IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 19, 2019, accepted September 3, 2019, date of publication September 12, 2019,
date of current version September 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940729

Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living

GARY M. WEISS, (Member, IEEE), KENICHI YONEDA, AND THAIER HAYAJNEH, (Member, IEEE)
Department of Computer and Information Science, Fordham University, Bronx, NY 10458, USA
Corresponding author: Gary M. Weiss (gaweiss@fordham.edu)

This work was supported in part by the National Science Foundation under Grant 1116124.

ABSTRACT Smartphones and smartwatches, which include powerful sensors, provide a readily available platform for implementing and deploying mobile motion-based behavioral biometrics. However, the few studies that utilize these commercial devices for motion-based biometrics are quite limited in terms of the sensors and physical activities that they evaluate. In many such studies, only the smartwatch accelerometer is utilized and only one physical activity, walking, is investigated. In this study we consider the accelerometer and gyroscope sensor on both the smartphone and smartwatch, and determine which combination of sensors performs best. Furthermore, eighteen diverse activities of daily living are evaluated for their biometric efficacy and, unlike most other studies, biometric identification is evaluated in addition to biometric authentication. The results presented in this article show that motion-based biometrics using smartphones and/or smartwatches yield good results, and that these results hold for the eighteen activities. This suggests that zero-effort continuous biometrics based on normal activities of daily living is feasible, and also demonstrates that certain easy-to-perform activities, such as clapping, may be a viable alternative (or supplement) to gait-based biometrics.

INDEX TERMS Authentication, biometrics, data mining, gait recognition, identification, sensors, smartphone, smartwatch, ubiquitous computing.

I. INTRODUCTION

The ability to identify or authenticate a person is critical to maintaining the security of digital and non-digital assets. This security is often provided by passwords or physical tokens (e.g., ID card), but these can easily be stolen or duplicated [1]. Biometric methods, which are tied to a person's unique physical or behavioral characteristic, generally do not share these disadvantages [2]. Common physical biometric systems are based on a person's fingerprints or iris. Such systems can sometimes be difficult to use and behavioral biometrics provides an important alternative. Motion-based biometrics is an especially attractive alternative because popular mobile devices, such as smartphones and smartwatches, contain motion sensors that can form the basis of a biometric system. Biometrics using these devices can be used as a primary mechanism for performing authentication and/or identification—or can function as part of a multi-factor system.

The associate editor coordinating the review of this manuscript and approving it for publication was Qiquan Qiao.

The study described in this article evaluates the use of the accelerometer and gyroscope sensors on commercially available smartphones and smartwatches for behavioral biometric authentication and identification, using eighteen activities of daily living. A different model is induced for each activity. This study is unique in that it evaluates a large number of activities of daily living for their biometric potential using a smartphone and smartwatch. Most other work focuses on a single activity (most often gait), while the few studies that assess more than one activity cover less than a half-dozen activities.

The evaluation of a large number of diverse activities for daily living is important for several reasons. One reason is to identify specific activities that form effective biometric signatures and hence can be used individually as the basis for a biometric system—just like gait is the basis of many existing biometric systems [3]–[9]. In this case, the subject would be asked to perform an action "on demand," which eliminates the issue of potentially not knowing which biometric model to apply. The interest in looking for new and novel activities to use for biometrics was motivated by a study that showed

that the acoustic signature from finger-snapping provides a useful biometric signature [10].

A second reason for evaluating a large number of activities is that if activities of daily living *generally* yield good biometric signatures, then it may be feasible to build a system that continuously performs biometric authentication (or identification) as the subject goes about his or her *normal daily routine*. Such a system would typically employ a two-stage approach, where an activity recognition system first recognizes the activity that is being performed, and then the biometric model for that activity is applied. Such an approach should be feasible since it was recently demonstrated, using the same data used in this study, that each of the eighteen activities can be recognized with an accuracy between 91.8% and 99.3% [11].

A biometric system based on the research described in this study can be implemented on a smartphone or smartwatch and thus could be used to secure those devices. Alternatively, these mobile devices can be used solely to collect the data, which is then sent to another entity (e.g., laptop, car, smart home, etc.), which subsequently applies the biometric model to validate the subject's identity. The wireless communication capabilities built into smartphones and smartwatches allow them to secure other devices.

This study makes several significant contributions, as it provides the most comprehensive study of smartphone and smartwatch-based biometrics using activities of daily living. Specifically, it answers the five following important research questions:

1. What is the relative value of a smartphone, smartwatch, and combination of the two devices for motion-based biometrics?
2. What combination of motion sensors perform best?
3. How well can motion-based biometrics perform for authentication and for identification?
4. How do diverse activities of daily living perform for biometrics? Do they perform well enough to permit continuous biometrics based on normal daily activities? Are any of them viable alternatives to gait?
5. How does the amount of training data impact biometric performance?

The focus of this article is on answering the five research questions just posed—not to advance the state of the art in biometric techniques. The techniques used in this article are relatively straightforward, but they nonetheless produce results that are competitive with other published studies. The current study makes very substantial extensions to a biometric study from 2010 [12], which employed only a smartphone and included only four simple activities, and a subsequent study from 2015 [7] that employed only a smartwatch and considered only the walking activity. This article also extends a preliminary study [13] that used the same devices and sensors but executed fewer experiments and provided much less analysis and discussion.

This article is organized as follows. Section II describes background and related work. Section III then describes the process for collecting the mobile sensor data and transforming it into a form suitable for data mining. The methodology used to build and evaluate the biometric models is described in Section IV, while the results from applying those models are described and analyzed in Section V. The biometric effectiveness of each of the eighteen activities is discussed in Section VI. The article concludes with Section VII, which summarizes the main conclusions, identifies areas for future work, and compares the current study to prior work to show that the current study is the most comprehensive on smartphone and smartwatch-based behavioral biometrics.

II. BACKGROUND AND RELATED WORK

This section provides context for the study described in this article by providing relevant background and related work. The discussion of related work is fairly extensive in order to demonstrate that the current study goes well beyond what was done in earlier studies—in terms of the number of activities evaluated, the types of biometrics tasks (authentication and identification), the number of subjects, and the use of the accelerometer and gyroscope on both a smartphone and smartwatch. Key characteristics of the related work are summarized in Table 9, which appears in the conclusion, and those summaries are used to show how the study described in this article goes beyond what was done in prior studies.

A. PHYSIOLOGICAL AND BEHAVIORAL BIOMETRICS

Biometric methods can be divided into physiological biometrics and behavioral biometrics [14], [15]. Physiological biometric methods rely on physiological traits such as fingerprints, iris, ear shape, DNA, vein pattern, or face [16]. The most popular physiological traits are fingerprints, which account for more than half of all commercial biometric systems [15]. Fingerprints are effective but are far from perfect—with EER values in the 2%-7% range [17]. The study described in this article concerns behavioral biometrics. Behavioral biometrics are based on behavioral characteristics that can be extracted from user actions, and include gait (i.e., walking), handwaving, keystrokes, signature, touchscreen contact, and voice [18]. Many of these user actions can be measured by smartphones and smartwatches and are covered in this section.

B. GAIT-BASED BEHAVIORAL BIOMETRICS

Gait is the most studied activity for behavioral biometrics. Gait biometrics [19] has been implemented using three different sensing modalities: vision-based, floor-based, and wearable sensor-based. Vision-based biometrics [16] requires video equipment but does not interfere with the subjects and is unobtrusive. However, the equipment requirement means this approach is only suitable for fixed locations (e.g., an airport). Floor-based systems also require special equipment, in the form of pressure sensors, and are even more geographically restrictive than the vision-based systems since the subjects

must walk on the instrumented area. One example of such a system was able to identify 15 subjects correctly 93% of the time [20].

Wearable sensor-based solutions are not limited to fixed locations but require each subject to be fitted with the sensors. This can be quite intrusive, especially when sensors are placed on multiple body locations. Gafurov and Snekkenes [21] attached custom-designed sensors with tri-axial accelerometers to the ankle, hip, pocket, and arm, and achieved an Equal Error Rate (EER) of 5% (ankle), 13% (hip), 7.3% (pocket), and 10% (arm). Several other studies, all conducted before 2009, similarly attached sensors to body position(s) to perform gait biometrics [22]–[24]. The intrusiveness issues can be largely avoided if the sensing is provided by commercial smartphones and smartwatches.

C. GAIT BIOMETRICS USING SMARTPHONES AND/OR SMARTWATCHES

This section covers research on gait biometrics using commercial smartphones and smartwatches. Four studies utilize only the smartphone accelerometer to perform biometric authentication using gait. Derawi *et al.* [3] achieved an EER of 20% using dynamic time warping to implement distance-based similarity, while Nickel *et al.* [4] achieved an EER of 10% using Support Vector Machines and Hidden Markov Models. Another study evaluated the impact of varying walking speed and achieved an EER of 3.6% with ''normal'' speed, an EER of 1.5% with ''fast'' speed, and an EER of 14.1% when ''normal'' speed was used for training and ''fast'' speed for evaluation [5]. Hoang *et al.* [6] developed a method to allow authentication models to be developed and deployed on different smartphone models, and achieved an authentication accuracy of 91% when training and testing on different phone models.

There is less research on smartwatch-based gait biometrics since smartwatches are a more recent development. One study used four different discriminant-based methods and achieved an EER of between 1.4% and 4.5% when using the accelerometer and an EER of between 6.3% and 9.6% when using the gyroscope [7]. Unlike the other studies mentioned in this section, this study also performed biometric identification. With 51 subjects and using only 10 seconds of data for evaluation, identification accuracies between 66.8% and 84.0% were achieved when using the accelerometer, and between 52.4% and 70.5% when using the gyroscope. Another study showed that by using a Microsoft Band 2 (not a smartwatch but worn in similar way), gait-based authentication can achieve an EER between 0.13% and 0.69% using the accelerometer sensor and an EER between 3.12% and 7.97% using the gyroscope sensor [8]. A final study placed a Shimmer 3 sensor unit on the wrist and in the pants pocket of 15 test subjects, thereby simulating a smartphone and smartwatch [9]. The system achieved an EER of 2.5% in the pocket and as low as 2.9% at the wrist.

D. NON-GAIT BEHAVIORAL BIOMETRICS USING SMARTPHONES AND/OR SMARTWATCHES

There are several common actions used for behavioral biometrics other than walking, and in this section we focus on those that utilize a smartphone or smartwatch. Most of the non-gait activities involve the subject touching the smartphone screen. One such study performed authentication using ''soft touchscreen'' (stouch) gestures such as: flick, spread, pinch, drag, and tap [25]. Using only the smartphone touch sensors the best performance yielded a false acceptance rate of 12% and a false rejection rate of 15%. Another study utilized the ''soft keystroke'' (skey) dynamics associated with the virtual keyboard to perform authentication, and used the detailed information provided by smartphones, such as the precise touch location, duration of touch, and pressure [26]. With 5 (15) keypresses the system achieved a false acceptance rate of 32.3% (14.0%) and false rejection rate of 4.6% (2.2%). Another keystroke-related study achieved a false acceptance rate and false rejection rate of 3-4%, when the only activity involved the subject inputting a 4-digit PIN code [27]. This study used the touch sensors, but also used the accelerometer, gyroscope and magnetometer to measure the motion and orientation of the smartphone as the PIN code was entered.

Two studies used the smartphone microphone to perform biometric authentication. One used samples from the smartphone phone calls and achieved an EER of 25% [28], while the other study used the acoustic properties of the finger snapping action to perform authentication and achieved an EER of about 6% [10].

One final non-gait-based study shares perhaps the most in common with the current study. It used smartwatch sensors to identify people based on their motion when handwriting specific prompts [29]. An accuracy of 90% was achieved when using the accelerometer, 85% when using the gyroscope, and 95% when using both sensors. The results were only slightly worse when each subject wrote out their signature instead of a predefined prompt. Authentication performance was not evaluated.

E. BIOMETRICS USING MULTIPLE ACTIVITIES OF DAILY LIVING

Most of the studies described thus far have limited overlap with the study described in this article in that they utilize only a single activity for biometrics. This section describes two studies that utilize a smartphone to perform biometrics using multiple activities; however, these studies are much more limited than the current study in that they analyze only a few activities and also do not employ a smartwatch.

The first study used the smartphone accelerometer to perform authentication and identification based on four activities: walking, jogging, climbing up stairs, and climbing down stairs [12]. Only limited experiments and analysis was completed for the authentication task. Authentication models were induced for only five subjects and the models were based on a combination of all four physical activities

without activity labels; the resulting average FAR was 14% and FRR was 5%. Identification experiments utilized the full set of 36 subjects with each of the four activities in isolation and then with all the activity data combined. The resulting accuracies were: 84% (walking), 83% (jogging), 66% (upstairs), 61% (downstairs), and 72% (combined activities).

The second study performed authentication using a smartphone accelerometer and gyroscope for six activities: walking, sitting, standing, running, climbing up stairs, and climbing down stairs [30]. A smartphone was placed on five body positions to assess the impact of position. The waist and thigh positions, which are the most reasonable positions for a smartphone, yielded authentication accuracies in the low-90s, while the arm and wrist positions, which simulate a smartwatch, yielded authentication accuracies in the mid-80s.

III. DATA COLLECTION AND TRANSFORMATION

This section describes all aspects of the data collection and transformation process. The data that was collected was used for this biometric study and for separate studies on activity recognition [11], [31]. This dual usage helped to defray the cost of the time-consuming data collection process.

A. THE EIGHTEEN PHYSICAL ACTIVITIES

This study includes eighteen routine physical activities, most of which are performed daily. For the purposes of this study, a physical activity is defined as a specific identifiable action with an associated starting and ending time. Some of the physical activities in this study (e.g., eating pasta) are not practical for a biometrics system that requests the subject to perform an activity on demand, but would be useful for a continuous biometrics system that operates as the subject performs their normal daily activities. Other activities, however, such as clapping and writing, are very easy to perform and could be done on an on-demand basis. Section VI, which discusses the overall biometric effectiveness of the activities, considers the practicality of the activities in the context of being performed on demand.

Table 1 lists the 18 activities included in this study. They are organized into three groups. The first grouping contains activities that are not primarily focused on hand movements. The other two groupings are primarily hand-based and are partitioned based on whether the activity involves eating. For each activity the appropriate equipment or food was provided. The eating activities, as well as the typing, writing, and sitting activities, were performed while seated; all other activities were performed while standing. The walking and jogging activities were performed outside, while the stairs activity was performed by repeatedly walking up and down several flights of stairs. The folding clothes activity was performed utilizing a table.

The activities listed in Table 1 were chosen for a variety of reasons. The following activities were chosen because they were included in prior studies: writing [29]; typing [26], [27]; walking [3]–[9]; and jogging, stairs, sitting, and standing [30]. The clapping activity was selected because it is easy to perform, can be done quickly, and the authors speculated it would yield good biometric performance. The remaining activities were selected partly because the data was already available from a prior activity recognition study [31]. Some of those activities represent basic sports activities (kicking a soccer ball, dribbling a basketball), while others (brushing teeth, folding clothes) were selected as representative activities of daily living. The five eating activities were selected to investigate the feasibility of automatic food tracking applications. While some of the activities included in this study were selected based on their suitability for activity recognition research, they nonetheless are relevant to the current biometrics study since they can help assess the viability of developing a continuous biometrics system based on a person's normal daily activities.

TABLE 1. Eighteen physical activities.

Non-Hand-Oriented Activities
• Walking
• Jogging
• Stairs (ascending & descending)
• Sitting
• Standing
• Kicking a Soccer Ball (two people)
Hand-Oriented Activities (General)
• Dribbling a Basketball
• Catch with a Tennis Ball (two people, underhand)
• Typing
• Writing
• Clapping
• Brushing Teeth
• Folding Clothes
Hand-Oriented Activities (Eating)
• Eating Pasta
• Eating Soup
• Eating a Sandwich
• Eating Chips
• Drinking from a Cup

B. THE DATA COLLECTION PROCESS

Smartphone and smartwatch sensor data were collected from 51 subjects, comprised mainly of undergraduate and graduate university students between the ages of 18 and 25. The data collection process was approved by the university's Institutional Review Board (IRB) and each subject provided written informed consent before participating in the study. Each subject performed the eighteen activities listed in Table 1 for 3 minutes each, with a smartphone in their right pants pocket and a smartwatch on their dominant hand. Participants used a Google Nexus 5/5X or Samsung Galaxy S5 smartphone running Android 6.0 (Marshmallow), and an LG G Watch running Android Wear 1.5. The data collection process for each subject took on average 70 minutes to collect the 54 (18×3) minutes of activity data.

The data collection was conducted primarily in a laboratory environment under the supervision of a researcher. However, for a few simple activities (walking, jogging, stairs), the subjects were given general instructions and allowed to perform the activities either outside the building or in the staircases within the building. The activities that require two people, kicking a soccer ball and playing catch with a tennis ball, were conducted with the active participation of the researcher. The use of the laboratory setting may introduce a bias into the data collection process, but was necessary to ensure a high-quality data set.

The time-series sensor data was collected by a customized Android application. The application logs data from any combination of the sensors available on an Android phone and/or Android Wear smartwatch. For this study the accelerometer and gyroscope data from both the smartphone and smartwatch were collected at a rate of 20 Hz (prior work [32], [33] shows that higher rates are not necessarily beneficial for motion-based predictive models). However, due to the nature of the Android OS, the sampling rate is only taken as a suggestion, so actual sampling rates sometimes differed. At the end of a data collection session, the raw time-series sensor data was transferred from the smartphone to a lab machine via a USB connection.

C. THE RAW SENSOR DATA

The raw time-series sensor data is stored in separate files. Each file contains the data from one sensor (accelerometer or gyroscope) on one device (smartphone or smartwatch) for one subject. Thus there are four files associated with each subject, although the sensor data from these four files will have been collected during the same time period and can be linked via the timestamp information. Each sensor measurement is recorded on a separate line in the data file with the following format:

$$< \text{subject-id, activity, timestamp, x, y, z} >$$

The subject-id identifies the test subject, activity is a code that identifies the physical activity performed, timestamp is the Unix time at which the sensor value was recorded, and the x, y, and z values represent the sensor values for the x, y, and z spatial axes. The format of the recorded sensor data is identical whether the data is from the accelerometer or gyroscope on either the smartphone or smartwatch. The accelerometer measures linear acceleration in meters/s^2 and the gyroscope measures angular velocity in radians/s. The raw sensor data used in this study is publicly available as the WISDM human activity recognition and biometrics data set from the UCI Repository [34]. Fig. 1 provides a graphical representation of the smartphone accelerometer data for the walking and jogging activities. The y-axis corresponds to the vertical direction and, as one would expect, has the largest magnitude. Also as expected, the jogging activity exhibits a higher frequency than the walking activity.

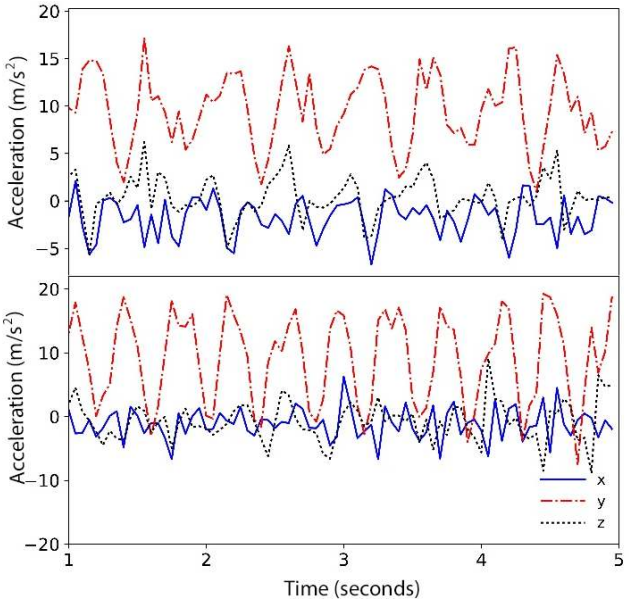

FIGURE 1. Graphical plot of the smartphone's triaxial accelerometer data for the walking activity (top) and the jogging activity (bottom).

D. THE TRANSFORMED DATA

Many classification algorithms do not handle time-series data directly, but instead require the data to be in the form of labeled examples, where each example is described by a fixed number of features. In order to generate data in this form, the data is partitioned into 10-second non-overlapping segments, and the time series data in each segment is described using a set of forty-three high level features. A 10-second window is used because it is sufficiently long to capture key elements of a person's movements, including several repetitions of basic movements like walking and stair climbing, and facilitates fast biometric identification. Prior activity recognition experiments also demonstrated that a 10-second window size outperforms other window sizes [35], and using the same window sizes for both activity recognition and biometrics facilitates the two-stage biometric approach discussed in Section I. Finally, longer periods of sensor measurements can still be utilized by basing the biometric decision on several examples—a strategy that is evaluated in this study.

The forty-three features are described below. The value in brackets denotes the total number of features of the given type (most features apply to each of the three spatial axes). Given the sampling rate of 20Hz and a window size of 10 seconds, there are 200 raw sensor readings for each example. In prior biometric work we did experiment with using more sophisticated features, such as by using Fourier analysis, but such features did not improve the results. None of the feature values were normalized.

- Average [3]: average sensor value (each axis)
- Standard deviation [3]: standard dev. (per axis)
- Average absolute difference [3]: average absolute difference between each of the 200 sensor readings and the mean of the 200 readings (per axis)

- Average resultant acceleration [1]: the average of the square root of the sum of the square of the x, y, z axis values.
- Binned distribution [30]: the range (max–min value) is determined for the window, 10 equal-sized bins are formed, and the fraction of the 200 values within each bin is recorded (per axis)
- Time between peaks [3]: Time between peaks in the sinusoidal waves formed by the data as deter-mined by a simple heuristic algorithm (per axis)

Each transformed example is tagged with the test subject's identifier and the activity that was performed, so that the appropriate activity is used for each experiment. The publicly available data set also includes the scripts for transforming the time-series data [34].

The transformation process just described generates an example using the data from only a single sensor. That process can be used to evaluate the four single-sensor configurations: *phone-accel*, *phone-gyro*, *watch-accel*, and *watch-gyro*. Given that one goal of this study is to identify the combination of sensors that yields the best results, the following five sensor combinations are also evaluated:

Phone:	*phone-accel* + *phone-gyro*
Watch:	*watch-accel* + *watch-gyro*
Accels:	*phone-accel* + *watch-accel*
Gyros:	*phone-gyro* + *watch-gyro*
All:	*phone-accel* + *phone-gyro* +*watch-accel* + *watch gyro*

The transformed examples for the various sensor combinations are generated by concatenating the 43 features associated with each sensor. Hence the *Phone*, *Watch*, *Accels*, and *Gyros* sensor combinations each contain 86 features, while the *All* sensor combination has 172 features.

IV. EXPERIMENT METHODOLOGY

This section describes the methodology used to execute all the experiments. Section IV-A describes the classification algorithms used to build the biometric models, while Section IV-B describes how data from a test subject are used to make an authentication or identification decision. The methodology used to construct and evaluate the authentication and identification experiments is described in Sections IV-C and IV-D, respectively.

A. CLASSIFICATION ALGORITHMS

Three classification algorithms are used to generate the authentication and identification models that are evaluated in this study: k Neighbors, Decision Tree, and Random Forest. We employ the implementations of these algorithms from Python's scikit-learn module, an open source library for data mining and analysis [36]. Unless otherwise specified, the default parameters are used. For k-Neighbors the number of neighbors is set to 5, using uniform weights and the Minkowski distance metric. For the Random Forest classifier, the maximum number of features considered is the square root of the number of features in the data, and the number of decision trees in the forest is set to 10.

B. DECISION MAKING USING EVALUATION DATA

The authentication and identification tasks both require a sample of data from a subject to make an authentication or identification decision. The simplest strategy is to use a single ''test'' example to make the decision. Given the method for transforming the raw time-series data into examples, these decisions are based on 10 seconds of sensor data. However, for some biometric applications it may be practical to utilize more than 10 seconds of data to make a decision. In this study we also evaluate authentication and identification performance using 50 seconds of data, where the decision is based on a simple majority voting scheme applied to five 10-second examples (using more than five samples does not significantly impact performance). Results are presented for both the non-voting and voting strategies. Based on the application demands one can decide whether to use voting or not—or perhaps use voting with only 30 seconds of data. The 10 second window size should be reasonable for most biometric applications, and if a faster decision time is required then behavioral biometrics is not appropriate.

C. AUTHENTICATION EXPERIMENT METHODOLOGY

The authentication task involves distinguishing an authorized subject from an imposter. Hence authentication is a classification problem involving two classes. For authentication, each authorized subject must have their own model, which means that in this study fifty-one authentication models are evaluated. Each authentication model is based on a single activity, so this is repeated 18 times (once per activity), and each experiment is replicated for the 3 classification algorithms and the 9 sensor combinations, so that a total of 24,786 ($51 \times 9 \times 18 \times 3$) experiments are executed.

Each model is trained using data from the subject to be authenticated and data from ''other'' subjects that are combined into a single class. In real world situations, data from actual imposters will not be available, so it is critically important to ensure that the ''imposters'' in the training set and test set do not overlap. Also, training data for the authentication models must be partitioned carefully, since a training set with a high degree of class imbalance will be biased against authenticating a valid user. We partition the data as follows. The data for the subject to be authenticated is divided equally between the training and test sets, such that each set has 90 seconds of the subject's data (i.e., for the selected activity). Then eighteen other subjects are randomly selected and 30 seconds of data for that activity are randomly chosen for each subject. Data from nine of these subjects is placed in the training set, while data from the other nine are placed into the test set; this yields 270 (9×30) seconds of ''other'' data for the training and test sets. The resulting training class ratio is 90:270, or 1:3.

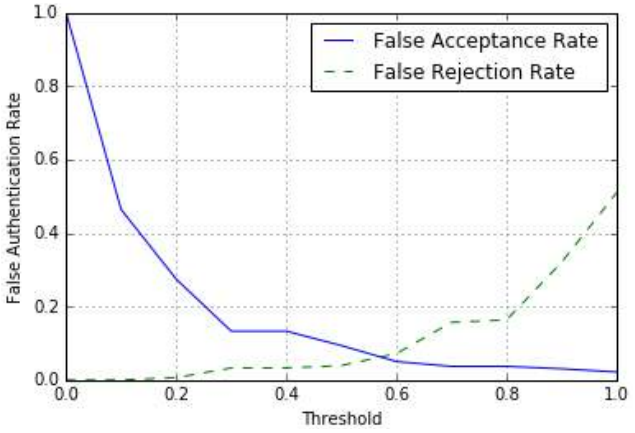

FIGURE 2. Graphical plot of FAR and FRR with an EER value of 0.068.

The training set utilizes a class ratio of 1:3, where the subject to be authenticated is the minority class. This was done for several reasons. First, this ratio is used in other biometric studies [7], [12]. Second, empirical results showed that the 1:3 class ratio performed best. Class ratios with lower levels of class imbalance, including 1:1 and 1:2, led to slightly poorer performance—probably because they included less imposter data. Meanwhile, class ratios with higher levels of class imbalance also led to poorer results, because the high level of imbalance led to too many predictions of the ''imposter'' class. One-class learning (e.g., one-class SVM), which would not require the use of any imposter data and hence avoid the class imbalance problem altogether, is worthy of future consideration.

In this study, authentication performance is evaluated using Equal Error Rate (EER), a common metric for com-paring authentication models [23]. This metric is calculated as the point where the False Acceptance Rate (FAR), the rate at which the model incorrectly accepts an imposter as a legitimate user, equals the False Rejection Rate (FRR), the rate at which the model incorrectly rejects a legitimate user. FAR and FRR can be altered by varying the probability threshold used for assigning a classification. Fig. 2 shows the actual FAR and FRR curves for the walking activity using *All* sensors and the Random Forest algorithm. In this case the EER value is approximately 0.068, which is the y-coordinate at which point the two curves intersect.

D. IDENTIFICATION EXPERIMENT METHODOLOGY

The methodology for executing the identification experiments is much simpler than for the authentication experiments, since we do not need to map the data from multiple subjects into a single class. Instead, each subject represents a different class—for the identification data set there are fifty-one classes. In this case the training set must have data from all the subjects and hence the subjects in the training and test set should overlap. The training and test data is partitioned using stratified 10-fold cross validation, so that the training and test sets both have the same percentage of data from each subject. Identification models are built per activity. As with the authentication experiments, each of these experiments are also executed for each of the three classification algorithms and for each of the nine sensor combinations.

TABLE 2. Summary authentication EER over 18 activities.

Alg.	Watch		Phone		Watch	Phone	Accels	Gyros	All
	accel	gyro	accel	gyro					
No Voting: single 10-second example									
RF	19.5	25.8	12.0	20.2	19.0	12.0	11.3	19.0	11.5
DT	61.8	35.5	90.1	51.6	63.7	90.5	91.8	56.9	91.4
kNN	36.2	14.6	76.3	24.9	34.2	75.0	77.8	34.9	75.3
Voting: 5 10-second examples									
RF	15.6	22.4	9.7	17.6	15.3	9.6	9.3	16.0	9.3
DT	20.0	30.1	13.7	23.5	19.9	12.9	13.6	21.4	13.0
kNN	22.4	29.0	11.9	22.4	22.5	11.4	11.8	19.3	11.5

V. RESULTS

This section presents and analyzes the results of all of the experiments conducted in the study. The results are presented first for the authentication experiments and then for the identification experiments.

A. AUTHENTICATION RESULTS

This section provides the results for all the authentication experiments. Table 2 aggregates the results over all eighteen activities in order to identify the best overall classification algorithm for authentication. The best EER value, for each sensor combination and labeling method (i.e., single example versus voting), is underlined. The results demonstrate that the Random Forest algorithm performs best for every sensor combination except for watch-gyro when there is no voting. Given the superiority of the Random Forest algorithm, the remainder of this section focuses on the experimental results for that algorithm. Some prior work on activity recognition similarly showed that Random Forest performs best [31].

The detailed authentication results for the Random Forest algorithm, where decisions are based on a single 10-second example (i.e., without voting), are provided in Table 3. Results are provided for each of the eighteen activities and each of the nine sensor combinations. The last row in the table provides the average performance over all eighteen activities. The relative value of each of the nine sensor configurations can be determined by comparing the values in the different columns.

Given that lower EER values are best, Table 3 shows that *Accels* and *All* are the best overall configurations, with *Accels* having a slight edge (11.3% vs. 11.5% average EER and better performance for 11 of the 18 activities). The next best sensor configurations are *Phone* and *Phone-accel*, which both have an average EER of 12.0%. The other five sensor configurations have a much higher average EER—at least 19%.

TABLE 3. Authentication EER using a one 10S example (RF).

Activity	Phone		Watch		Phone	Watch	Accels	Gyros	All
	accel	gyro	accel	gyro					
Walking	11.2	11.3	17.5	18.8	9.3	16.1	12.6	10.2	7.9
Jogging	11.5	13.2	18.1	19.3	10.3	15.1	11.3	13.8	9.8
Stairs	12.3	16.4	24.3	26.1	11.8	21.6	13.9	16.5	13.5
Sitting	13.6	26.3	21.8	33.4	12.8	22.3	10.7	27.2	13.0
Standing	14.7	26.0	22.6	33.3	15.6	23.0	11.9	27.9	15.4
Kicking	12.5	18.5	21.8	26.7	11.5	21.1	13.8	16.7	14.0
Dribbling	12.2	19.9	18.9	21.0	12.7	17.9	11.2	15.7	12.0
Catch	10.8	20.3	20.6	20.8	13.4	16.7	12.1	17.2	12.2
Typing	11.5	19.4	16.8	26.2	11.3	18.0	10.4	19.0	8.7
Writing	13.3	19.4	15.3	27.1	12.3	15.6	11.2	18.5	10.8
Clapping	11.3	20.5	15.8	20.8	11.7	19.2	9.7	14.6	10.6
Teeth	11.8	19.7	18.6	22.7	12.1	17.2	11.4	19.9	12.2
Folding	11.4	16.6	19.6	24.7	12.3	17.1	8.3	17.0	10.9
Pasta	12.4	23.0	18.4	28.8	14.4	20.4	12.3	22.6	10.9
Soup	9.6	22.4	17.6	24.6	10.1	17.5	8.6	21.7	9.8
Sandwich	11.4	22.6	24.1	30.2	10.4	22.1	10.1	23.6	12.3
Chips	12.3	23.3	19.2	29.5	11.7	20.3	11.3	20.4	10.2
Drinking	12.0	24.2	20.0	30.1	12.9	20.1	11.8	19.7	12.4
Ave	12.0	20.2	19.5	25.8	12.0	19.0	**11.3**	19.0	**11.5**

TABLE 4. Authentication EER using voting with 5 examples (RF).

Activity	Phone		Watch		Phone	Watch	Accels	Gyros	All
	accel	gyro	accel	gyro					
Walking	9.4	9.8	13.2	17.2	8.8	13.9	11.3	10.0	6.8
Jogging	7.8	10.8	16.2	15.2	9.7	12.7	9.0	11.2	8.3
Stairs	13.4	12.5	19.3	23.9	9.3	18.9	8.4	14.1	6.9
Sitting	10.4	23.7	14.5	32.1	8.8	17.0	10.0	21.1	10.2
Standing	12.1	22.1	16.7	31.6	10.9	15.2	10.0	21.5	7.7
Kicking	10.6	19.4	21.0	24.1	11.0	16.6	10.1	18.8	11.0
Dribbling	10.3	21.0	16.4	16.1	9.7	14.5	10.0	11.8	11.5
Catch	9.7	19.3	16.3	15.5	10.0	14.9	9.3	13.9	10.0
Typing	8.3	15.4	13.0	20.7	8.9	14.0	8.6	13.3	8.8
Writing	8.7	15.7	10.7	21.3	9.2	11.6	9.0	16.0	10.1
Clapping	9.4	13.4	12.9	17.2	10.1	13.2	8.1	14.8	8.5
Teeth	10.1	14.0	13.3	20.0	10.2	14.4	10.8	14.9	8.2
Folding	7.9	18.6	17.0	23.4	10.0	17.3	8.1	16.2	7.1
Pasta	8.0	23.7	14.3	26.6	8.9	18.5	9.0	19.6	5.4
Soup	7.3	19.2	17.0	22.3	6.1	13.3	7.8	17.5	8.0
Sandwich	9.9	17.9	17.5	25.7	11.4	17.7	8.2	16.2	9.3
Chips	9.9	21.5	14.7	25.9	10.3	18.1	8.5	17.2	8.0
Drinking	11.3	19.2	16.6	25.1	10.2	13.9	10.9	19.9	8.1
Ave	9.7	17.6	15.6	22.4	9.6	15.3	**9.3**	16.0	**9.3**

From these results we conclude that if both a phone and watch are available, the best choice is to use the accelerometers from both devices. However, if only a phone is available, one can still do nearly as well by using either both phone sensors or only the phone accelerometer. If only a watch is available, then both watch sensors should be used, but performance will suffer significantly. The results also show that the gyroscope on either device performs more poorly than the accelerometer.

Performance can be improved by using a longer sample of data for making the authentication decision. Table 4 provides the results for experiments identical to those used to populate Table 3, except that the authentication decision is based on majority voting using five examples (i.e., 50 seconds of data). The results largely parallel those of Table 3, with the main difference being that the EER values are lower in Table 4. As before, over the 18 activities the *All* and *Accels* sensor configurations yield the best results. Voting improves the average results for *Accels* by 18% (9.3% vs. 11.3%), while for *All* it improves performance by 20% (9.3% vs. 11.5%). The walking activity, which is the most commonly used activity for motion-based biometrics, improves by 10% for *Accels* and 14% for *All*.

The voting results demonstrate that having a larger sample of data from a subject for evaluation (i.e., test data) allows for better authentication performance. Similarly, more training data should also have a positive impact. Fig. 3 provides a learning curve for authentication performance, averaged over all eighteen activities, where the x-axis measures the amount of training data per activity and the y-axis measures performance in terms of EER. The results show that performance

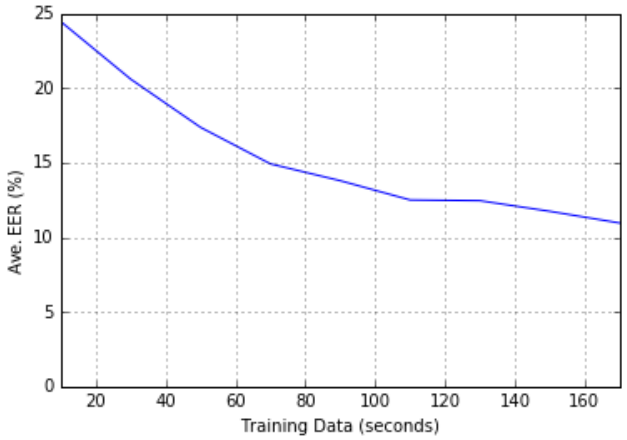

FIGURE 3. Learning curve that plots EER averaged over all eighteen activities versus the amount of training data per activity. The authentication models are generated using the Random Forest algorithm and the phone and watch accelerometer data (Accels) without voting.

is quite sensitive to the amount of training data. Notably, even with 170 seconds of training data the performance has not reached a plateau—suggesting that substantial improvements might be possible if additional training data could be acquired.

B. IDENTIFICATION RESULTS

The identification task is a multi-class classification problem, which for this study includes fifty-one classes. Table 5 presents the identification results that are based on one 10-second test example. The results are only for the Random Forest algorithm since the other algorithms produced inferior results. Specifically, the average accuracy over

TABLE 5. Identification accuracy using one 10-second example (RF).

Activity	Phone		Watch		Phone	Watch	Accels	Gyros	All
	accel	gyro	accel	gyro					
Walking	96.1	94.7	75.1	67.0	96.8	78.9	96.5	95.3	97.4
Jogging	94.7	92.5	75.0	74.3	96.0	82.1	95.7	95.2	98.0
Stairs	90.8	81.2	52.4	39.2	92.7	58.7	92.6	80.9	95.1
Sitting	90.1	56.3	70.4	30.1	91.5	69.3	93.1	55.9	92.0
Standing	85.8	47.1	64.1	27.0	86.8	61.2	90.5	46.6	89.9
Kicking	87.4	67.2	54.3	38.3	88.6	59.8	92.1	72.7	92.1
Dribbling	88.3	66.0	72.3	74.8	89.5	80.3	93.9	82.1	94.4
Catch	90.0	67.2	69.1	71.3	90.3	75.4	94.1	82.0	93.7
Typing	94.8	71.7	81.2	51.2	94.6	84.2	95.6	76.5	95.7
Writing	92.8	69.1	79.6	47.6	93.1	79.1	94.2	73.0	93.9
Clapping	94.8	72.8	83.4	73.9	93.8	85.3	96.6	86.1	96.7
Teeth	92.2	69.5	70.0	56.3	93.7	76.1	95.2	74.5	95.4
Folding	90.7	65.8	60.0	38.8	92.0	63.0	93.6	72.7	93.8
Pasta	94.1	56.9	67.2	38.1	94.0	71.6	96.6	61.1	96.3
Soup	94.3	56.5	74.1	50.4	95.8	76.6	96.3	66.9	96.6
Sandwich	92.9	62.8	61.9	37.6	92.6	62.1	95.9	68.5	95.2
Chips	93.3	56.8	62.6	38.7	93.2	62.4	96.0	66.3	94.9
Drinking	93.9	57.4	63.9	41.3	93.8	65.3	95.4	60.6	94.7
Ave	92.1	67.3	68.7	49.8	92.7	71.7	**94.7**	73.2	**94.8**

TABLE 6. Identification accuracy using voting (RF).

Activity	Phone		Watch		Phone	Watch	Accels	Gyros	All
	accel	gyro	accel	gyro					
Walking	100	100	94.1	80.4	100	90.2	100	100	100
Jogging	100	100	90.0	88.0	100	98.0	100	100	100
Stairs	98.0	90.0	70.0	43.8	96.0	75.0	100	91.7	100
Sitting	100	62.7	88.2	33.3	98.0	86.3	100	64.7	100
Standing	98.0	39.2	82.4	20.0	94.1	84.0	100	50.0	100
Kicking	96.1	68.6	76.0	32.0	100	82.0	100	80.0	98.0
Dribbling	96.1	68.6	98.0	90.2	98.0	86.3	96.1	96.1	100
Catch	100	86.3	96.1	90.2	100	98.0	100	100	98.0
Typing	100	89.8	94.0	50.0	100	100	100	95.9	100
Writing	96.1	80.0	94.1	58.8	100	98.0	100	90.0	100
Clapping	100	86.3	96.1	90.2	100	98.0	100	100	98.0
Teeth	98.0	82.4	94.1	62.7	100	96.1	100	94.1	100
Folding	100	76.5	64.7	39.2	96.1	86.3	100	78.4	100
Pasta	100	56.0	84.0	48.0	100	84.0	100	71.4	98.0
Soup	100	66.7	88.2	62.0	100	88.0	100	80.0	100
Sandwich	98.0	68.0	84.0	38.0	100	82.0	100	73.5	98.0
Chips	100	76.0	82.4	41.2	98.0	82.4	98.0	80.0	100
Drinking	100	58.8	86.3	41.2	100	80.4	100	60.8	100
Ave	98.8	74.4	85.8	55.8	98.9	88.3	**99.7**	83.2	**99.6**

the eighteen activities, without voting and using the *Accels* sensor combination, is 94.7% for Random Forest, 91.8% for Decision Tree, and 77.8% for kNN.

As with the authentication results, Table 5 shows that averaged over the 18 activities, the *Accels* and *All* sensor combinations perform best, with identification accuracies of 94.7% and 94.8%, respectively. The phone device alone can provide fairly good performance (92.7%), but the watch alone provides much weaker performance (71.7%). As with authentication, the gyroscope sensors consistently perform worse than the accelerometer sensor. The identification results are quite promising, given that with fifty-one classes the strategy of random guessing would yield an accuracy below 2%. Note that while the biometric performance does vary by activity, the differences are quite modest: for the *Accels* sensor configuration the performance per activity ranges from 90.5% (standing) to 96.6% (clapping, eating pasta) and for the *All* sensor configuration the values range from 89.9% (standing) to 98.0% (jogging). The relative performance of each activity for biometric identification is discussed in additional detail in Section VI.

Table 6 provides the identification results when using a voting strategy with five 10-second examples. As was the case with the authentication results, the voting strategy yields a substantial improvement for identification performance. In fact, for both the *Accels* and *All* configurations, the majority of the activities yield perfect (100%) performance. The *Accels* and *All* sensor configurations also again perform best over the eighteen activities. With the voting strategy, one can achieve substantially better results when just using the phone—a 98.9% accuracy versus only 92.7% without voting. Performance with the watch is also much more competitive (88.3% with voting versus 71.7% without voting), but still does not perform nearly as good as with the phone sensors.

Identification accuracy will be impacted by the amount of training data available. The learning curve for the identification task is provided in Fig. 4, which plots accuracy versus the number of minutes of training data per activity. The learning curve results are based on averages over all eighteen activities and all fifty-one subjects, using Random Forest and the *Accels* sensor configuration. The figure shows that diminishing returns begin to set in once there is about one minute of training data per activity—but nonetheless the performance continues to gradually improve past this point and does not reach a plateau even with 140 seconds of training data. These results are encouraging since it indicates that good performance is possible with a modest amount of training data.

VI. BIOMETRIC EFFECTIVENESS OF ACTIVITIES

One contribution of this research is that it evaluates a large number of physical activities, whereas almost all motion-based biometrics research, as described in Section II and demonstrated later in Table 9, focuses on a single activity. In this section we evaluate the relative value of each activity for use in biometrics. In making this determination the *practicality* of the activity is considered along with its performance as a biometric signature. In the context of assessing practicality we assume that the activity will be performed on-demand. In this context an activity is practical if it is easy to perform

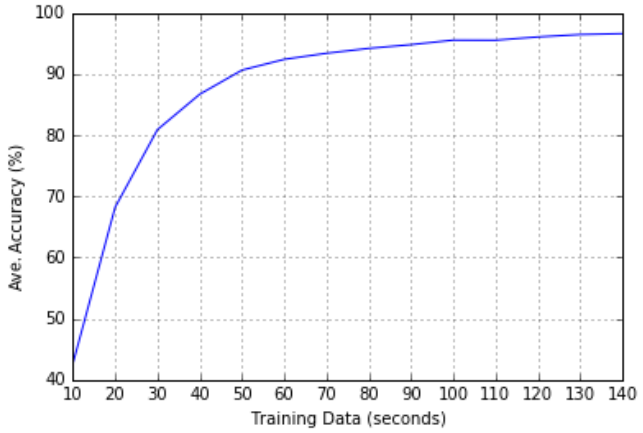

FIGURE 4. Learning curve that plots identification accuracy averaged over all eighteen activities versus the amount of training data per activity. Identification models are generated using the Random Forest algorithm and the phone and watch accelerometer (Accels) data without voting.

TABLE 7. Identification accuracy without voting (RF).

AUTHENTICATION PERFORMANCE (EER)
Non-hand-based activities:
Walking (10.25), Jogging (10.55), Sitting (11.85), Standing (13.65), Stairs (13.7), Kicking (13.9)
Hand-oriented Activities (General)
Typing (9.55), Folding (9.6), Clapping (10.15), Writing (11.0), Dribbling (11.6), Teeth (11.8), Catch (12.15)
Hand-Oriented Activities (Eating)
Soup (9.2), Chips (10.75), Sandwich (11.2), Pasta (11.6), Drinking (12.1)
IDENTIFICATION PERFORMANCE (ACCURACY)
Non-hand-based activities:
Walking (96.95), Jogging (96.85), Stairs (93.85), Sitting (92.55), Kicking (92.1), Standing (90.2)
Hand-oriented Activities (General)
Clapping (96.65), Typing (95.65), Teeth (95.3), Dribbling (94.15), Writing (94.05), Catch (93.9), Folding (93.7)
Hand-Oriented Activities (Eating)
Pasta & Soup (96.45), Sandwich (95.55), Chips (95.45), Drinking (95.05)

by a large segment of the population and does not require unusual equipment.

Biometric effectiveness is determined separately for authentication and identification, since each yields different biometric performance results. Performance is based on the results without voting, since the improvements associated with voting obscure many of the underlying performance differences. In all cases biometric performance is measured using Random Forest and by averaging the results for the *Accels* and *All* sensor configurations, since these are consistently the two best-performing sensor configurations. Since the types of activities in each of the three main activity groupings are quite different, we rank the activities within each group separately. Table 7 provides the rankings, from best to worst, for the activities in each activity group, with the actual performance value denoted in parentheses.

From the non-hand-based activity group, walking performs best and jogging performs second best for both authentication and identification. This is notable since walking is currently the most commonly used activity for motion-based biometrics. While jogging performs nearly as well, it is not nearly as practical as walking due to the amount of physical exertion required. Interestingly, sitting performs reasonably well for both tasks even though it does not entail much motion—the minor shifts in position that occur must be distinctive.

The next activity grouping is the general hand-oriented activities. From this grouping, teeth, dribbling, catch, and folding are the least practical on-demand activities because they require equipment or materials that are not routinely available. Typing, clapping, and writing are more practical, especially since typing and writing are common in the business environments that most often require authentication. For example, it certainly would be practical to type on a computer keyboard in order to authenticate one's identity to gain access to the associated computer. Clapping, because it requires no equipment and can be performed by almost everyone, is intriguing as a potential new activity for biometrics. Clapping is similar in spirit to the finger snapping activity studied elsewhere [10], although that work relied on acoustic properties to form the biometric signature.

The eating activities are not useful for "on-demand" biometrics, and hence are not discussed in detail. But it is important to note that these activities have biometric performance that is consistent with the other activities, and hence they should help enable a system that performs continuous biometrics based on activities of daily living. The activities that are most commonly performed further demonstrate that continuous biometrics based on one's normal daily routine should be feasible. For example, walking is an activity that is generally performed every day and yields good biometric performance, while typing and sitting are very common activities for today's workforce.

In summary, for on-demand biometrics, the best activities, based on performance and practicality, are: walking, typing, and clapping—with writing lagging a bit due to poorer performance. The good biometric performance of the sitting and typing activities suggest that it will be possible to perform continuous biometric authentication while someone is seated at a computer. The relatively good biometric performance of all eighteen activities further suggests that it may be possible to perform continuous biometrics while a person performs their normal daily routine (although the eighteen activities certainly do not cover all possible activities).

As mentioned earlier, a system that performs continuous biometrics as one performs their normal daily activities will typically require two stages. The first stage identifies the activity from the sensor data using an activity recognition model, while the second stage applies the biometric model

TABLE 8. Activity recognition performance (accuracy %).

Activity	Phone		Watch		Phone	Watch	Accels	Gyros	All
	accel	gyro	accel	gyro					
Walking	95.8	92.3	87.8	85.6	96.6	89.1	96.8	94.4	97.0
Jogging	95.5	94.3	96.9	93.6	98.6	97.3	99.3	98.1	99.3
Stairs	89.9	84.1	85.5	70.4	92.7	84.0	93.7	88.3	93.8
Sitting	86.7	59.7	87.3	62.8	87.0	84.0	91.9	70.8	91.8
Standing	90.0	68.2	90.7	59.0	90.2	89.7	94.8	75.1	94.7
Kicking	87.8	80.4	82.9	72.7	90.6	84.4	93.3	86.1	92.7
Dribbling	84.9	75.7	91.2	90.6	88.2	96.1	95.2	94.7	95.6
Catch	83.2	73.9	90.5	88.7	85.8	94.4	94.3	94.2	94.7
Typing	90.3	69.2	94.1	83.3	92.3	92.9	95.8	83.3	95.4
Writing	89.7	67.6	89.9	77.6	90.8	91.2	92.4	81.2	92.9
Clapping	88.7	72.6	95.0	92.7	91.0	96.6	96.8	94.1	97.6
Teeth	90.0	69.6	91.9	81.6	90.4	94.8	96.2	86.1	95.2
Folding	88.4	82.9	89.8	85.3	92.1	95.2	95.9	95.6	96.1
Pasta	84.4	48.0	83.3	68.3	85.8	84.1	92.2	70.4	92.6
Soup	86.3	52.6	86.6	69.1	85.5	87.3	93.0	74.3	93.9
Sandwich	86.7	48.1	72.7	50.5	84.7	70.9	91.1	59.1	90.4
Chips	82.9	50.1	78.8	60.6	83.0	80.0	92.0	69.9	92.4
Drinking	85.5	50.0	80.9	65.2	85.2	80.8	92.7	69.9	92.1
Ave	87.8	69.6	87.8	75.9	89.7	88.8	94.4	83.1	94.4

for the corresponding activity. A recent study [11] utilizes the same data set and activities as the study described in this article. The activity recognition results from that study for the Random Forest algorithm are summarized in Table 8. The results are based on personal activity recognition models built using data from the intended subject (personal models vastly outperform impersonal/universal models built from a panel of subjects [31]). The collection of labeled training data can be done by the subjects themselves, as was demonstrated by the self-training mode implemented in the Actitracker activity recognition system [37]. Table 8 shows that if the subject has a smartphone and smartwatch then an average accuracy of 94.4% can be achieved using either the *All* or *Accels* sensor configuration—and that each activity can be recognized with at least 91% accuracy. With only one of the two devices it is still possible to achieve an average accuracy of about 89%.

VII. CONCLUSION

This study demonstrates that motion-based biometrics using activities of daily living is feasible using a commercially available smartwatch and/or smartphone. It also answers the five research questions posed in Section I. To establish the research contributions of this study, Section VII-A analyzes the differences between the current and prior studies, and shows that the prior studies are not as comprehensive and do not answer the five research questions. The main conclusions of the study are then summarized in Section VII-B and areas for future work are discussed in Section VII-C.

TABLE 9. Summary of the most relevant biometrics work.

Study	Num	Device(s)			Sensor			Activities		Task	
	Subj.	phone	watch	other	accel	gyro	other	gait	other	auth	id
[3]	51	✓	-	-	✓	-	-	✓	-	✓	-
[4]	36	✓	-	-	✓	-	-	✓	-	✓	-
[5]	36	✓	-	-	✓	-	-	✓	-	✓	-
[6]	14	✓	-	-	✓	-	-	✓	-	✓	-
[7]	59	-	✓	-	✓	✓	-	✓	-	✓	✓
[8]	60	-	✓	-	✓	✓	-	✓	-	✓	-
[9]	15	✓	✓	-	✓	-	-	✓	-	✓	-
[25]	40	✓	-	✓	-	-	touch	-	stouch	✓	-
[26]	13	✓	-	-	-	-	touch	-	skey	✓	-
[27]	12	✓	-	-	✓	✓	magn.	-	skey	✓	-
[28]	14	✓	-	-	-	-	audio	-	talk	✓	-
[10]	76	✓	-	-	-	-	audio	-	snap	✓	-
[29]	24	-	✓	-	✓	✓	-	-	write	-	✓
[12]	36	✓	-	-	✓	-	-	✓	3	✓	✓
[30]	10	✓	-	-	✓	✓	-	✓	5	✓	-
This	**51**	✓	✓		✓	✓	-	✓	**17**	✓	✓

A. DIFFERENCES WITH PRIOR RESEARCH STUDIES

The key characteristics of the most relevant related work are summarized in Table 9. The table is organized into four main groupings that are separated by solid lines. Dashed lines identify subgroupings. The last line in the table summarizes the characteristics of the current study.

The first grouping, which consists of the seven studies described in Section II-C, covers gait biometrics research that utilizes a smartphone or smartwatch. The first four entries utilize only a smartphone, while the last three utilize a smartwatch. Only one of these seven studies utilizes both a smartphone and a smartwatch—and in that case the two devices are only simulated. Also, only two of the studies include the gyroscope, and even the studies that utilize more than one device and sensor only considered each in isolation. Furthermore, only one of these studies [7] considered the identification task and none of them analyzed the impact of the amount of training data on biometric performance. The seven studies also only considered the walking activity, whereas the current study considers eighteen activities. We can conclude that none of these seven studies fully cover any of the five research questions posed in Section I.

The six studies included in the second grouping, which are described in Section II-D, cover non-gait biometrics that uses a smartphone and/or smartwatch. These studies differ from the current study in several important ways. Four of the six studies do not use the accelerometer or gyroscope and do not implement motion-based biometrics, so they have little in common with the current study. Beyond that, only one study evaluates the identification task [29] or utilizes a smartwatch, and none of the six studies analyze more than one activity. Hence none of these six studies fully addresses any of the five research questions that are the focus of the current study.

The two studies most similar to the current study include the walking activity and either three [12] or five [30]

additional activities of daily living. This is much less that the eighteen activities evaluated in the current study and, more importantly, none of the additional activities include diverse hand-oriented activities such as typing, writing, brushing teeth, and eating. Overall, one can conclude that the current study is very different from prior research and is unique in its ability to address five research questions posed in Section I.

B. SUMMARY OF CONCLUSIONS

This study shows that the best biometric performance occurs when using the smartphone and smartwatch together, with the accelerometer sensor on both devices performing about as well as when the accelerometer and gyroscope on both devices are used. The study also demonstrates that substantial improvements in biometric performance are achieved by using 50 seconds of data for evaluation rather than just 10 seconds of data. Biometric performance is also quite sensitive to the amount of training data. The performance for the authentication task improves rapidly as more training data is added, and the improvement was continuing when the maximum of 170 seconds of data per activity was reached. Identification performance similarly improved rapidly but began to reach a plateau with 2 minutes of training data per activity. Overall, it appears that good performance is achievable using relatively little training data. Also, while most studies focus solely on authentication, this study showed that identification is feasible using activities of daily living—at least with fifty-one subjects.

The research in this article also showed that while different activities have varying levels of biometric effectiveness, all perform reasonably well and within a relatively narrow range. Thus, biometric authentication and identification is feasible with activities other than walking. When considering the practicality of the activity, walking, typing, and clapping are the best overall activities for on-demand biometrics. Given that all the eighteen activities that were evaluated are useful for biometrics, continuous biometrics using the naturally occurring activities of daily living should be feasible. Prior activity recognition research demonstrated that it is possible to identify the eighteen activities used in this study with good accuracy [11], [31], so it is feasible to implement a two-stage approach—where activity recognition is used to identify an activity and the associated biometric model is then applied to authenticate or identify the subject.

C. FUTURE WORK

The research described in this article can be extended in several important ways. One key way would be to include many more activities and implement the two-stage biometric system described earlier, which would employ activity recognition in the first stage. A further step would be to fully implement continuous biometric authentication as a subject performed their normal daily activities. Regarding the research methods employed, it would be interesting to apply and evaluate the use of one-class learning for biometric authentication using the smartphone and smartwatch. The current methodology employed neither feature normalization nor feature selection, and there would be value in trying both of these (although neither is likely to have any effect on Random Forest, the best performing algorithm). Finally, there has been great success in applying deep learning to complex problems, and while the amount of data employed in this study is not enormous, deep learning has the potential to improve the biometric performance by automatically learning new feature representations.

One of the goals of this research is to progressively move to a behavioral biometrics system that operates continuously as a subject performs their normal daily activities. The presumed approach involves a two-stage process, where an activity is recognized and then the biometrics model for that activity is applied. An alternative approach is possible, where a generalized activity-based model is formed that works for all activities. This would be an interesting area of future research.

ACKNOWLEDGMENT

The authors would like to thank all of the WISDM lab members who assisted with the data collection effort.

REFERENCES

[1] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proc. IEEE*, vol. 91, no. 12, pp. 2021–2040, Dec. 2003. doi: 10.1109/JPROC.2003.819611.

[2] A. Jain, A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Boston, MA, USA: Springer, 2011.

[3] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2010, pp. 306–311.doi: 10.1109/IIHMSP.2010.83.

[4] C. Nickel, H. Brandt, and C. Busch, "Benchmarking the performance of SVMs and HMMs for accelerometer-based biometric gait recognition," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2011, pp. 281–286. doi: 10.1109/ISSPIT.2011.6151574.

[5] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, and M. Savvides, "Gait-id on the move: Pace independent human identification using cell phone accelerometer dynamics," in *Proc. IEEE 5th Int. Conf. Biometrics, Theory, Appl. Syst.*, Sep. 2012, pp. 8–15. doi: 10.1109/BTAS.2012.6374552.

[6] T. Hoang, T. D. Nguyen, C. Luong, S. Do, and D. Choi, "Adaptive cross-device gait recognition using a mobile accelerometer," *J. Inf. Process. Syst.*, vol. 9, no. 2, pp. 333–348, 2013. doi: 10.3745/JIPS.2013.9.2.333.

[7] A. H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," in *Proc. IEEE 7th Int. Conf. Biometrics Theory, Appl. Syst.*, Sep. 2015, pp. 1–6. doi: 10.1109/BTAS.2015.7358794.

[8] N. Al-Naffakh, N. Clarke, F. Li, and P. Haskell-Dowland, "Unobtrusive gait recognition using smartwatches," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2017, pp. 1–5.doi: 10.23919/BIOSIG.2017.8053523.

[9] G. Cola, M. Avvenuti, F. Musso, and A. Vecchio, "Gait-based authentication using a wrist-worn device," in *Proc. 13th Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, Nov. 2016, pp. 208–217. doi: 10.1145/2994374.2994393.

[10] Y. Yang, F. Hong, Y. Zhang, and Z. Guo, "Person authentication using finger snapping—A new biometric trait," in *Proc. Chin. Conf. Biometric Recognit.* Lecture Notes in Computer Science, vol. 9967, Z. You, Ed., 2016, pp. 765–774.

[11] G. M. Weiss and A. E. O'Neill, "Smartphone and smartwatch-based activity recognition," Dept. Comput. Inf. Sci., Fordham Univ., Bronx, NY, USA, Tech. Rep., Jul. 2019.

[12] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Cell phone-based biometric identification," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory Appl. Syst.*, Sep. 2010, pp. 1–7. doi: 10.1109/BTAS.2010.5634532.

[13] K. Yoneda and G. M. Weiss, "Mobile sensor-based biometrics using common daily activities," in *Proc. IEEE 8th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf.*, Oct. 2017, pp. 584–590. doi: 10.1109/UEMCON.2017.8249001.

[14] K. Delac and M. Grgic, "A survey of biometric recognition methods," in *Proc. 46th Int. Symp. Electron. Mar.*, Jun. 2004, pp. 184–193.
[15] A. K. Jain and A. Kumar, "Biometric recognition: An overview," in *Second Generation Biometrics: The Ethical, Legal and Social Context* (The International Library of Ethics, Law and Technology), vol. 11, E. Mordini and D. Tzovaras, Eds. Dordrecht, The Netherlands: Springer, 2012. doi: 10.1007/978-94-007-3892-8_3.
[16] I. Bouchrika, "A survey of using biometrics for smart visual surveillance: Gait recognition," in *Surveillance in Action* (Advanced Sciences and Technologies for Security Applications), P. Karampelas and T. Bourlai, Eds. Cham, Switzerland: Springer, 2018. doi: 10.1007/978-3-319-68533-5_1.
[17] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Performance evaluation of fingerprint verification systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 3–18, Jan. 2005. doi: 10.1109/TPAMI.2006.20.
[18] A. Alzubaidi and J. Kalita, "Authentication of smartphone users using behavioral biometrics," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1998–2026, 3rd Quart., 2016. doi: 10.1109/COMST.2016.2537748.
[19] D. Gafurov, "A survey of biometric gait recognition: Approaches, security, and challenges," in *Proc. Annu. Norwegian Comput. Sci. Conf.*, Nov. 2007, pp. 1–12.
[20] R. J. Orr and G. D. Abowd, "The smart floor: A mechanism for natural user identification and tracking," in *Proc. Extended Abstr. Hum. Factors Comput. Syst.*, Apr. 2000, pp. 275–276. doi: 10.1145/633292.633453.
[21] D. Gafurov and E. Snekkenes, "Gait recognition using wearable motion recording sensors," *EURASIP J. Adv. Signal Process.*, vol. 2009, p. 7, Jan. 2009. doi: 10.1155/2009/415817.
[22] A. Annadhorai, E. Gutenberg, J. Barnes, K. Harage, and R. Jafari, "Human identification by gait analysis," in *Proc. 2nd Int. Workshop Syst. Netw. Support Health Care Assist. Living Environ.*, Jun. 2008, p. 11.
[23] D. Gafurov, K. Helkala, and T. Søndrol, "Biometric gait authentication using accelerometer sensor," *J. Comput.*, vol. 1, no. 7, pp. 51–59, Nov. 2006.
[24] J. Mantyjarvi, M. Lindholdm, E. Vildjounaite, S.-M. Makela, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2005, pp. 973–976.
[25] T. Feng, Z. Liu, K.-A. Kwon, W. Shi, B. Carbunar, Y. Jiang, and N. Nguyen, "Continuous mobile authentication using touchscreen gestures," in *Proc. IEEE Conf. Technol. Homeland Secur.*, Nov. 2012, pp. 451–456.
[26] B. Draffin, J. Zhu, and J. Zhang, "KeySens: Passive user authentication through micro-behavior modeling of soft keyboard interaction," in *Mobile Computing, Applications, and Services*. Cham, Switzerland: Springer, 2014, pp. 184–201. doi: 10.1007/978-3-319-05452-0_14.
[27] A. Buriro, B. Crispo, F. Del Frari, and K. Wrona, "Touchstroke: Smartphone user authentication based on touch-typing biometrics," in *Proc. Int. Conf. Image Anal. Process.*, 2015, pp. 27–34.
[28] M. Kunz, K. Kasper, H. Reininger, M. Möbius, and J. Ohms, "Continuous speaker verification in realtime," in *Proc. Special Interest Group Biometrics Electron.*, Sep. 2011, pp. 79–88.
[29] F. Ciuffo and G. M. Weiss, "Smartwatch-Based transcription biometrics," *Proc. 8th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf.*, Oct. 2017, pp. 145–149. doi: 10.1109/UEMCON.2017.8249014.
[30] M. Ehatisham-ul-Haq, J. Loo, K. Shuang, S. Islam, U. Naeem, and Y. Amin, "Authentication of smartphone users based on activity recognition and mobile sensing," *Sensors*, vol. 17, no. 9, p. 2043, Sep. 2017. doi: 10.3390/s17092043.
[31] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform.*, Feb. 2016, pp. 426–429. doi: 10.1109/BHI.2016.7455925.
[32] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, "Design considerations for the WISDM smart phone-based sensor mining architecture," in *Proc. 5th Int. Workshop Knowl. Discovery Sensor Data*, Aug. 2011, pp. 25–33. doi: 10.1145/2003653.2003656.
[33] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *Proc. Int. Workshop Wearable Implant. Body Sensor Netw.*, Apr. 2006, pp. 1–4. doi: 10.1109/BSN.2006.6.
[34] D. Dua and E. K. Taniskidou. (2017). UCI Machine Learning Repository. University of California. School of Information and Computer Science. Irvine, CA, USA. [Online]. Available: http://archive.ics.uci.edu/ml
[35] JR. Kwapisz, G.M. Weiss, and S.A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explor. Newslett.*, vol. 12, no. 2, pp. 74–82, Dec. 2010. doi: 10.1145/1964897.1964918.
[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[37] G. M. Weiss, J. W. Lockhart, T. T. Pulickal, P. T. McHugh, I. H. Ronan, and J. L. Timko, "Actitracker: A smartphone-based activity recognition system for improving health and well-being," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal.*, Montreal, QC, Canada, Oct. 2016, pp. 682–688. doi: 10.1109/DSAA.2016.89.

GARY M. WEISS received the B.S. degree from Cornell University, in 1985, the M.S. degree from Stanford University, in 1986, and the Ph.D. degree from Rutgers University, in 2003, all in computer science. He was with AT&T Labs, from 1985 to 2004. In 2004, he joined Fordham University, where he is currently an Associate Professor with the Department of Computer and Information Science. He also directs the Wireless Sensor Data Mining Lab, which explores how smartphones and other mobile devices can support human activity recognition and biometrics. His work has been funded by the U.S. National Science Foundation, Google, and several industry partners. He has published over 70 articles in machine learning and data mining. He has received the Innovative Application of Artificial Intelligence Award from AAAI, and in 2015, he received the Five-Year Highest Impact Award for a biometrics article at the IEEE BTAS-2010 Conference. He is an Associate Editor of the *Knowledge and Information Systems* journal and serves on the Editorial Board of several journals.

KENICHI YONEDA received the B.S. and M.S. degrees in computer science from Fordham University, in 2015 and 2017, respectively. While at Fordham University, he participated in research in mobile activity recognition and mobile biometrics, as part of the Wireless Sensor Data Mining (WISDM) Lab.

THAIER HAYAJNEH received the B.S. and M.S. degrees in electrical and computer engineering from the Jordan University of Science and Technology, in 1997 and 1999, respectively, and the M.S. and Ph.D. degrees in information sciences with specialization in cybersecurity and networking from the University of Pittsburgh, in 2005 and 2009, respectively. He is currently a University Professor of computer science, the Founder and the Director of the Fordham Center for Cybersecurity, and the Director of the Cybersecurity Graduate Program at Fordham University. He has published over 80 articles in reputable journals and conferences. His research interests include system's security, wireless security, applied cryptography, blockchain and cryptocurrency, and the IoT security, privacy, and forensics. He served on several NSF Cybersecurity Review Panels and serves as a CAE Reviewer and Mentor for NSA. He is serving/served as the Editor-in-Chief, an Editor, the Program Chair, and a Guest Editor for several prestigious journals and leading conferences.