AUToSen: Deep Learning-based Implicit Continuous Authentication Using Smartphone Sensors

Mohammed Abuhamad[†‡], Tamer Abuhmed[◇], David Mohaisen[‡], *Senior Member, IEEE*, and DaeHun Nyang[¶]
[†]INHA University, [‡]University of Central Florida, [◇]Sungkyunkwan University, [¶]Ewha Womans University.

Abstract*—Smartphones have become crucial for our daily life activities and are increasingly loaded with our personal information to perform several sensitive tasks including mobile banking, communication, and are used for storing private photos and files. Therefore, there is a high demand for applying usable authentication techniques that prevent unauthorized access to sensitive information. In this work, we propose AUTo*Sen*, a deep learning-based active authentication approach that exploits sensors in consumer-grade smartphones to authenticate a user. Unlike conventional approaches, AUTo*Sen* is based on deep learning to identify user distinct behavior from the embedded sensors with and without the user's interaction with the smartphone. We investigate different deep learning architectures in modeling and capturing users' behavioral patterns for the purpose of authentication. Moreover, we explore the sufficiency of sensory data required to accurately authenticate users. We evaluate AUTo*Sen* on a real-world dataset that includes sensors data of 84 participants' smartphones collected using our designed data-collection application. The experiments show that AUTo*Sen* operates accurately using readings of only three sensors (accelerometer, gyroscope, and magnetometer) with a high authentication frequency, e.g., one authentication attempt every 0.5 seconds. Using sensory data of one second enables an authentication F1-score of approximately 98%, FAR of 0.95%, FRR of 6.67%, and EER of 0.41%. While using sensory data of half a second enables an authentication F1-score of 97.52%, FAR of 0.96%, FRR of 8.08%, and EER of 0.09%. Moreover, we investigate the effects of using different sensory data at variable sampling periods on the performance of the authentication models under various settings and learning architectures.

***Index Terms*—Active authentication, continuous authentication, mobile sensing, deep learning-based authentication, smartphone authentication.**

I. Introduction

To date, the primary method for accessing smartphones is restricted to methods addressing the Point-of-Entry authentication, which rely on knowledge (*e.g.,* token or password) or physiological biometrics (*e.g.,* fingerprint or face). Knowledge-based techniques include several modalities such as passwords, pattern-based passwords, and password-based on biometric features [39]. Despite the simplicity and reliability of knowledge-based techniques, they are inherently vulnerable to several attacks such as shoulder surfing attack [34] and smudge attack [6]. Such vulnerabilities are tackled with physiological biometrics (*e.g.,* fingerprint and face), which enabled various stronger authentication techniques on smartphones [30], [39]. However, most of these techniques raise several privacy and usability concerns [50]. Both Knowledge-based and physiological biometrics-based methods fail to offer security and access control beyond the point-of-entry, *i.e.,* allowing for continuous authentication of the user. Moreover, such methods assume the same level of security across all applications once access to the device is granted, posing a significant vulnerability [12]. Therefore, there is an increasing need for efficient authentication methods to operate transparently and continuously based on behavioral biometrics of users. Most of today's smartphones are equipped with many built-in sensors, e.g., accelerometer, gyroscope, magnetometer, proximity sensor, barometer, *etc.* These sensors enable implicit authentication techniques that capture the user's behavioral characteristics from the readings of these sensors. Since this data can be obtained with or without the user's interaction, the sensors-based authentication techniques provide a convenient and intuitive access control for legitimate users. These authentication techniques are often referred to as "*transparent*, *continuous*, *implicit*, *active*, *passive*, *non-intrusive*, *non-observable*, *adaptive*, *unobtrusive*, and *progressive*" [30]. The implicit authentication works as a support method rather than an alternative to the point-of-entry conventional authentication techniques (using either knowledge-based or physiological biometrics-based authentication), where the primary task of the continuous authentication module is to detect any adversary, who attempts to control the smartphone, to prompt the user for reauthentication and regaining control. Therefore, both explicit and implicit authentication methods should co-exist to ensure device security. An alternative setting for utilizing implicit authentication alongside the point-of-entry approaches is the two-factor authentication (2FA), where the implicit authentication is used as an additional factor to the primary authentication modality.

In this work, we propose AUTo*Sen*, a deep learning-based implicit authentication technique. AUTo*Sen* exploits data from the smartphone embedded sensors to capture users' behavioral patterns for authentication. Our proposed approach has the following advantages. ❶ Unlike the conventional authentication techniques which prompt users on a specific

This work was supported in part by NRF grant NRF- 2016K1A1A2912757 and by the Collaborative Seed Award Program of Cyber Florida. M. Abuhamad is with the Department of Computer Engineering at Inha University, South Korea, and the Department of Computer Science at the University of Central Florida, FL, USA. T. Abuhmed is with the college of computing at Sungkyunkwan University, South Korea (tamer@skku.edu; corresponding author). D. Mohaisen is with the Department of Computer Science and the University of Central Florida, FL, USA (mohaisen@ucf.edu; corresponding author). D. Nyang is with the Department of Cyber Security at Division of Software Science and Engineering, Ewha Womans University, South Korea. (nyang@ewha.ac.kr; corresponding author).

time, our approach keeps implicitly authenticating the user in the background. ❷ This research measures the performance of AUTo*Sen* in the task of continuous authentication in a realistic scenario, where the users use their own smartphones freely without any usage constraints, through a non-invasive background service that keeps authenticating users with or without their interactions with their smartphones. ❸ The proposed approach does not require any sensitive software or hardware permissions that could invade the user's privacy. ❹ We have conducted comprehensive experiments and illustrated the performance of our proposed approach in different settings.

We explore the use of different sensors to capture distinct user's behavioral characteristics. Starting with five sensors, our experiments show that readings from only three sensors—accelerometer, gyroscope, and magnetometer—are sufficient to model users' behavior with high accuracy for the authentication purpose. Considering a sampling rate of 64Hz, readings of sensory data within one second provides an F1-score of approximately 98%. Even when using readings for half a second, the three sensors data allowed an authentication F1-score higher than 97%.

Several works investigated user's identification based on specific activities (*i.e.,* walking, standing, sitting, running, walking upstairs, and walking downstairs) and only when the user is interacting with the smartphone [16], [36]. However, *to the best of our knowledge, this work is the first to propose a high-frequency deep learning-based approach for continuous user authentication on smartphone using built-in sensors without setting any constraints on the user interaction or activity types.* After building the authentication model, AUTo*Sen* is capable of operating in a high-frequency manner as it requires the readings of a user's sensors data for a short period (*e.g.,* 0.5 seconds or one second) to passively authenticate the user during the daily activities. We exploit the sensors embedded in smartphones to design and build an accurate, efficient, and continuous authentication system and thoroughly evaluate it against several options; e.g., number of sensors, model size, and deep learning architectures.

Contributions. Our contributions are summarized as follows.

- We propose AUTo*Sen* of a deep learning-based implicit and continuous authentication system using built-in smartphone sensors. The experiments are conducted using a real-world dataset that includes data of 84 participants. The collection of data is conducted using our data-collection application.
- The proposed approach is light-weight and does not require a complicated feature extraction process that can be computationally demanding and energy consuming, by exploiting the capabilities of LSTM to capture users' behavioral traits directly from the readings of sensory data.
- We explore the effects of using a collection of different embedded sensors in the user authentication task modeled by different LSTM architectures and settings. Using AUTo*Sen*, the experiments show that readings of three sensors, namely, accelerometer, gyroscope, and magnetometer, are adequate for a highly accurate user authentication process.
- We investigate the sampling period required to enable accurate active user authentication. Our experimental results show that AUTo*Sen* is capable of operating in real-time with the requirement of readings collected within half a second.
- We conducted extensive experiments to evaluate AUTo*Sen* with different evaluation metrics and under different settings. We show that our approach authenticates users with an average F1-score of roughly 98% across all users using the reading of three sensors. Moreover, we report state-of-the-art results in terms of FAR, FRR, and EER since AUTo*Sen* achieves an average FAR of 0.95%, FRR of 6.67%, and EER 0.41% across all users when the sampling period is one second. While a sampling period of half a second enables an average FAR of 0.96%, FRR of 8.08%, and EER 0.09%.
- Using simulation experiments, we demonstrate the usability of AUTo*Sen* through our reported FAR and FRR scores. For example, using the data of three sensors (accelerometer, gyroscope, and magnetometer) with a sampling period of one second, legitimate users are expected to be authenticated within two seconds with a probability of 99.56% and within three seconds with a probability of 99.97%. When the sampling period is set to half a second, AUTo*Sen* is expected to authenticate legitimate users within the first second with a probability of 99.34% and with a probability of 99.99% for authentication within only two seconds.

Organization. The remainder of the paper is structured as follows. We review related works in Section II. In Section III, we present our deep learning-based approaches for continuous and implicit authentication on smartphones. We proceed with detailed experimental results from our approaches in Section IV. We provide our conclusion in Section V.

II. Related work

Recent studies show the significance of sensory data collected from mobile devices in a variety of applications, such as modeling human behavior [16], user authentication [46], [5], activity recognition [7], [16], and healthcare assessment and monitoring [14], [10], among others [42], [44]. User authentication using mobile sensory data has shown remarkable results in exploiting users' physiological and behavioral biometrics [30]. Various techniques have been developed to authenticate a mobile user using behavioral characteristics, where a background process continuously captures user's usage of the device to perform an active and transparent authentication, e.g., using hand-waving [15], gait [18], [13], [31], touchscreen [6], [15], [43], electrocardiography (ECG) [5], keystroke [43], [28], [20], voice [33], [26], signature [11], [27], and general profiling [3], [35]. These approaches continuously authenticate mobile users by measuring their behavioral characteristics while interacting with their mobile.

Keystroke dynamics are used for continuous authentication utilizing keystroking features. Those features include time, *i.e.,* the latency between the press and release of a key (called dwell or hold time), and latency between the release of one key and the press of next key (called flight time). Other features are also obtained while stroking the keys, such as the device orientation, finger pressure size, and accelerometers [28], [20]. An advantage of this approach is being noninvasive and transparent, although it achieves lower accuracy compared

TABLE I
A summary of the related work. In the features, all data is extracted from, 'Ac' : Accelerometer, 'Gr': Gyroscope, 'Ma': Magnetometer, 'El': Elevation, 'Co': Compass, 'Ga': Gravity, 'Or': Orientation of smartphones.

Reference	Technique	Features-level	Participants	Performance	Authentication Time
Draffin *et al.* [15]	Behavioural biometric	Keystrokes	13	FAR = 14.0%; FRR = 2.2%	5∼15 keystrokes
Mondal *et al.* [28]	Behavioural biometric	Keystrokes	64	90%	500 keystrokes
Song *et al.* [38]	Behavioural biometric	Eye movement	10	88.73%	10s
Zhang *et al.* [46]	Behavioural biometric	Eye movement	30	90.3-93.1%	130s
Juan *et al.* [5]	Biometric	Mobile ECG sensors	10	81.82%	4 ∼ 10 s
Juan *et al.* [29]	Biometric	Wearable Sensors	37	EER =1.9%	1 ∼ 4 min.
Lee *et al.* [22]	Physical activity	(Ac,Ma,Or)	4	90%	20s
Kayacik *et al.* [21]	Physical activity	(WiFi, Ac, GPS, light)	7	53.2∼ 99.4%	>122s
Ehatisham *et al.* [16]	Physical activity	(Ac,Gr,Ma)	10	97.95%	180s
Zhu *et al.* [48]	Physical activity	(Ac,Gr,Co)	20	TPR = 71.30%, FPR= 31.1%	4.96s
Lee *et al.* [23]	Physical activity	(Ac,Gr)	35	FRR = 0.9%, FAR = 2.8%	6s
Amini *et al.* [4]	Physical activity	(Ac,Gr)	47	F1-score = 96.7%	20s
Shen *et al.* [36]	Physical activity	(Ac, Gr, Touch)	48	FRR = 6.85%, FAR = 5.01%	2.89s - 3.31s
Sitova *et al.* [37]	Physical activity	(Ac,Gr,Ma)	100	EER = 7.16% ∼ 10.05%	∼ 60s
Fenu *et al.* [17]	Physical activity	(Camera,Ac,Gr,Ma,Touch)	100	EER = 5.95% ∼ 17.56%	n.a.
Li *et al.* [25]	Physical activity	(Ac,Gr)	100	EER = 4.66%	5s
Shen *et al.* [35]	Physical activity	(Ac,Or,Ma,Gr)	102	FRR = 5.03%, FAR = 3.98%	8s
Zhu *et al.* [49]	Physical activity	(Ac,Gr, Ga)	1,513	Accuracy = 95.6%	3.2s
This work	Physical activity	(Touch, Ac,Gr, Ma, El)	84	FRR = 8.47%, FAR = 1.72%	1s
This work	Physical activity	(Ac,Gr, Ma, El)	84	FRR = 7.62%, FAR = 2.31%	1s
This work	Physical activity	(Ac,Gr, Ma)	84	**FRR = 6.67%**, **FAR = 0.95%**	1s
This work	Physical activity	(Touch, Ac,Gr, Ma, El)	84	FRR = 9.16%, FAR = 1.53%	0.5s
This work	Physical activity	(Ac,Gr, Ma, El)	84	FRR = 9.87%, FAR = 1.65%	0.5s
This work	Physical activity	(Ac,Gr, Ma)	84	**FRR = 8.08%**, **FAR = 0.96%**	0.5s

to other techniques. Moreover, it only works when the user interacts with the keyboard [3].

Using other biometric modalities, Zhang *et al.* [46] proposed an eye movement-based continuous authentication technique by tracking the eye movements of a VR headset wearer when there are visual stimuli. Their system can detect 91.2% of all imposters within 130 seconds. Similar work has been done on smartphones to track human eye movement through the built-in front camera to explore gaze patterns for authentication [38]. Their experiments show that the system achieves 88.73% accuracy after tracking users' eyes for 10 seconds. Juan *et al.* [5] proposed a mobile biometric authentication algorithm based on electrocardiogram (ECG) collected from ECG electrodes installed on the mobile cover. Their approach collects users' heartbeats for four seconds to achieve an accuracy of 81.82%. Processing and reaction, feasible deployment, and robustness to changes in environmental conditions (*e.g.,* light and noise conditions) are all bottlenecks of the approach.

Since most smartphones are equipped with a variety of sensors, including motion (*i.e.,* gravity, accelerometer, gyroscope, and magnetometer), environmental (*i.e.,* light, temperature, barometer, and proximity), and position sensors (*i.e.,* GPS and compass), recent works have used these sensors for authentication [35], [16], [4], [23]. Ehatisham *et al.* [16] designed a continuous authentication scheme that recognizes smartphone users based on their physical activity patterns using accelerometer, gyroscope, and magnetometer sensors. In their experiment, an analysis of the motion sensors was conducted when the smartphone is strictly placed at five different positions on the user's body. Amini *et al.* [4] proposed *DeepAuth*, an LSTM-based user authentication method, which uses accelerometer and gyroscope data to capture behavioral patterns. The results of their experiments, conducted on data collected from 47 participants with 10-13 minutes each, have shown an authentication accuracy of 96.7% in a window of 20 seconds. Zhu *et al.* [49] proposed *RiskCog*, a method that requires 3.2 seconds to validate users using data collected from accelerometer, gyroscope and gravity sensors with an accuracy of 93.8% and 95.6% for steady and moving users, respectively. Li *et al.* [25] proposed five sensory data augmentation processes to authenticate users with *SensorAuth*, and have shown an EER of 4.66% with a time window of 5 seconds.

Zhu *et al.* [47] proposed a method based on users' phone-skating behavior using the motion sensory data. The experiments were conducted using data of 20 participants and have shown an average EER of 1.2%. Using gaits as a feature, several works are proposed; *e.g.,* Damaševičius *et al.* [13] and Fernandez-Lopez *et al.* [18]. Li and Bours [24] proposed an authentication scheme for smartphones using WiFi and accelerometer to enable users accessing an application within three seconds, with an average EER of 9.19%. Buriro *et al.* [8] proposed an authentication method based on the user's hand micro-movements and timing differences while users entering 10-digit strokes. The experiments included 97 participants performing several activities. The results show that their approach was capable of authenticating users with True Acceptance Rate (TAR) of 85.77% and False Acceptance Rate (FAR) of 7.32%. Lee *et al.* [22] proposed an SVM based approach for implicit user authentication. They fed their system with three mobile sensor readings to train with 7 seconds of mobile data for training and 20 seconds detection, with an accuracy of 90%. Lee *et al.* [23] showed that combining sensors' readings from the user's smartphone and other wearable devices can enhance the authentication accuracy; they reported accuracy of 98.1%, FRR of 0.9%, and FAR of 2.8% by combining reading from smartphone and smartwatch of 35 users within

6 seconds. Kayacik *et al.* [21] designed a data-driven approach that is temporally and spatially aware of the mobile user using several hard and soft sensors. However, their approach required more than 122 seconds to authenticate a user and more than 717 seconds to detect an imposter. Zhu *et al.* [48] proposed a gesture-based authentication mechanism when using the smartphone. The model generates a sureness score to evaluate the necessity of an authentication. Their approach achieved 75% accuracy for identifying users and 71.3% accuracy for detecting non-owners with 13.1% false alarms.

A summary of related works and a comparison between them, including ours, across multiple variables, are in Table I.

III. AUTo*Sen*: An Overview

We propose AUTo*Sen*, a deep learning-based active authentication system using smartphone sensors. We explore the distinctive usage and activity patterns of smartphone users captured by the sensory data for authentication. AUTo*Sen* is a real-time authentication system, which collects readings from the embedded sensors and feeds the data to an authentication model to validate the smartphone users. The authentication models are designed using LSTM-based architectures to process sequential sensory data records to capture the behavioral patterns of users when holding their smartphones. Regardless of users' activities (*e.g.,* texting, voice or video chatting, internet surfing, jogging, exercising, *etc.*), AUTo*Sen* aims to exploit the distinctive behavioral patterns of users for the authentication. Using sensory data to capture such behavioral patterns requires performing several tasks, including data preprocessing, temporal alignment, feature extraction, and sequential modeling. Performing this task is very challenging as continuous authentication entails real-time user validation, robust feature extraction, an accurate authentication, and acceptable usability. This study addresses these challenges.

Assuming that each user with own pattern for activities, we focus on validating users based on accumulative sensors data collected from their mobile during a period of time. The main process flow of AUTo*Sen* includes ❶ collecting the internal sensors data of the smartphone for individual users using background service during their activities, ❷ preprocessing the collected data, ❸ feeding this data to the authentication models, which are trained to capture the behavioral patterns of individual users, for user validation. Figure 1 illustrates the steps of AUTo*Sen* system which mainly consists of: data collection of readings from the built-in sensors of the smartphone, data preprocessing (removing corrupted data, temporal alignment of all sensors data, and constructing sequences to feed them to the authentication model), user authentication model, and finally the user verification step in real-time using short-period sequences fed to the built model.

A. Data Collection

We conduct our experiments on Android smartphones to implement and evaluate AUTo*Sen* using different approaches and sensors data. For sensor data collections, we implemented an Android data-collection application that transparently operates in the background to collect data from five embedded sensors, namely: screen touch, accelerometer, gyroscope, magnetometer, and elevation. The data-collection application operates continuously as long as it runs to record sensor data with timestamps. We collected mobile sensor data of 84 participants including students, staff personnel, and professors. They ranged in age from 19 to 37, with an average age of 25 years (SD = 4.49). All participants were skilled smartphone users with at least one year's experience. All participants conducted their usual routine usage of smartphones for the data collection task and remained in the study for the authentication evaluation. For the data collection task, participants were asked to run the data-collection application for five days to obtain large data samples to investigate the effects of different sensors' data in capturing the behavioral patterns of users. The collected sensor data includes: ❶ *screen touches* (*i.e.,* touch sliding or touch tapping) when users touch their smartphones during smartphone usage, which covers a wide range of applications such as web surfing, document/email reading, call making, picture browsing, instant chatting, *etc.*, ❷ *sensor data* (*i.e.,* accelerometer, gyroscope, magnetometer, and elevation), and ❸ *timestamps* for data readings.

The accelerometer indicates the mobile orientation and measures gravitational acceleration in meter per second squared. Each accelerometer reading is a vector $Ac \in \mathbb{R}^3$ of x, y, and z values that represent the phone coordinates. The gyroscope sensor provides a three-dimensional vector $Gr \in \mathbb{R}^3$ for the angular rotation in radians per second along the axes. The magnetometer was used to report the magnetic field in micro Tesla, and each magnetometer reading is a vector $Ma \in \mathbb{R}^3$. We explore the effect of different sensors in capturing the behavioral patterns of users. We created four different datasets:

1) **Five-Sensor Dataset (ToAcGrMaEl):** This dataset includes data readings of five sensors (namely, Touch actions, accelerometer, gyroscope, magnetometer, and elevation).
2) **Four-Sensor Dataset (AcGrMaEl):** This dataset includes data readings of four sensors (namely, data of accelerometer, gyroscope, magnetometer, and elevation).
3) **Three-Sensor Dataset (AcGrMa):** This dataset includes data readings of three sensors (namely, accelerometer, gyroscope, and magnetometer).
4) **Two-Sensor Dataset (AcGr):** This dataset includes data readings of two sensors (accelerometer and gyroscope).

B. Data Preprocessing

Sensor data collected from smartphones requires a preprocessing stage for possible noise handling and temporal alignment for sequence generation. The targeted sensory data includes readings of five sensors, which are touch actions, accelerometer, gyroscope, magnetometer, and elevation. Denote the collected data reading as $X_i^{(t)} \in \mathbb{R}^m$ for the user i at time step t, where m as the total dimension of collected data. For example $m = 11$ when using the five sensors readings, given the touch data $To \in \mathbb{R}^1$ as the frequency of touch actions within time t, accelerometer reading $Ac \in \mathbb{R}^3$, gyroscope reading $Gr \in \mathbb{R}^3$, magnetometer reading $Ma \in \mathbb{R}^3$, and elevation reading $El \in \mathbb{R}^1$.

Handling Missing Values. Given the nature of the sensory data, missing values are handled at the individual sensor's

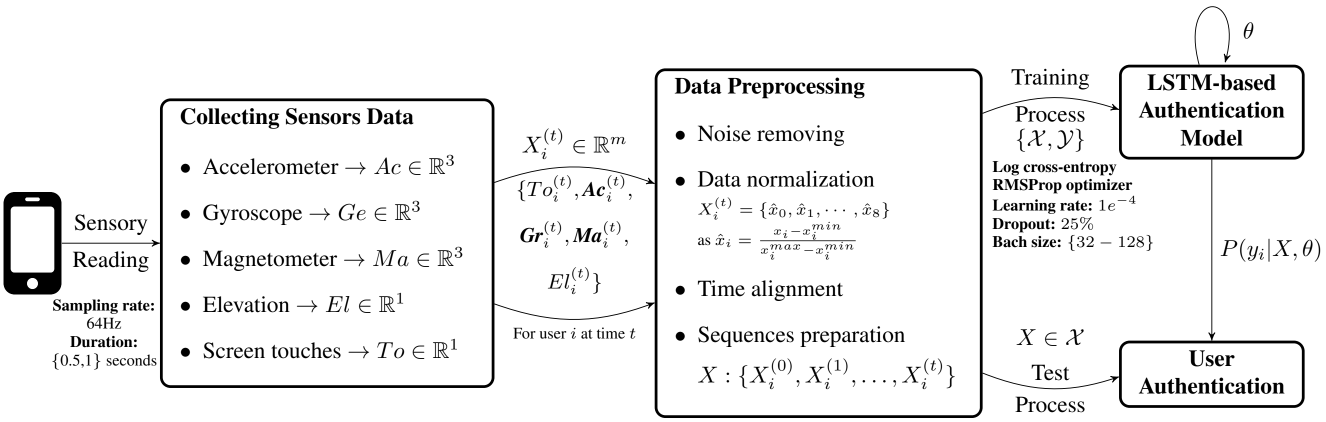

Sensory
Reading
Sampling rate:
64Hz
Duration:
{0.5,1} seconds
Collecting Sensors Data
Accelerometer → $Ac \in \mathbb{R}^3$
Gyroscope → $Ge \in \mathbb{R}^3$
Magnetometer → $Ma \in \mathbb{R}^3$
Elevation → $El \in \mathbb{R}^1$
Screen touches → $To \in \mathbb{R}^1$
$X_i^{(t)} \in \mathbb{R}^m$
$\{To_i^{(t)}, \boldsymbol{Ac}_i^{(t)}, \boldsymbol{Gr}_i^{(t)}, \boldsymbol{Ma}_i^{(t)}, El_i^{(t)}\}$
For user i at time t
Data Preprocessing
Noise removing
Data normalization
$X_i^{(t)} = \{\hat{x}_0, \hat{x}_1, \cdots, \hat{x}_8\}$
as $\hat{x}_i = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}$
Time alignment
Sequences preparation
$X : \{X_i^{(0)}, X_i^{(1)}, \ldots, X_i^{(t)}\}$
Training
Process
$\{\mathcal{X}, \mathcal{Y}\}$
Log cross-entropy
RMSProp optimizer
Learning rate: $1e^{-4}$
Dropout: 25%
Bach size: $\{32 - 128\}$
θ
LSTM-based
Authentication
Model
$P(y_i|X, \theta)$
$X \in \mathcal{X}$
Test
Process
User
Authentication

Fig. 1. AUTo*Sen*: an implicit authentication system overview.

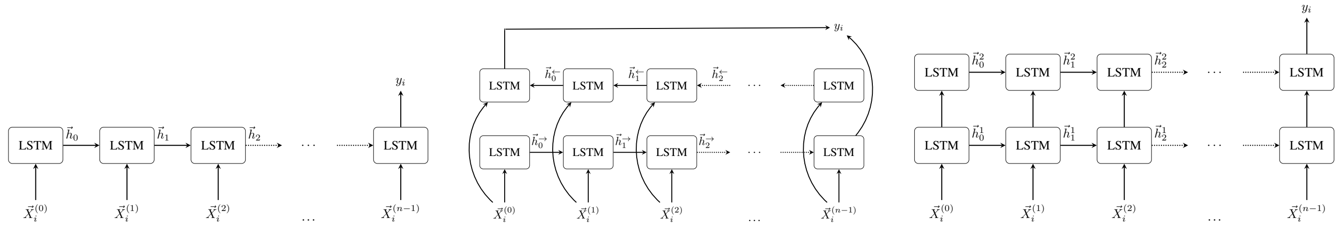
level using the specified time step. For a specific timeframe, *i.e.,* the sampling period (*e.g.,* one or half a second), missing values of touch actions indicate empty records while missing values from other sensors might be due to a cold start, reading stability, or sensor malfunction. Missing values from the touch sensor are assigned a zero value. Missing values from the elevation readings are set to the last observed value. Missing values from other sensors are imputed using the mean value of the previously observed data points within a window of size five (sliding window). This process of data imputation is a common practice for handling missing values based on the statistics (*e.g.,* minimum, maximum, mean, or median) of surrounding data points, which generate adequate estimates of missing data points [45].

Data Normalization. We normalize the collected data to minimize the effect of noise during the data collection stage. Moreover, since most of the dimensions of an input data $X_i^{(t)} = x_0, x_1, \cdots, x_{m-1} \in \mathbb{R}^m$ scale the values of readings from different sensors to a range between zero and one ($X_i^{(t)} \in \mathbb{R} \mid \forall\ x_i \in [0, 1]$), it would help the machine learning to weigh the effects of separate dimensions. The normalization process operates on data collected within a period of time (*e.g.,* one or five seconds), in which normalized data points are calculated $X_i^{(t)} = \hat{x}_0, \hat{x}_1, \cdots, \hat{x}_{m-1}$ where $\hat{x}_i = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}} \in \mathbb{R} \mid \forall x_i \in [0, 1]$, and x_i^{min} and x_i^{max} are the minimum and maximum values of the dimension i within the specified timeframe, respectively. In our experiments, we use five seconds as the timeframe for data normalization (*i.e.,* continuously observing the maximum and minimum values of each dimension for five or ten readings when using either the one or half a second sampling period, respectively)—because it shows the best trade-off between effectiveness and efficiency.

Sequences Generation. The readings of sensory data are collected with a sampling rate of 64Hz, *i.e.,* obtaining 64 readings per second. Even though most current devices support higher sampling rates, using 64Hz as a sampling rate provides a sufficient amount of data for authentication without exhausting the device resources (energy and computation). For generating sequences of sensory data, all sensor readings from all sensors are aligned. Assuming we used the data from all sensors, the aligned data can be represented as $X_i^{(t)} = To_i^{(t)}, Ac_i^{(t)}, Gr_i^{(t)}, Ma_i^{(t)}, El_i^{(t)}$. In our experiments, we use a sampling period of 0.5 seconds and one second to generate sequences of sensory data with a length of 32 and 64 readings per sequence, respectively. We note that the data collection app is designed to collect data continuously with the specified sampling rate for as long as it is running. For our experiments, we considered sensory readings that occur with active usage status, *i.e.,* readings are considered when a change is recorded by at least one of the motion sensors (accelerometer and gyroscope) as an indication of user activity (even without directly handling the phone). The readings are considered in inactive usage status if no changes are recorded for five seconds. Eliminating readings from inactive usage status provides a more practical and realistic scenario for modeling behavioral patterns of users.

C. LSTM-based User Authentication

Typically, the assigned task is to authenticate an owner of a smartphone using his behavioral patterns extracted from the sensors of the smartphone. To capture these behavioral patterns from a sequence of sensors' readings makes the RNN as the best candidate for this task [45], [1]. AUTo*Sen* utilizes recurrent neural network with Long Short-Term Memory (LSTM) [19] models for user authentication. The authentication models estimate the probability of assigning input data to one of two classes, *i.e.,* binary classification over the legitimate user and impostors. For an input data $\{X_i^{(0)}, X_i^{(1)}, \cdots, X_i^{(n-1)}\}$, the authentication model of user i estimates the probability $P(y_i|X : \{X_i^{(0)}, X_i^{(1)}, \cdots, X_i^{(n-1)}\}, \theta)$, where θ is the LSTM parameters and $y_i = \{0, 1\}$. LSTM is a variant of RNN, proposed to overcome the problem of "vanishing" or "exploding" of gradients when processing long sequences [19], [9]. LSTM uses gating mechanism and memory cells $C_i^{(t)}$ to process sequences at different time steps by propagating relevant information and removing the irrelevant information. Given input data $X_i^{(t)}$, the previous hidden state $h_i^{(t-1)}$ and the previous memory cell $X_i^{(t-1)}$, an LSTM unit calculates the current hidden state $h_i^{(t)}$ and memory cell $C_i^{(t)}$. First, the LSTM calculates the values of four gates, namely: the



(a) Simple LSTM model.

(b) Bidirectional LSTM model.

(c) Multi-layers LSTM model.

Fig. 2. Different LSTM model architectures used for the authentication task.

input, forget, output, and input modulation gates ($\mathfrak{i}$, $\mathfrak{f}$, $\mathfrak{o}$, and $\mathfrak{g}$, respectively) as follows:

$$\begin{aligned}
\mathfrak{i} &= \text{sigmoid}(W_{x\mathfrak{i}}X_i^{(t)} + W_{h\mathfrak{i}}h_i^{(t-1)}),\\
\mathfrak{f} &= \text{sigmoid}(W_{x\mathfrak{f}}X_i^{(t)} + W_{h\mathfrak{f}}h_i^{(t-1)}),\\
\mathfrak{o} &= \text{sigmoid}(W_{x\mathfrak{o}}X_i^{(t)} + W_{h\mathfrak{o}}h_i^{(t-1)}),\\
\mathfrak{g} &= \tanh(W_{x\mathfrak{g}}X_i^{(t)} + W_{h\mathfrak{g}}h_i^{(t-1)}).
\end{aligned}$$

Where $\text{sigmoid}(x) = (1+e^{-x})^{-1}$ and $\tanh(x) = (e^{2x}-1)(e^{2x}+1)^{-1}$. Then, $C_i^{(t)}$ and $h_i^{(t)}$ are calculated as follows:

$$\begin{aligned}
C_i^{(t)} &= \mathfrak{f} \odot C_i^{(t-1)} \odot \mathfrak{g},\\
h_i^{(t)} &= \mathfrak{o} \odot \tanh(C_i^{(t)}).
\end{aligned}$$

Where $\odot$ is the element-wise multiplication. This mechanism allows propagating the gradient across long time sequences, since only relevant data passes through the input modulation gate, and filtered data propagates through the output gate after using the forget gate to remember necessary data. At the last time step $t = n-1$, the output probability is calculated as $P(y_i|X,\theta) = \text{sigmoid}(W_{hy}h_i^{(n-1)})$, and the output $y_i = 1$ if $P(y_i|X,\theta) \geq 0.5$.

For our experiments, we investigate the performance of the authentication task using three different LSTM-based architectures, namely, simple LSTM, bidirectional LSTM, and multi-layers LSTM. Figure 2(a) illustrates the first architecture, which is a simple LSTM-based authentication model. This architecture consists of a single layer of LSTM units. The second architecture is the bidirectional LSTM-based authentication model. Unlike the simple authentication model, where each RNN unit i uses the information from the previous RNN unit $i-1$ to generate its current state as well as propagate this information to the next state, the bidirectional LSTM-based authentication model at any time step access the past and future information to form the state of any given unit. By accessing the information of the previous time steps and future time steps, the model learns to better understand the context and eliminate ambiguity. The model incorporates two LSTMs trained to make the output decision. The first LSTM works on sequences $\{X_i^{(0)}, X_i^{(1)}, \cdots, X_i^{(n-1)}\}$, while the other RNN operates in these sequences from the opposite direction $\{X_i^{(n-1)}, X_i^{(n-2)}, \cdots, X_i^{(0)}\}$ as illustrated in Figure 2(b). LSTM-based architectures with multiple hidden layers are popular for their exceptional scalable capability of capturing complex patterns on a given data [1]. In this work, we also explore the capability of the multi-layers LSTM, shown in Figure 2(c), to authenticate a given user based on the behavioral patterns captured from the sensory data.

LSTM-based Model Architecture. The sensory data are fed into the LSTM-based model for user authentication. The simple LSTM model consists of one recurrent layer, while multi-layers LSTM consists of only two layers. Bidirectional LSTM model consists of one bidirectional layer, *i.e.,* two LSTMs operating in opposite directions. All hidden recurrent layers of all adopted architectures consist of a number of LSTM units ranging from 16 to 256 to evaluate the performance of models considering the breadth of their hidden layers. For all adopted architectures, the output layer is a sigmoid layer with one unit signaling the probability of authenticating the legitimate user.

Dataset Usage. The user authentication model is trained using data from the assigned user (the legitimate user) and other users (as impostors' data). Data from the legitimate user are labeled as positive, while data from the wrong users are labeled as negative. When training the model, we use data from the legitimate user and data from ten randomly selected users as impostors' signals with a size of five folds larger than the data size of the legitimate user (*i.e.,* the number of imposters' data records are five times larger than the legitimate records). We understand that this approach introduces the class-imbalance problem; however, it is essential to emphasize the distinction between legitimate and imposter behaviors. To address the class imbalance, we use class weights (percentage) to penalize the incorrect predictions and to weigh the loss during the training process. Since the training process of the authentication models follows a data-driven approach, we use stratified 10-fold cross-validation for the evaluation. The using 10-fold cross-validation is straightforward, where the model is evaluated on each fold while using the other nine folds for training. The reported results are the average of the ten results obtained in each fold. This method is adopted for training LSTM-models in all experiments.

Authentication Models Training Process. The training of the LSTM model is an optimization process to find a set of model's parameters that allows performing a specific task. Starting from random weights, the optimization process guided by the minimization of the log cross-entropy loss enables adjusting the model's weights in a supervised manner. The log cross-entropy, also known as binary cross-entropy, is defined as follows:

$$\text{loss}(\theta) = \frac{-1}{N}\sum_{n=1}^{N}[y_i \times \log(P_n) + (1-y_i) \times \log(1-P_n)],$$

where P_n refers to the conditional probability $P(y_i|X, \theta)$. We use RMSProp [41] optimization algorithm to train the authentication models. The RMSProp algorithm is set with a fixed learning rate of $1e^{-4}$ to scale the estimated gradient of every training step. For the regularization, we adopt *dropout* [40], which enhances the generalization capabilities of the model.

Training with Mini-batch Approach. For efficient training, we adopt a mini-batch approach, where a number of input samples are packed into a tensor of predefined dimensions, as $\left[batch\_size,\ sequence\_length,\ sample\_length\right]$, where $batch\_size$ is ranging from 32 to 128 samples with (64 or 32) readings per sample ($sequence\_length$) when using (one or half a second for data sampling, respectively), and 11, 10, 9, or 6 dimensions per readings ($sample\_length$) when using five, four, three, or two sensors for the input data, respectively.

D. Authentication Evaluation Metrics

Classification Accuracy Metrics. We report the F1-score in evaluating the performance of the authentication models. The F1-score is calculated as *F1-score* $= 2\times$(Recall $\times$ Precision)$\div$ (Recall $+$ Precision), where *Recall = (TP)* $\div$ *(TP* $+$ *FN)* and *Precision* $=$ *(TP)* $\div$ *(TP* $+$ *FP)*. The precision and recall emphasize on the model's performance regarding false positive (FP) and false negative (FN), respectively.

Authentication Evaluation Metrics. Unlike password-based authentication systems where the user authentication claim could pass or fail, the biometric authentication approaches are subject to authentication errors that can be evaluated using False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER). FAR is the rate of accepting an imposter biometric samples as a legitimate user and calculated as $(FP) \div (FP + TN)$. FRR is the rate of mistakenly rejecting a legitimate user as if the user is an imposter and calculated as $(FN) \div (FN + TP)$. Equal error rate (EER) is the rate where the FAR is equal to FRR. In our experiment, we calculate FAR, FRR, and EER. The FAR indicates the likelihood that the proposed authentication approach authorizes an imposter as a legitimate user; FRR indicates the likelihood that the proposed authentication approach incorrectly unauthorizes a legitimate user as an impostor. Since the authentication problem here is to distinguish between impostors and legitimate users, we also provide the ROC curve as a preferable method to visualize the relation between the true positive rate (TPR) against the false positive rate (FPR) and to illustrate the classifier performance.

E. AUToSen's Operations

AUTo*Sen* follows the general operational design for authentication systems by including the two phases, namely, the *enrollment* and *continuous authentication* phase.

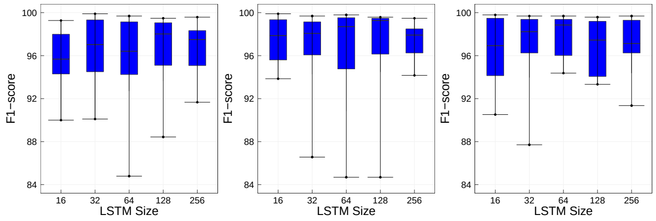
User enrollment. The enrollment phase incorporates: ❶ the data collection of users' sensory readings from five sensors, ❷ data cleaning and preprocessing, and ❸ authentication model training and evaluation. Due to the computational requirements of the enrollment phase, the enrolment is performed on the *authentication server*, responsible for training and periodically updating users' models. Once the model is trained and selected by a cross-validation process, it is secured to authenticate the assigned user. The trained model is considered fully-operational as long as it maintains a high authentication accuracy. Considering possible behavioral changes, a *model retraining process* is considered if the user reports several false alarms (the number of false alarms can be set by a design choice). Another approach to consider retraining the model is by observing the model confidence score for authenticating the legitimate user [23]. The degradation of the model confidence score (from the average score of different activities) indicates behavioral changes that require retraining the model.

User Continuous Authentication. There are two possible approaches for AUTo*Sen* to operate on smartphones: using a local authentication module or using a client/server design. Using the client/server design, users can access the authentication service through the server, where the authentication models are deployed. The data records are collected, cleaned, preprocessed, and sent to the cloud, where the authentication model responds with the authentication decision. Note that the client/server system design is a framework for experimental settings to examine the validity of our approach in adopting high-frequency continuous authentication using deep learning. Even with the requirement of network data communication between the user's smartphone and the authentication server, data transfer was minimal and sufficiently convenient for a real-time scenario. However, adopting such a design requires delivering alternatives, *e.g.,* explicit authentication, in case the service access is interrupted due to malicious activity or connection and systems failure. Using a local authentication module that incorporates a trained and continuously updated authentication model is a preferred choice. The expansion of storage and computational resources in current smartphones and the rapid development of machine learning tools allow the authentication modules to run locally on the device without requiring a connection to the server. Recently, *TensorFlow*, an open-source platform for machine learning, has launched *TensorFlow lite* that enables trained deep learning models to be deployed and run on smartphones for inference. In this work, we provide an analysis of different sensory data to train several deep learning architectures for user authentication on smartphones. We assume a client/server framework for our experiments and leave the deployment of authentication models on the user's smartphone for future work.

IV. Experiments and Evaluation

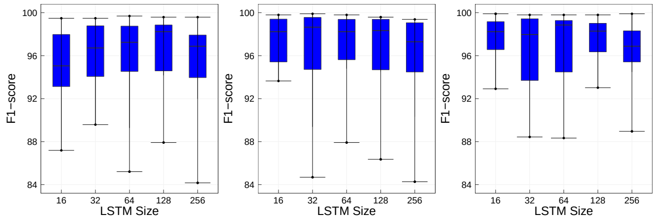
A. The Effects of Sensors Data

In this experiment, we investigate the effects of including different sensory data on the performance of authentication models. The experiment includes the evaluation of the implicit continuous user authentication task using various LSTM-based architecture models, namely, simple LSTM-based model architecture of Figure 2(a), bidirectional LSTM-based model architecture of Figure 2(b), and multi-layers LSTM-based model architecture of Figure 2(c). All models are trained with sensory data sampled with one second with 64Hz sampling rate *i.e.,* sequences of 64 readings per second from different sensors. The results show the accuracy metrics for the authentication models of all users included in the experiment.



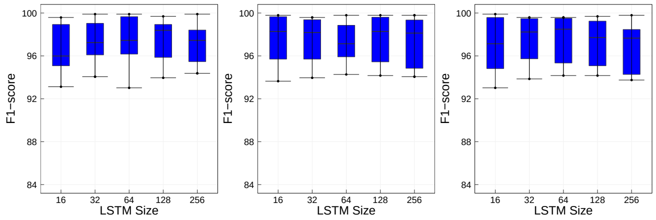
(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

Fig. 3. The accuracy of different LSTM model architectures when we feed the authentication model with five sensors (ToAcGrMaEl) data sequences collected within a second sampling period.



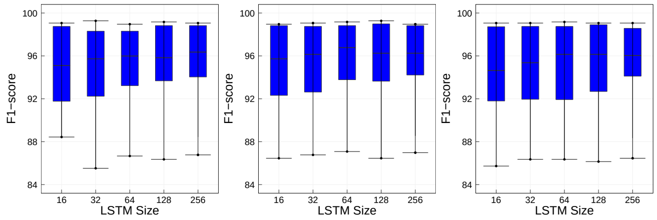
(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

Fig. 4. The accuracy of different LSTM model architectures when we feed the authentication model with four sensors (AcGrMaEl) data sequences collected within a second sampling period.



(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

Fig. 5. The accuracy of different LSTM model architectures when we feed the authentication model with three sensors (AcGrMa) data sequences collected within a second sampling period.



(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

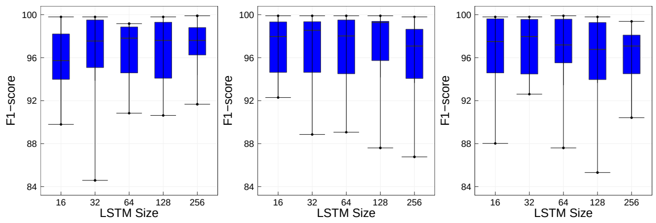
Fig. 6. The accuracy of different LSTM model architectures when we feed the authentication model with two sensors (AcGr) data sequences collected within a second sampling period.

Five Sensors (ToAcGrMaEl). Using the dataset of five sensors (ToAcGrMaEl), Figure 3 shows the results obtained by adopting different models architectures with different LSTM sizes. All model architectures show high F1-scores, with considerable results improvement when adopting a multi-layer LSTM model. Using simple LSTM, Figure 3(a) shows an average F1-score of 95.58% with the simplest model of simple LSTM with 16 units. The F1-score increased to reach 96.59% when increasing the model size to 128 units. Figure 3(b) shows the F1-scores of adopting bidirectional LSTM with different sizes, ranging from 16 units with an average F1-score of 97.34% to 256 units with average F1-score of 97.23%. Figure 3(c) shows the improvement of the authentication performance of multi-layer LSTM models over other architectures since it achieves an average F1-score of 96.44% with 16-units models and 97.84% with 64-units models. These results demonstrate the capabilities of LSTM models in capturing behavioral patterns of users' usage of smartphones.

Four Sensors (AcGrMaEl). The success of sensory data modeling with five sensors encourages the investigation of using data of fewer sensors. Since the touch actions require calculating the frequencies of actions, we explored working with other sensors, namely, accelerometer, gyroscope, magnetometer, and elevation. Figure 4 shows the results achieved by adopting different model architectures. When using simple LSTM, Figure 4(a) shows that a model with 16 units achieves an average F1-score of 96.18%, and 97.01% when using 128 units. Adopting bidirectional LSTM architecture improves the results to achieve an average F1-score of 97.39% with 16-units models and an average F1-score of 96.62% for the 128-units models. Similar results are obtained when adopting multi-layer LSTM models. Figure 4(c) shows that multi-layer LSTM models with 16-units achieve an average F1-score of 95.95%. Even when removing the feature of touch actions, LSTM models successfully capture the behavioral patterns that enable high authentication accuracy.

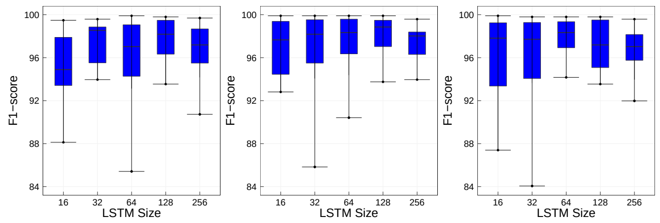
Three Sensors (AcGrMa). Since several previous works, such as [37], [17], and [16], have shown the success of modeling users' activities using readings of three sensors namely, accelerometer, gyroscope, and magnetometer, this experiment uses data collected from only three sensors (accelerometer, gyroscope, and magnetometer). Figure 5 shows the results of adopting different architectures with different model sizes. Generally, using readings from the three sensors (AcGrMa) has shown better performance of the authentication task across all users. One explanation for these improvements is that the touch actions and elevation readings are more sensitive and event-oriented than other sensors. Figure 5(a) shows the results obtained when using simple LSTM models with different model sizes. Using 16 units as a model size achieves an average F1-score of 96.59% and for some users as high as 99.58%. Similar results are obtained when using bidirectional LSTM models. Figure 5(b) shows the results of authentication models using bidirectional LSTM with an average F1-score of 97.49% when using 16 units, and an average of 97.57% when using 128 units. The results obtained from using multi-layer LSTM models are shown in Figure 5(c).

Two Sensors (AcGr). Using only two sensors (accelerometer and gyroscope), this experiment evaluates the performance of the authentication task. Figure 6 shows the results of using different architectures to model users' behavioral patterns for the purpose of authentication. The results show surprisingly high F1-score even when using readings of two sensors, with a slight degradation in comparison to results from other datasets. Figure 6(a) shows an average F1-score of 93.64% with the simplest LSTM model of 16 units. Bidirectional LSTM models achieve similar results shown in Figure 6(b), while the results obtained by multi-layer models are shown in Figure 6(c).



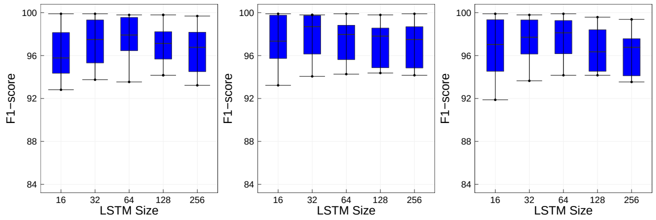
(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

Fig. 7. The accuracy of different LSTM model architectures when we feed the authentication model with five sensors (ToAcGrMaEl) data sequences collected within half a second sampling period.



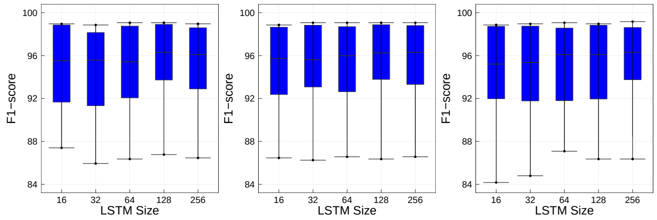
(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

Fig. 8. The accuracy of different LSTM model architectures when we feed the authentication model with four sensors (AcGrMaEl) data sequences collected within half a second sampling period.



(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

Fig. 9. The accuracy of different LSTM model architectures when we feed the authentication model with three sensors (AcGrMa) data sequences collected within half a second sampling period.



(a) Simple LSTM (b) Bidirectional LSTM (c) Multi-layer LSTM

Fig. 10. The accuracy of different LSTM model architectures when we feed the authentication model with two sensors (AcGr) data sequences collected within half a second sampling period.

B. The Effects of Data Sampling Period

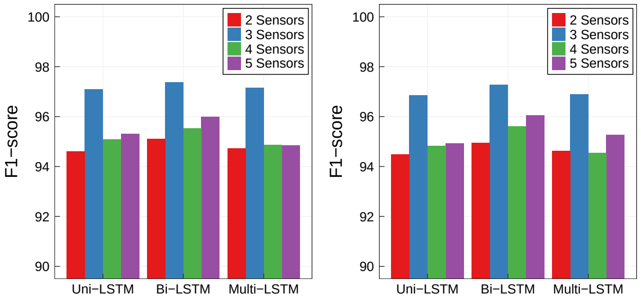
Continuous authentication requires a high user validation frequency. In the previous experiments, we show that LSTM-based models are capable of modeling users' behavior for authentication with high accuracy using sensory data collected with a one-second sampling period. In this experiment, we investigate the effects of using sensory data for a higher frequency authentication, such as 0.5 seconds. Considering the sensors' sampling rate at 64Hz, the size of readings within 0.5 seconds is 32 readings per sensor. Similar to the previous experiments, we investigate the effects of using higher authentication frequency, *i.e.,* shorter data sampling period, on the performance of different model architectures.

Five Sensors (ToAcGrMaEl). Starting with the Five-Sensor dataset, Figure 7 shows the F1-scores achieved by different LSTM model architectures. Figure 7(a) shows that simple LSTM model achieves 95.65% and 96.52% with 16 and 64 units, respectively. Using bidirectional LSTM models, Figure 7(b) shows the achieved F1-scores for the authentication task with an average of 97.03%, 96.72%, and 96.31% when using 16, 32, and 64 units, respectively. Similar results are obtained when using multi-layer LSTM models (see Figure 7(c)).

Four Sensors (AcGrMaEl). With Four-Sensor dataset, Figure 8 shows a slight improvement in the F1-scores in comparison with the results of the Five-Sensor dataset, especially when using bidirectional LSTM models. Simple LSTM models are still capable of modeling users' behavioral patterns with high F1-scores across different users, as shown in Figure 8(a). Using simple LSTM models achieve an average of 94.72% for models with 16 units, and 97.21% for models with 32 units. Figure 8(b) shows an improvement of F1-scores achieved by bidirectional LSTM with an average of 97.79% and 97.85% for 32-units and 128-units models, respectively. Similar results are reported in Figure 8(c) for multi-layer LSTM.

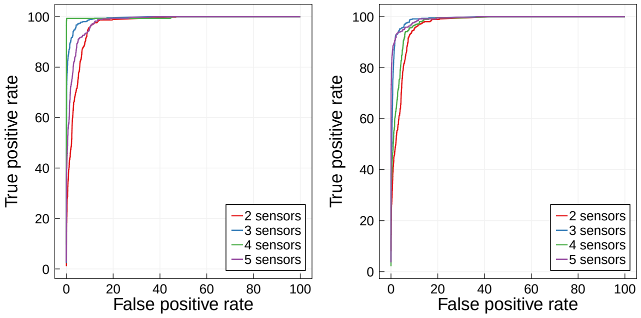
Three Sensors (AcGrMa). Figure 9 shows the results obtained by using the Three-Sensor dataset with readings collected within 0.5-second periods. Similar to results obtained by using the three sensors (accelerometer, gyroscope, and magnetometer) when sampling sensors' reading form one-second period, the results obtained using higher authentication frequency represent the best results among other datasets. Figure 9(a) shows an average F1-score of 96.18%, 97.14%, 97.51%, 97.01%, and 96.42% when using simple LSTM with 16, 32, 64, 128, and 256 units, respectively. When using bidirectional LSTM models, the F1-scores reported in Figure 9(b) show an average of 97.31%, 97.70%, 97.32%, 97.06%, and 97.01% when using simple LSTM with 16, 32, 64, 128, and 256 units, respectively. Similar results are achieved when using multi-layer architectures of the authentication models as shown in Figure 9(c) with the best average F1-score of 97.52% is reported for the 128-units multi-layer models.

Two Sensors (AcGr). Using only the accelerometer and gyroscope readings, Figure 10 shows a slight degradation in F1-scores in comparison to those achieved using the Three-Sensor dataset. However, the results show that using only readings from the accelerometer and the gyroscope are sufficient to model users' behavioral patterns even when collected with high frequency such as within a half-second period. Figure 10(a) shows the F1-scores achieved by using simple LSTM models with an average F1-scores ranging from 94% to 96% across different model sizes. Similar results are reported for other architectures in Figure 10(b) and Figure 10(c) for bidirectional and multi-layer LSTM architecture, respectively.



(a) Sensors data from one second. (b) Sensors data from half a second.

Fig. 11. The average F1-scores of different LSTM architecture when we feed the models with sequence readings of one second and a half-second period. The scores are achieved with different datasets (5 Sensors: ToAcGrMaEl, 4 Sensors: AcGrMaEl, 3 Sensors: AcGrMa, 2 Sensors: AcGr).



(a) Sensors data from one second. (b) Sensors data from half a second.

Fig. 12. ROC curves with multi-layers LSTM-based architecture using readings of various datasets and sampling periods. (5 Sensors: ToAcGrMaEl, 4 Sensors: AcGrMaEl, 3 Sensors: AcGrMa, 2 Sensors: AcGr).

C. Performance Summary and Discussion

Figure 11 shows a summary of all model performance under different configurations. Figure 11(a) shows the average F1-scores of different model architectures performing the authentication task using different sensory datasets collected with a one-second sampling period. While Figure 11(b) shows the average F1-scores of different model architectures when using sensor readings of half a second. Using only three sensors, *i.e.,* accelerometer, gyroscope, and magnetometer, is sufficient to capture users' behavioral patterns for authentication. The best results are obtained using the (Three-Sensor data: AcGrMa) in all configuration and settings. For example, Figure 11(a) shows the average F1-score across models of simple LSTM architecture with different sizes is 94.62%, 97.10%, 95.08%, and 95.31% for the Two-Sensor dataset: AcGr, Three-Sensor dataset: AcGrMa, Four-Sensor dataset: AcGrMaEl, and Five-Sensor dataset: ToAcGrMaEl, respectively. The difference in F1-scores when considering fewer sensors is (97.10 - 94.62 = 2.48%) with only two sensors and one-second readings.

We observe that using four or five sensors does not provide better performance than using three sensors (*i.e.,* accelerometer, gyroscope, and magnetometer). Figure 11(a) shows a difference in the average F1-scores of (97.10 - 95.08 = 2.02%) and (97.10 - 95.31 = 1.79%) when using the simple LSTM with datasets of four and five sensors, respectively, against the performance of using data of three sensors in AcGrMa dataset. This can be due to several factors, such as using the elevation readings and touch frequency data makes the input more sensitive to circumstantial changes in users' behavioral patterns when performing different tasks on their smartphones, *e.g.,* changing from Internet surfing to texting within the sampling period. Using readings from accelerometer, gyroscope, and magnetometer are robust for capturing distinctive behavioral patterns in a short sampling period, especially when there are no constraints on the user activity, while the additional sensors do not add any significant discriminative features to the authentication modality. Another observation is that the sampling period does not greatly affect the performance of the models as the results achieved using sensors' data of half a second is comparable to those of one second, indicating the validity of our approach for active user authentication with frequency as low as half a second.

TABLE II
THE AVERAGE PERCENTAGE OF FAR, FRR, AND EER OF AUTO*Sen* PERFORMING THE AUTHENTICATION TASK FOR ALL USERS IN THE COLLECTED DATASET USING MULTI-LSTM MODEL.

Dataset	Sampling frequency	FAR	FRR	EER
Two-Sensor (AcGr)	1s	1.83	43.83	5.02
Three-Sensor (AcGrMa)	1s	**0.95**	**6.67**	**0.41**
Four-Sensor (AcGrMaEl)	1s	2.31	7.62	0.72
Five-Sensor (ToAcGrMaEl)	1s	1.72	8.47	0.37
Two-Sensor (AcGr)	0.5s	1.97	42.24	3.85
Three-Sensor (AcGrMa)	0.5s	**0.96**	**8.08**	**0.09**
Four-Sensor (AcGrMaEl)	0.5s	1.65	9.87	0.36
Five-Sensor (toAcGrMaEl)	0.5s	1.53	9.16	0.35

Among different LSTM architectures, bidirectional LSTM achieved the best average results in almost every setting. Simple Uni-LSTM model achieves similar results to that achieved by bidirectional LSTM. Multi-LSTM has shown a slight degradation in F1-scores due to the fixed time given to training different models since multi-LSTM architectures include a larger number of parameters to be optimized during the training process than the one of simple or bidirectional LSTM, and thus multi-layers models require more training time. However, the experiments have shown that simple LSTM and bidirectional LSTM are sufficient to model users' behavioral patterns for the authentication task.

D. Authentication Analysis

The previous experiments evaluate the performance of AUTo*Sen* using different settings. The results show that using bidirectional and multi-layer LSTM models achieve remarkable authentication results with different model sizes. In this experiment, we aim to evaluate the authentication models in terms of FAR, FRR, and EER. We choose a multi-layer LSTM model with 64 units to be the baseline architecture model as the best trade-off between performance and efficiency. Each user authentication model is trained, in the same manner as in the previous experiments, to perform the authentication task using sensory data collected from the model's legitimate user and other ten randomly selected users as impostors' signals. After the training process, the authentication models are evaluated using sensory readings from the legitimate users to calculate the FRR, and sensory readings from the other ten random users as impostor signals to calculate the FAR. Table II shows the average results of the authentication performance of AUTo*Sen* using three evaluation metrics, FAR, FRR, and EER across all users. Moreover, Figure 12 shows the ROC curves for the true positive rate and false positive rate across different datasets. The best average FAR is reported when using the Three-Sensor dataset (AcGrMa) with the one-second sensory data sampling period, which is a FAR of 0.95%. Achieving this FAR score shows the resilience of our approach against authenticating impostors, which is the lowest among other works in the literature. Using a higher authentication frequency of 0.5 seconds, AUTo*Sen* achieves an average FAR of 0.96% which is only 0.01% different from authenticating with one-second frequency. This result shows that AUTo*Sen* is resilient to adversaries even when users' behavioral patterns are modeled with 0.5 seconds readings. For FRR, AUTo*Sen*achieves the best results using the Three-Sensor data (AcGrMa) with an average of 6.67% when the sensor data is collected with a sampling period of one second, and an average of 8.08% when the sampling period is 0.5 seconds. The best EER score is also achieved using the Three-Sensor dataset (AcGrMa) with an average of 0.41% and 0.09% when using the sampling period of one and 0.5 seconds, respectively. The authentication performance of AUTo*Sen* shows that three sensors, *i.e.,* accelerometer, gyroscope, and magnetometer, are sufficient to model users' behavior for the authentication tasks. Including the elevation readings and the touch actions also helps the modeling process; however, such readings may introduce more sensitive inputs that influence the authentication final decision. Moreover, using a sampling period of 0.5 seconds enables sufficient information for capturing users' behavior for the authentication task since AUTo*Sen* achieves the best EER score of 0.09% using readings from three sensors collected in 0.5 seconds. However, the FAR and FRR scores in Table II show that the model performance when using readings of one second long is relatively better than when using readings of half a second. The explanation of this result could be related to the longer sequences fed to the model at the training phase that makes the model generates more stable and accurate user profiles.

Considering the FRR score, AUTo*Sen* is expected to deliver convenient usability for authenticating legitimate users by increasing the authentication frequency. Using AUTo*Sen* with a one-second sampling period, users are expected to be authenticated with a probability of 99.56% and 99.97% within two and three seconds, respectively. When using the AUTo*Sen* with half a second sampling period, the usability increases and the legitimate users are expected to be authenticated with a high probability such as 99.99% within two seconds. Moreover, considering the FAR results achieved by the two examined sampling period, the half a second sampling period shows outstanding usability and performance.

E. Temporal Behavioral Changes and Model Retraining

Behavioral Changes of Legitimate User. Temporal behavioral changes of the legitimate user may negatively influence the authentication model performance by increasing the false rejection rate. Such changes are reported by previous work [23]. One way to handle temporal behavioral changes is by updating the model periodically or/and when behavioral changes are observed through the decline of the model confidence in authenticating the legitimate user. Observing the temporal behavioral changes of users, we show that the authentication models of AUTo*Sen* are robust to temporal changes and can operate with high authentication accuracy if they updated periodically every three days. Figure 13 shows the confidence score, *i.e.,* the positive authentication probability, of authenticating a user in high frequency, *e.g.,* every one second, during five separate minutes from three consecutive days. Similar patterns of behavioral changes were observed in different users' data. However, we selected a random user's readings (from three sensors) during five random minutes for three consecutive days to demonstrate the existence of behavioral changes and to show that a retraining process might be needed after three days. Setting a confidence threshold of 0.2, the AUTo*Sen*' model only authenticated the legitimate user three times during the third day. This demonstrates the robustness of AUTo*Sen* to temporal changes and its smooth user experience when the models are periodically updated.

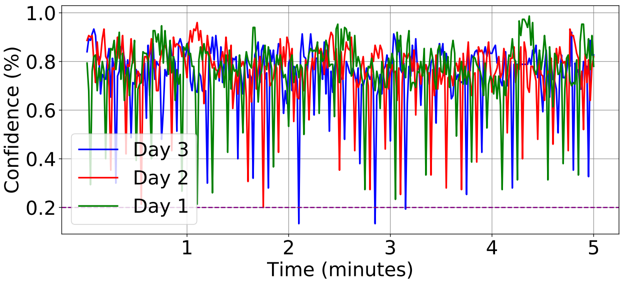


Fig. 13. Temporal changes observed by AUTo*Sen*'s confidence score.

Behavioral Changes of Other Users. Proper system design should consider scenarios where the device is shared among multiple users. Addressing such scenarios can be by enabling ❶ application-specific continuous authentication, where the continuous authentication is enabled for specific applications, ❷ multi-user continuous authentication, where each user is enrolled and the authentication module should detect the user and grant access privileges based on specified policies, or ❸ simple straightforward switching on and off the module when sharing the device with trusted others.

F. Robustness Assessments for Adversarial Settings

Designing authentication methods requires establishing both security and usability assessment. Further, using machine learning-based techniques for the authentication module requires addressing aspects related to adversarial machine learning, where adversaries may launch attacks by manipulating the input for the system to force the model to generate the wrong output, *i.e.,* leading the model to misclassification [32], [2]. Such manipulations require minimal changes to the input data so the resulting adversarial samples are very similar to the original, therefore posing a serious threat. Further, the authentication system design, *e.g.,* local module or client/server design, requires analyzing different levels of adversary's knowledge and capabilities to launch either white-box or black-box attacks. We leave addressing the robustness of models against such adversarial settings for future work.

V. Conclusion

Smartphone becomes crucial for our daily life activities and increasingly loaded with our personal information to perform several sensitive tasks including mobile banking, communication, and storing private photos and files. Therefore, there is a high demand for applying secure and usable authentication technique that prevents unauthorized access to sensitive information. This research proposes AUTo*Sen*, an active authentication approach for smartphone users using sensors data. Our approach exploits LSTM to model users' behavioral patterns using readings of smartphones' sensors. AUTo*Sen* is simple and efficient in handling and processing sensors' data in a real-time manner that enables validating users without the requirement to interact with their phones. We conducted comprehensive experiments to evaluate AUTo*Sen* on a real dataset collected with our data-collection application from 84 participants. We show the AUTo*Sen* is capable of authenticating users with high F1-score using readings from different sensors. However, using the three sensors namely, accelerometer, gyroscope, and magnetometer, shows the best impact of the authentication performance. Moreover, we show that AUTo*Sen* is capable of processing and modeling users' behavior in real-time and with high frequency as 0.5 seconds since the authentication performance indicates the sufficiency of sensors readings within 0.5 seconds to model users' behavior for authentication purposes. AUTo*Sen* shows new state-of-the-art results for active authentication for smartphones using sensors data in terms of FAR, FRR, and EER.

References

[1] M. Abuhamad, T. AbuHmed, A. Mohaisen, and D. Nyang, "Large-scale and language-oblivious code authorship identification," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2018, pp. 101–114.

[2] A. Abusnaina, H. Alasmary, M. Abuhamad, S. Salem, D. Nyang, and A. Mohaisen, "Subgraph-based adversarial examples against graph-based iot malware detection systems," in *Computational Data and Social Networks*. Springer International Publishing, 2019, pp. 268–281.

[3] A. Alzubaidi and J. Kalita, "Authentication of smartphone users using behavioral biometrics," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 1998–2026, 2016. [Online]. Available: https://doi.org/10.1109/COMST.2016.2537748

[4] S. Amini, V. Noroozi, A. Pande, S. Gupte, P. S. Yu, and C. Kanich, "Deepauth: A framework for continuous user re-authentication in mobile apps," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018, pp. 2027–2035.

[5] J. S. Arteaga-Falconi, H. A. Osman, and A. El-Saddik, "ECG authentication for mobile devices," *IEEE Trans. Instrumentation and Measurement*, vol. 65, no. 3, pp. 591–600, 2016. [Online]. Available: https://doi.org/10.1109/TIM.2015.2503863

[6] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens," in *Proceedings of the 4th USENIX Conference on Offensive Technologies*, ser. WOOT'10. USENIX Association, 2010, pp. 1–7.

[7] G. Biegel and V. Cahill, "A framework for developing mobile, context-aware applications," in *Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the*. IEEE, 2004, pp. 361–365.

[8] A. Buriro, B. Crispo, S. Gupta, and F. Del Frari, "Dialerauth: A motion-assisted touch-based smartphone user authentication scheme," in *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*. ACM, 2018, pp. 267–276.

[9] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *CoRR*, vol. abs/1409.1259, 2014.

[10] S. Choi, D. J. Kim, Y. Y. Choi, K. Park, S.-W. Kim, S. H. Woo, and J. J. Kim, "A multisensor mobile interface for industrial environment and healthcare monitoring," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2344–2352, 2017.

[11] N. L. Clarke and A. Mekala, "The application of signature recognition to transparent handwriting verification for mobile devices," *Information management & computer security*, vol. 15, no. 3, pp. 214–225, 2007.

[12] H. Crawford, K. Renaud, and T. Storer, "A framework for continuous, transparent mobile device authentication," *Computers & Security*, vol. 39, pp. 127–136, 2013.

[13] R. Damaševičius, R. Maskeliūnas, A. Venčkauskas, and M. Woźniak, "Smartphone user identity verification using gait characteristics," *symmetry*, vol. 8, no. 10, p. 100, 2016.

[14] S. Das, D. Guha, and B. Dutta, "Medical diagnosis with the aid of using fuzzy logic and intuitionistic fuzzy logic," *Applied Intelligence*, vol. 45, no. 3, pp. 850–867, 2016.

[15] B. Draffin, J. Zhu, and J. Y. Zhang, "Keysens: Passive user authentication through micro-behavior modeling of soft keyboard interaction," in *Mobile Computing, Applications, and Services - 5th International Conference, MobiCASE 2013, Paris, France, November 7-8, 2013, Revised Selected Papers*, 2013, pp. 184–201. [Online]. Available: https://doi.org/10.1007/978-3-319-05452-0_14

[16] M. Ehatisham-ul Haq, M. Awais Azam, U. Naeem, Y. Amin, and J. Loo, "Continuous authentication of smartphone users based on activity pattern recognition using passive mobile sensing," *J. Netw. Comput. Appl.*, vol. 109, no. C, pp. 24–35, May 2018. [Online]. Available: https://doi.org/10.1016/j.jnca.2018.02.020

[17] G. Fenu and M. Marras, "Controlling user access to cloud-connected mobile applications by means of biometrics," *IEEE Cloud Computing*, vol. 5, no. 4, pp. 47–57, 2018. [Online]. Available: https://doi.org/10.1109/MCC.2018.043221014

[18] P. Fernandez-Lopez, J. Liu-Jimenez, C. Sanchez-Redondo, and R. Sanchez-Reillo, "Gait recognition using smartphone," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2016, pp. 1–7.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[20] G. Kambourakis, D. Damopoulos, D. Papamartzivanos, and E. Pavlidakis, "Introducing touchstroke: keystroke-based authentication system for smartphones," *Security and Communication Networks*, vol. 9, no. 6, pp. 542–554, 2016.

[21] H. G. Kayacik, M. Just, L. Baillie, D. Aspinall, and N. Micallef, "Data driven authentication: On the effectiveness of user behaviour modelling with mobile device sensors," *CoRR*, vol. abs/1410.7743, 2014. [Online]. Available: http://arxiv.org/abs/1410.7743

[22] W. Lee and R. B. Lee, "Multi-sensor authentication to improve smartphone security," in *ICISSP 2015 - Proceedings of the 1st International Conference on Information Systems Security and Privacy, ESEO, Angers, Loire Valley, France, 9-11 February, 2015.*, 2015, pp. 270–280. [Online]. Available: https://doi.org/10.5220/0005239802700280

[23] W.-H. Lee and R. B. Lee, "Implicit smartphone user authentication with sensors and contextual machine learning," in *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2017, pp. 297–308.

[24] G. Li and P. Bours, "Studying wifi and accelerometer data based authentication method on mobile phones," in *Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications*. ACM, 2018, pp. 18–23.

[25] Y. Li, H. Hu, and G. Zhou, "Using data augmentation in continuous authentication on smartphones," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 628–640, Feb 2019.

[26] H. Lu, A. B. Brush, B. Priyantha, A. K. Karlson, and J. Liu, "Speakersense: Energy efficient unobtrusive speaker identification on mobile phones," in *International conference on pervasive computing*. Springer, 2011, pp. 188–205.

[27] M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia, "Towards mobile authentication using dynamic signature verification: useful features and performance evaluation," in *2008 19th International Conference on Pattern Recognition*. IEEE, 2008, pp. 1–5.

[28] S. Mondal and P. Bours, "Person identification by keystroke dynamics using pairwise user coupling," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1319–1329, June 2017.

[29] A. Mosenia, S. Sur-Kolay, A. Raghunathan, and N. K. Jha, "CABA: continuous authentication based on bioaura," *IEEE Trans. Computers*, vol. 66, no. 5, pp. 759–772, 2017. [Online]. Available: https://doi.org/10.1109/TC.2016.2622262

[30] T. J. Neal and D. L. Woodard, "Surveying biometric authentication for mobile device security," *Journal of Pattern Recognition Research*, vol. 1, pp. 74–110, 2016.

[31] C. Nickel, H. Brandt, and C. Busch, "Classification of acceleration data for biometric gait recognition on mobile devices," *BIOSIG 2011–Proceedings of the Biometrics Special Interest Group*, 2011.

[32] N. Papernot, P. D. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *Proceedings of the IEEE European Symposium on Security and Privacy*, 2016, pp. 372–387.

[33] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[34] F. Schaub, R. Deyhle, and M. Weber, "Password entry usability and shoulder surfing susceptibility on different smartphone platforms," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '12, 2012, pp. 13:1–13:10. [Online]. Available: http://doi.acm.org/10.1145/2406367.2406384

[35] C. Shen, Y. Li, Y. Chen, X. Guan, and R. A. Maxion, "Performance analysis of multi-motion sensor behavior for active smartphone authentication," *IEEE Trans. Information Forensics and Security*, vol. 13, no. 1, pp. 48–62, 2018. [Online]. Available: https://doi.org/10.1109/TIFS.2017.2737969

[36] C. Shen, T. Yu, S. Yuan, Y. Li, and X. Guan, "Performance analysis of motion-sensor behavior for user authentication on smartphones," *Sensors*, vol. 16, no. 3, p. 345, 2016.

[37] Z. Sitova, J. Sedenka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani, "HMOG: new behavioral biometric features for continuous authentication of smartphone users," *IEEE Trans. Information Forensics and Security*, vol. 11, no. 5, pp. 877–892, 2016. [Online]. Available: https://doi.org/10.1109/TIFS.2015.2506542

[38] C. Song, A. Wang, K. Ren, and W. Xu, "Eyeveri: A secure and usable approach for smartphone user authentication," in *35th Annual IEEE International Conference on Computer Communications, INFOCOM*, 2016, pp. 1–9. [Online]. Available: https://doi.org/10.1109/INFOCOM.2016.7524367

[39] R. Spolaor, Q. Li, M. Monaro, M. Conti, L. Gamberini, and G. Sartori, "Biometric authentication methods on smartphones: A survey." *Psychnology Journal*, vol. 14, no. 2, 2016.

[40] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[41] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, 2012.

[42] J. Wu and R. Jafari, "Orientation independent activity/gesture recognition using wearable motion sensors," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1427–1437, April 2019.

[43] J. Wu and Z. Chen, "An implicit identity authentication system considering changes of gesture based on keystroke behaviors," *IJDSN*, vol. 11, pp. 470 274:1–470 274:16, 2015. [Online]. Available: https://doi.org/10.1155/2015/470274

[44] Z. Yu, E. Xu, H. Du, B. Guo, and L. Yao, "Inferring user profile attributes from multidimensional mobile phone sensory data," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5152–5162, June 2019.

[45] Y. Zhang, P. J. Thorburn, W. Xiang, and P. Fitch, "Ssim—a deep learning approach for recovering missing time series sensor data," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6618–6628, Aug 2019.

[46] Y. Zhang, W. Hu, W. Xu, C. T. Chou, and J. Hu, "Continuous authentication using eye movement response of implicit visual stimuli," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 177:1–177:22, Jan. 2018. [Online]. Available: http://doi.acm.org/10.1145/3161410

[47] H. Zhu, J. Hu, S. Chang, and L. Lu, "Shakein: Secure user authentication of smartphones with single-handed shakes," *IEEE transactions on mobile computing*, vol. 16, no. 10, pp. 2901–2912, 2017.

[48] J. Zhu, P. Wu, X. Wang, and J. Zhang, "Sensec: Mobile security through passive sensing," in *International Conference on Computing, Networking and Communications, ICNC 2013, San Diego, CA, USA, January 28-31, 2013*, 2013, pp. 1128–1133. [Online]. Available: https://doi.org/10.1109/ICCNC.2013.6504251

[49] T. Zhu, Z. Qu, H. Xu, J. Zhang, Z. Shao, Y. Chen, S. Prabhakar, and J. Yang, "Riskcog: Unobtrusive real-time user authentication on mobile devices in the wild," *IEEE Transactions on Mobile Computing*, 2019.

[50] V. Zimmermann and N. Gerber, ""if it wasn't secure, they would not use it in the movies"–security perceptions and user acceptance of authentication technologies," in *International Conference on Human Aspects of Information Security, Privacy, and Trust*. Springer, 2017, pp. 265–283.