08), [0.08-0.15), [0.15-0.25), [0.25+) with an accuracy of 89.45% (24.89% more accurate than accelerometer features only as in prior work). As future work, we would like to gather intoxicated gait data for a large number of subjects who are also breathalyzed and explore the effects of differences in alcohol tolerances, walking patterns and confounding factors such as fatigue and mood, which also affect gait.

Table 1: Features Generated from Gyroscope Data

Feature Name	Feature Description	Formula	Reference
XZ Sway Area	Area of projected gyroscope readings from Z (yaw) and X (pitch) axes	$XZ\ Sway\ Area = \pi r^2$	[3]
YZ Sway Area	Area of projected gyroscope readings from Z (yaw) and Y (roll) axes	$YZ\ Sway\ Area = \pi r^2$	Our contribution
XY Sway Area	Area of projected gyroscope readings from X (pitch) and Y (roll) axes	$XY\ Sway\ Area = \pi r^2$	[13, 14]
Sway Volume	Volume of projected gyroscope readings from all three axes (pitch, roll, yaw)	$Sway\ Volume = \frac{4}{3}\pi r^3$	Our contribution

Table 2: Features Generated from Accelerometer Data

Feature Name	Feature Description	Calculation	Ref.
Steps	Number of steps taken	calculation of signal peaks above one standard deviation away from mean of gravity corrected magnitude of signal [15]	[15]
Cadence	Number of steps taken per minute	$cadence = \frac{\#\ steps}{minute}$	[15]
Skew	Lack of symmetry in one's walking pattern	$skewness = \frac{\frac{1}{n}\sum(x_i - \mu_x)^3}{\left[\frac{1}{n}\sum(x_i - \mu_x)^2\right]^{3/2}}$ Where x_i is the data sequence, and μ_x is the average of all x_i [16]	[15]
Kurtosis	Measure of how outlier-prone a distribution is	$kurtosis = \frac{\frac{1}{n}\sum(x_i - \mu_x)^4}{\left[\frac{1}{n}\sum(x_i - \mu_x)^2\right]^{2}}$ Where x_i is the data sequence, and μ_x is the average of all x_i [16]	[15]
Average gait velocity	Average steps per second divided by average step length	average gait velocity $= \frac{(\ average\ steps\ /\ sec\ )}{step\ length}$	[15]
Residual step length	Difference from the average in the length of each step	residual step length $= \frac{distance}{\#\ steps}$	[15]
Ratio	Ratio of high and low frequencies	$harmonic\ ratio = \frac{\sum_{i=1,3,5,\ldots} V_i}{\sum_{j=2,4,6,\ldots} V_j}$ Where V_i is the amplitude of odd-ordered harmonic frequency and V_j is the even-	[15]

		ordered harmonic frequency [16]	
Residual step time	Difference in the time of each step	$\text{residual step time} = \frac{\sqrt{\frac{1}{n}\Sigma(interval_i - \mu_{interval})^2}}{\mu_{interval}}$ Where $interval_i$ is a sequence of stride intervals and $\mu_{interval}$ is average of all $interval_i$ [16]	[15]
Bandpowe r	Average power in the input signal	$\text{bandpower} = \text{bandpower(x)}$ Where x is a matrix of the magnitude, and bandpower calculates the average power in each column independently [14]	[15]
Signal to noise ratio	Estimated level of noise within the data	$snr = \frac{power_{signal}}{power_{noise}}$ [16]	[15]
Total harmonic distortion	"Determined from the fundamental frequency and the first five harmonics using a modified periodogram of the same length as the input signal" [22]	$thd = \frac{\sqrt{\sum_{i=2,3,4,5} V_i^{\,2}}}{V_1}$ Where V_1 is energy contained within peak of PSD at the fundamental frequency and V_i are the energy contained within the harmonics [15]	[15]

	Classifier	Configurations	Accuracy	ROC Area
219171	J48	Cross-validation 10 folds	76.70%	0.843
219171	J48	Percent Split 66% train, 33% test	75%	0.768
219171	Random Forest	Cross-validation 10 folds	80%	0.967
219171	Random Forest	Percent Split 66% train, 33% test	75%	0.915
1506627	J48	Cross-validation 10 folds	93.30%	0.976
1506627	J48	Percent Split 66% train, 33% test	75%	0.773
1506627	Random Forest	Cross-validation 10 folds	97%	0.998
1506627	Random Forest	Percent Split 66% train, 33% test	75%	0.964
1520109	J48	Cross-validation 10 folds	76%	0.86
1520109	J48	Percent Split 66% train, 33% test	80%	0.888
1520109	Random Forest	Cross-validation 10 folds	90%	0.942
1520109	Random Forest	Percent Split 66% train, 33% test	85%	0.972

Figure 3 Results of exploring personalization. Users (highlighted in green) showed improvement with personalization

Classification Configuration			Results					
Attributes Classified	Classifier	Test Set	Accuracy When Normalized	Precision	Recall	F-Measure	ROC Area	Accuracy (Unnormalized)
Accelerometer features, gyroscope features, ID, height, weight, gender	J48	Cross-validation, 10 folds	69.53%	0.695	0.695	0.695	0.817	69.26%
	J48	Percentage split, 66% train 33% test	63.28%	0.63	0.633	0.631	0.786	65.74%
	J48	Percentage split, 95% train 5% test	73.12%	0.735	0.731	0.731	0.835	72.22%
	J48	Percentage split, 99% train 1% test	89.45%	0.912	0.895	0.895	0.916	88.89%
	Random Forest	Percentage split, 66% train 33% test	72.66%	0.723	0.727	0.721	0.892	69.18%
	Random Forest	Percentage split, 95% train 5% test	81.72%	0.816	0.817	0.809	0.924	75.56%
	Random	Percentage split, 99%	73.68%	667	0.737	0.674	0.946	77.78%

	Forest	train 1% test						
	Random Forest	Cross-validation, 10 folds	73.74%	0.735	0.737	0.731	0.91	74.79%
	Random Tree	Cross-validation, 10 folds	67.69%	0.68	0.68	0.68	0.782	66.26%
	Random Tree	Percentage split, 66% train 33% test	66.77%	0.672	0.668	0.669	0.771	63.44%
	JRip	Cross-validation, 10 folds	50.29%	0.503	0.503	0.435	0.616	50.31%
	Bayes Net	Cross-validation, 10 folds	43.60%	0.405	0.436	0.41	0.696	44.34%
	Bayes Net	Percentage split, 66% train 33% test	45.47%	0.348	0.455	0.386	0.691	42.46%
	Bagging	Cross-validation, 10 folds	67.53%	0.673	0.675	0.674	0.792	70.94%
	Bagging	Percentage split, 66% train 33% test	60.25%	0.606	0.603	0.604	0.761	65.25%

Figure 10 Results of training various classifiers on our dataset

REFERENCES

[1] Gutierrez, M, Fast M, Ngu A, Gao, B, Real-Time Prediction of Blood Alcohol Content using Smartwatch Sensor Data, Volume 9545 of the series Lecture Notes in Computer Science pp 175-186

[2] Tokyoflash Japan, "Kisai Intoxicated LCD Watch," Tokyoflash Japan, 2014. [Online]. Available: http://www.tokyoflash.com/en/watches/kisai/intoxicated/. [Accessed 09 December 2014].

[3] Nieschalk, M., Ortmann, C., West, A., Schmäl, F., Stoll, W., & Fechner, G. (1999). Effects of alcohol on body-sway patterns in human subjects. Int'l Journal of Legal Medicine, 253-260.

[4] SCRAM Systems, "SCRAM Continuous Alcohol Monitoring," Alcohol Monitoring Systems, Inc., 2014. [Online]. Available: http://www.scramsystems.com/index/scram/continuous-alcoholmonitoring.[Accessed 09 December 2014].

[5] Intoxicheck® Phone App - Buzzed Driving App | Impaired Driving App | Impairment Assessment. (n.d.). Retrieved Dec 1, 2015, from http://fatalvision.com/intoxicheck.html

[6] Lifeloc Technologies - Impairment Goggles. (n.d.). Retrieved Dec 1, 2015, from https://lifeloc.com/c-60-impairment-goggles.aspx

[7] Brown, B, Good Question: How many people still wear watches, CBS Minnesota, July 22, 2015. http://minnesota.cbslocal.com/2015/07/22/good-question-how-many-people-still-wear-watches/

[8] 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.

[9] 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.

[10] 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 ACM Ubicomp 2013, Seattle, WA, 2013.

[11] Dietterich, Thomas. Ensemble Methods in Machine Learning, (1-5). Retrieved Dec 4, 2015, http://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf

[12] BACtrack, Long term DUI consequences, http://www.bactrack.com/blogs/expert-center/35042309-long-term-dui-consequences

[13] Hall, M. A. (1999). *Correlation-based feature selection for machine learning* (Doctoral dissertation, The University of Waikato).

[14] Mathworks. (n.d.). Bandpower. Retrieved April 20, 2016, from http://www.mathworks.com/help/signal/ref/bandpower.html

[15] Arnold, Z, LaRose, D, and Agu, E Smartphone Inference of Alcohol Consumption Levels from Gait, in Proc. IEEE ICHI 2015

[16] Qi, Muxi. "A Comprehensive Performance Comparison of Signal Processing Features in Detecting Alcohol Consumption from Gait Data." WPI Graduate Thesis, April 2016.

[17] Substance Abuse and Mental Health Services Administration (SAMHSA) National Survey on Drug Use and Health. Underage Alcohol Use among Full-Time College Students. Issue 31, 2006.

[18] Chambers, M., Liu, M., Moore, C.: Drunk driving by the numbers. United States Department of Education.

[19] Fact Sheets - Alcohol Use and Your Health. (2015, December 17). Retrieved January 8, 2016, http://www.cdc.gov/alcohol/factsheets/alcohol-use.htm

[20] Impaired Driving: Get the Facts. (2015, November 24). Retrieved January 8, 2016, http://www.cdc.gov/motorvehiclesafety/impaired_driving/impaired-drv_factsheet.html

[21] Tjondronegoro, D.; Drennan, J.; Kavanagh, D.J.; Zhao, E.J.; White, A.M.; Previte, J.; Connor, J.P.; Fry, M.-L., "Designing a Mobile Social Tool that Moderates Drinking," in IEEE Pervasive Computing, vol.14, no.3, pp.62-69, July-Sept. 2015 doi: 10.1109/MPRV.2015.62

[22] Mathworks. (n.d.). Total Harmonic Distortion. Retrieved April 20, 2016, from http://www.mathworks.com/help/signal/ref/thd.html

[23] Wang, K., Huang, M., Hsieh, Y., Lau, S., Yen, C., Kao, H,. Chen, Y. (2014). SoberDiary. In Adjunct Proceedings ACM Ubicomp 2014

[24] National Highway Traffic Safety Administration. (n.d.). The ABCs of BAC A Guide to Understanding Blood Alcohol Concentration and Alcohol Impairment. Retrieved April 20, 2016, from http://www.nhtsa.gov/links/sid/ABCsBACWeb/page2.htm

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

[26] Smartphone users worldwide 2014-2019 | Statistic. (n.d.). Retrieved April 20, 2016, http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[27] Kaewkannate, K, The Correlation Among Body Sway Parameters from the Gyroscope Data Sets, in Proc Biomedical Engineering Conference (BMEiCON) 2013.