Combating False Data Injection Attacks on Human-Centric Sensing Applications

JINGYU XIN, Syracuse University, USA

VIR V. PHOHA, Syracuse University, USA

ASIF SALEKIN, Syracuse University, USA

The recent prevalence of machine learning-based techniques and smart device embedded sensors has enabled widespread human-centric sensing applications. However, these applications are vulnerable to false data injection attacks (FDIA) that alter a portion of the victim's sensory signal with forged data comprising a targeted trait. Such a mixture of forged and valid signals successfully deceives the continuous authentication system (CAS) to accept it as an authentic signal. Simultaneously, introducing a targeted trait in the signal misleads human-centric applications to generate specific targeted inference; that may cause adverse outcomes. This paper evaluates the FDIA's deception efficacy on sensor-based authentication and human-centric sensing applications simultaneously using two modalities - accelerometer, blood volume pulse signals. We identify variations of the FDIA such as different forged signal ratios, smoothed and non-smoothed attack samples. Notably, we present a novel attack detection framework named Siamese-MIL that leverages the Siamese neural networks' generalizable discriminative capability and multiple instance learning paradigms through a unique sensor data representation. Our exhaustive evaluation demonstrates Siamese-MIL's real-time execution capability and high efficacy in different attack variations, sensors, and applications.

CCS Concepts: • **Security and privacy → Intrusion detection systems**; • **Human-centered computing** → Ubiquitous and mobile computing design and evaluation methods.

Additional Key Words and Phrases: Injection Attack, False Data Injection Attack, Multiple Instance Learning, Siamese Network, Mobile, Wearable, Authentication, Sensor Attack, Defense, Deep Learning

1 INTRODUCTION

In recent years, machine learning-based techniques and Internet of Things sensors are advancing rapidly, enabling various human-centric sensing applications on smart devices, such as motion sensor-based authentication [13, 14, 19, 38], physiological sensing-based health monitoring [36, 40, 47, 50, 51], human activity recognition [6, 8, 21, 37], etc. Due to the invasive nature of these applications, privacy breaches may lead to severe adverse effects. Recent studies have focused on attacks targeting to deceive the authentication systems [39, 41, 58], however, no study has addressed an attack on smart device sensory streams that can deceive the continuous authentication and manipulate human-centric systems simultaneously to generate adverse effects.

This paper evaluates such an attack as the **false data injection attack (FDIA)**, its deception efficacy on sensor-based authentication and human-centric sensing applications simultaneously, and presents a novel attack detection approach. By injecting forged data into the sensory data stream, the inference of the human-centric applications can be altered. Furthermore, no knowledge of the victim or victim's signal sample is needed to launch this attack. Take a health monitoring system as an example that utilizes smartwatch physiological sensory data to assess a patient's health. FDIA attackers can inject other unhealthy individuals' physiological signals into the genuine healthy user's signals to cause a wrong assessment that may lead to unnecessary interventions and even harm the patient's health. Using such targeting misinformation generation through the FDIA, attackers can control the outputs of the sensing applications as they desire. Some practical FDIA on smart-device scenarios are further discussed in the threat model Section 2.2.

Considering the significant adverse outcomes generation capability of the FDIA, this paper evaluates and demonstrates the deception efficacy of FDIA on the sensor-based authentication and human-centric sensing applications. The novelty of this paper comes from formulating the FDIA detection problem as a multiple instance learning (MIL) problem and developing a framework named Siamese-MIL that combines multi-head Siamese networks with MIL paradigms to detect FDIA accurately. Furthermore, we performed a comprehensive evaluation that shows the high efficacy of Siamese-MIL in FDIA detection.

Paper Outline: Section 2 discusses the FDIA threat model. Paper contributions are summarized in Section 3. Section 4 discusses the datasets and data processing. Section 5 discusses the FDIA's efficacy in deceiving smart device authentication systems and human-centric sensing applications. We present the Siamese-MIL, the FDIA attack detection approach in Section 6. Section 7 discusses the exhaustive evaluation of Siamese-MIL on multiple datasets, modalities, and applications. Section 8 discusses how the variations of FDIA affect the Siamese-MIL's performance. Section 9 evaluates the Siamese-MIL's real-time execution capability on resource constraint smart devices. Section 10 summarizes the related works and the security concerns of smart device sensors. Lastly, we discuss the study observations, insights, and limitations in Section 11 and give conclusions in Section 12.

2 THREAT MODEL AND ANATOMY OF THE ATTACK

2.1 Threat Model

In this threat model, the attacker injects targeting information into the sensory data stream to mislead the human-centric sensing system. False data injection attack (FDIA) [57] modifies the sensory data, such that the sensory data stream of another user is injected in the target user's sensory stream (victim), so no knowledge of the victim is assumed. If the attacker wants to generate a specific output from the sensing application, they can use signals with targeting information to influence the application's outcome. This paper particularly focuses on such targeting FDIAs aiming to misinform human-centric sensing applications with certain forged information.

The attack can be launched by manipulating the sensor signal using malware infection (e.g., SMASheD framework [33], Spy-sense [16]); or compromising the communication channel (e.g., using BlueDoor [62], Fit Bite, or GarMax [44]). Such a threat model is practical and has been commonly employed in studies about the privacy threat of smart device sensors [29, 30, 61]. The prevention of such manipulations of sensors and communication is out of the scope of the paper. Rather, following the state-of-the-art attack detection studies [26, 32, 43], this paper focuses on the FDIA's deception efficacy on continuous authentication and human-centric sensing system, and our presented attack detection approach's performance.

2.2 Attack Scenarios

This section gives some FDIA attack scenarios and their impact on the human-centric sensing systems.

Scenario 1: Consider the victim as a diabetes patient, and the caregiver monitors his daily exercise through a smartphone activity detection system. However, the victim is tricked into installing some malicious application

on the smartphone that can manipulate the accelerometer sensory streams. E.g., SMASheD [33] can enable a malicious app with only the INTERNET permission to manipulate motion sensors on unrooted Android devices. Installed malicious application can mislead the activity detection system to infer incorrect activities. E.g., the attacker can inject other people's jogging data portions into the user's accelerometer sensory stream to mislead the activity detection system that the victim is jogging. Such attack will inform misleading exercise measurements to the caregiver, which may lead to wrong follow-up interventions. If diabetes is not treated properly, it may further harm the patient's heart, eyes, kidneys, etc. [5, 34], and increase the risk of death [48].

Scenario 2: A post-traumatic stress disorder (PTSD) patient wears a smart band/watch to monitor their stress levels. Though simple analysis such as step counting is performed on the wearable device itself, complicated applications like stress detection need the wearable to send raw sensing data to another device to get the data processed. E.g., Empatica E4, Embrace, HeartGuide, MobileHelp Smart are smart bands that connect to smartphones and transmit the collected physiological data through Bluetooth for advanced processing. During the data transmission, the FDIA attacker may access the Bluetooth communication and modifies the data packets. A tool like BlueDoor [62] can break the confidentiality of Bluetooth and alter the communicated information. The attacker can mix the victim's physiological data with other individual's data under stress to cause an illusion that the victim is under long-term stress. Such an attack may result in a misdiagnosis which can bring inappropriate or even harmful treatment to the patient. It may even result in worsened suicidal or homicidal ideation [59]. It also would lead to an additional cost (e.g., per person/year PTSD treatment cost is $16,750 in USA (2016) [63]).

2.3 Assumptions

Each smart device has a unique owner in the threat model. Smart device embedded sensors are generating continuous data streams. We assume that the attacker has a large amount of sensory data of various traits (e.g., motion sensory data of different activities) from other individuals. These sensory data segments will be utilized as the signal portions injected on the target (victim) user's sensory streams to generate the FDIA samples.

Recent studies [28, 35, 52, 54] have evaluated continuous authentication systems (CAS) that verify the authenticity of the sensory data streams. CASs identify the characteristics of the sensory data indicative of the respective user's identity and repeatedly examine the authenticity of the continuous sensory streams. Our threat model considers a harder situation where a CAS is working in the background. Only the signal verified by the CAS can be further accepted by other sensing applications. Hence, a successful FDIA sample needs to deceive the CAS into thinking it is an authentic signal and the human-centric application to generate a misled inference. Additionally, FDIA considers that the attackers have no information about the victim, background sensing or CAS approaches.

2.4 Attack Sample Generation

A challenge for FDIA samples generation is that, the replacement of the whole sensory data stream with signals containing misinformation from others can maximize the misleading effectiveness, but the CAS will easily reject such signal due of its high inconsistency with the legitimate user's data; thus, the attack signal cannot even reach the targeting sensing application. Therefore, we generate the attack sample as a mixture of legit and forged data, containing both legitimate patterns and misinformation. Such a mixture makes it harder to be detected by the CAS and can still mislead the sensing applications.

This paper evaluates the attack on n-length signals under the continuous sensing settings. For example, the original sensor reading $X = \{x_1, x_2, ..., x_n\}$ is n-length sequence. The attack can use k ($k \geq 1$) forged data segments from other individuals to replace k legitimate data portions. Consider $k = 1$, and a m-length ($m < n$) forged data sequence $F = \{f_1, f_2, ..., f_m\}$, is injected into the legitimate signal X. The generated attack sample will be $A = \{x_1, x_2, ..., x_i, f_1, f_2, ..., f_m, x_{i+m+1}, ..., x_n\}$ where a legitimate m-length signal portion is replaced by F at a random position (i.e., $i + 1$ to $i + m$). When $k > 1$, there can be multiple legitimate signal portions to be replaced

and the replacement signals can be from different individuals. This paper evaluates different characteristics of the attack samples that affect the FDIA's deception efficacy on authentication and sensing applications and the difficulty of detecting the attack samples. The characteristics are discussed below:

(1) *Forged signal ratio (FSR)*. Consider a n-length sensory signal, where in total t_s-length signal is replaced with other individuals' forged data. In this attack sample, the forged signal ratio (FSR) is $\frac{t_s}{n}$. Less injected signal portions in an attack sample generate a smaller FSR attack. Attack samples with smaller FSR have higher consistency with the legitimate data, making the attack detection more challenging. One trade-off for the attacker is, smaller FSR means a smaller portion of misinformation is injected into the signal. Hence, it can be less effective in misleading human-centric sensing applications. To ensure the robustness of the attack detection approach, we evaluate attack samples with FSRs from 10% to 90%. When FSR = 0%, no attack signal is injected, we consider it a pure sample; when FSR = 100%, the whole signal is replaced by others' data, it is considered a zero-effort attack [22] sample which is detected by the CAS easily (discussed in Section 5.1). Thus, we do not include evaluation for 0% and 100% FSR samples.

(2) *Effect of smoothing the boundaries between forged and legit signal*. When the forged signal from different individuals are injected, the transition between the forged and legit signals can be inconsistent, thus distinguishable. Smoothing the boundaries may remove such inconsistency and make the attack samples harder to be identified. Therefore, we evaluate both smoothed and non-smoothed attack samples.

Figure 1 shows an example of the attack signal generation to deceive an activity detection system. The legitimate user's accelerometer signal during walking activity is shown in the second row. This attack aims to misinform the activity detection system that the user is jogging. $k = 2$ forged jogging samples (top row) totaling t_s length are selected from different individuals. These samples randomly replace $k = 2$ portions (shown as red boxes on second-row figure) of the original walking accelerometer signal, thus generate a synthetic attack sample (third row of Figure 1). The signal transition on the boundary of the between the inserted signal portions is easily distinguishable (e.g., left boundary of red box). Hence, a smoothing operation is applied. The smoothed accelerometer attack sample is shown in the bottom row of Figure 1. Such a mixture of legitimate user's data with other individuals' data makes it difficult to be detected by the CASs and misinforms the background activity monitoring system as a jogging activity.
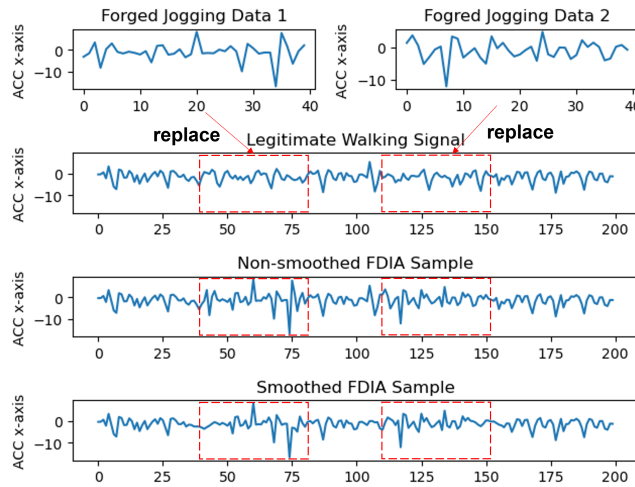

Fig. 1. An example of FDIA sample generation

3 CONTRIBUTIONS

This paper is the first to address the FDIA on deceiving and misleading smart device authentication and human-centric applications with attack samples. The paper's novelty comes from formulating the FDIA detection problem as a multiple instance learning (MIL) problem. The FDIA detection identifies if a signal sample (subject to inspection) comprises at least a pair of segments belonging to different individuals. This paper performs this task through a novel framework named Siamese-MIL that leverages the MIL paradigms, Siamese network structure, and a unique sensor data representation. Unlike supervised learning or voting mechanisms, the presented approach learns to identify any segment pair containing signals from different individuals without such data annotations during training. Moreover, the MIL training paradigm effectively avoids potential bias due to the disproportional ratio of legit and forged signal segments in the data. Siamese-MIL approach is discussed in detail in Section 6.2.

In particular, this paper addresses the following research questions:

(1) Are smart device continuous authentication systems (CASs) effective in detecting FDIA?
(2) Can FDIA deceive human-centric sensing applications?
(3) How does the FSR of FDIA samples and smoothing operation affect the FDIA's deception efficacy?
(4) What is the performance of the Siamese-MIL detection approach against smart device sensor FDIA?
(5) How does the FSR of FDIA samples and smoothing operation affect the Siamese-MIL's performance?

Using three datasets (BB-MAS [42], WISDM [64] and WESAD [53]) and two signal modalities (accelerometer (ACC) and blood volume pulse (BVP)), we have generated FDIA variations with different FSRs and smoothing operations. We evaluate FIDA's efficacy in deceiving two ACC-based authentication systems, ACC-based activity detection systems, and a BVP-based stress detection system, indicating that FDIA with 50-60% FSR is highly effective in deceiving both the CASs and the human-centric sensing applications. Additionally, smoothing operations does not increase FDIA's deception capability.

Siamese-MIL achieves an average 92.66% F1-score on the three datasets, showing its high efficacy in FDIA detection. The further evaluation shows that the Siamese-MIL *and its integration to authentication (Appendix C)* achieve a high attack detection accuracy against all FSR attack samples, and the smoothing operation does not significantly influence its performance. Additionally, we evaluate and demonstrate Siamese-MIL's real-time execution capability and resource usage in resource constraint smart devices.

4 GENERATED FALSE DATA INJECTION ATTACK DATASETS AND DATA PROCESSING

To our knowledge, there is no existing human-centric sensing dataset that contains injection attack data. Following the recent false data injection attack work [17, 32], we develop synthetic attack datasets to simulate the FDIA. We use three publicly available datasets - BB-MAS [42], WISDM [64] and WESAD [53] to generate FDIA samples on different applications: continuous authentication, activity detection and emotional stress detection.

For each of the evaluations, we follow the **person-disjoint hold-out method** [9]. Each dataset contains data from different individuals. To avoid personal bias and make the models generalizable, we separate each dataset into person-disjoint training, validation, and test subsets. Every subset only contains data from some specific individuals, and one user's data won't appear in two subsets. This separation is performed randomly five times; hence we have five groups of person-disjoint training, validation, and test subset combinations. All presented results are averaged over the five groups to reduce contingency and avoid overfitting of the model. Detail discussion of each dataset is below:

(1) *BB-MAS Dataset* [42] contains 3-axis smartphone ACC sensor readings (100 Hz) from 117 participants' gait. Each individual participated in two 2-min (on average) data collection sessions where the smartphone was placed in the participant's pant pocket. After filtering out participants with missing data, we obtain gait data sessions from 96 individuals. BB-MAS dataset is used to investigate how FDIA samples impact on a gait-based

authentication system. For evaluation, the participants are randomly divided into - 70 in training, 10 in validation, and 16 in testing. We generate $n = 80$ data samples (pure and attack data) from each participant.

(2) *WISDM Dataset* [64] contains 3-axis smartphone ACC data (20 Hz) from 51 individuals performing 18 different daily living activities. During data collection, the smartphone was placed in the participant's pant pocket. WISDM dataset is used to investigate FDIA samples' impacts on a motion-based authentication system and a human activity detection system. We choose 4 activities - walking, jogging, taking-stairs and kicking a ball. The reasoning for selecting these 4 activities is that they are related and frequent daily human activities. For evaluation, the 51 participants are randomly split into three groups - 30 in training, 6 in validation, and 15 in testing. We generate $n = 80$ data samples (pure and attack) from each individual.

(3) *WESAD Dataset* [53] contains Empatica E4 wristband [31] BVP (64 Hz) signals collected from 15 individuals. Each participant was exposed to stress stimuli through the Trier Social Stress Test (TSST) [25]. For each participant, the dataset contains stress and non-stress physiological data totaling 35 mins. WESAD dataset is used to study how FDIA samples impact on an emotional stress detection detection system. We choose to evaluate on BVP signals among the physiological modalities, since previous studies [4, 51, 55] have shown that the BVP signals contain identifiable human traits. 15 subjects are randomly split into three groups - 7 in training, 4 in validation, and 4 in testing. We generate $n = 850$ data samples (pure and attack) from each individual.

For all datasets: For each subject, we generate a same number of synthetic FDIA and pure data samples. An equal number of different attack data with FSR ranging from 10% to 90% is generated. To evaluate the effect of smoothing, we generate two synthetic dataset variations for each dataset: one with smoothing and the other without smoothing. For BB-MAS, WISDM and WESAD datasets, the period used in exponential moving average smoothing are 0.1-s, 0.2-s, and 0.125-s.

5 DECEPTION RESULT OF FALSE DATA INJECTION ATTACK

This section evaluates the FDIA's efficacy in deceiving smart device CASs and human-centric sensing applications such as activity and emotional stress detection through accelerometer (ACC) and blood volume pulse (BVP) signals. Detailed descriptions of the models used are in Appendix A.

5.1 Evaluation of FDIA on Authentication Systems

This section investigates the question **"Are continuous authentication systems (CASs) effective in detecting false data injection attacks?"**. We evaluate FDIA samples' efficacy in deceiving a gait-based CAS using ACC signal from BB-MAS dataset and a daily-activity motion-signal-based CAS using ACC data from WISDM dataset.

Authentication Systems: Following recent work [1, 11] on continuous ACC-based smart device authentication systems, we developed Siamese convolutional neural networks that learn to differentiate authentic user's ACC data from others. Siamese authentication models take two 10-s ACC signals as inputs: one legitimate user's reference signal and one for authentication and identifies if the inputs are from the same or different individuals. If matched, the authentication is verified, and rejected otherwise. For *BB-MAS* and *WISDM* datasets, the authentication model achieves 90% and 87% F1-scores, 93% and 89% true acceptance rates (consider FSR = 0% samples), and 87% and 84% true rejection rates on differentiating authentic vs. other individuals' signals.

FDIA on Authentication Systems: We evaluate FDIA's deception efficacy against the developed authentication systems considering different factors: forged signal ratio (FSR) and effect of smoothing operation.

As mentioned in Section 2.4, if FDIA replaces the whole legitimate sensory data stream with signals from other individuals (consider it a FSR = 100% sample), this authentication model will have a high possibility (87% on BB-MAS dataset and 84% on WISDM dataset) to detect the attack sample. Hence, the attack sensory signals will

fail to reach the targeting human-centric sensing application. *Therefore, instead of replacing the whole legitimate signal, we consider a stealthier FDIA, generating a mixture of legitimate and forged data, keeping both legitimate patterns and misinformation.*

For BB-MAS dataset, FDIA samples are generated by inserting other individuals' gait-ACC data into the legitimate user's signal; For WISDM dataset, attack samples are generated by inserting others' activity jogging (B) and taking-stairs (C) ACC signals into the legitimate user's activity walking (A) data, noted as A/B and A/C respectively. Attack samples have FSRs ranging from 10% to 90%, and we evaluated both smoothed and non-smoothed variations.

Table 1. ACC-based authentication models' rejection rate on FDIA samples with different FSRs and smoothing variations.

(a) Rejection rate of FDIA variations from BB-MAS dataset

FSR	smoothed	non-smoothed
10%	8.89%	9.86%
20%	15.92%	17.48%
30%	26.95%	29.30%
40%	44.53%	43.46%
50%	55.67%	54.89%
60%	67.97%	68.48%
70%	75.39%	77.05%
80%	84.08%	83.11%
90%	85.36%	84.96%

(b) Rejection rate of FDIA variations from WISDM dataset

FSR	A/B		A/C	
	smoothed	non-smoothed	smoothed	non-smoothed
10%	18.10%	18.12%	16.52%	16.34%
20%	32.12%	31.84%	27.74%	27.41%
30%	48.51%	48.38%	41.57%	41.51%
40%	62.79%	63.79%	55.25%	55.28%
50%	72.53%	71.98%	65.90%	66.61%
60%	77.93%	77.38%	72.86%	72.92%
70%	81.76%	81.33%	76.40%	77.72%
80%	83.49%	83.73%	80.18%	80.16%
90%	85.59%	86.14%	82.34%	82.19%

Table 1a and 1b show the true rejection (i.e., attack detection) rate of the authentication models on different FDIA variations from BB-MAS and WISDM dataset. On low FSR attacks samples where only a small fraction (10-20%) of the ACC signal is forged, the authentication systems successfully reject only 8-32% of the FDIA samples. This is due to the high similarity of the majority of the attack signal portions with the legitimate user's movements. As a higher ratio of forged information is injected, the authentication systems achieve a higher efficacy in capturing the attack. This evaluation demonstrates that, *when the FSR is moderate (30-60%), FDIA samples have a good chance (23-73%) to deceive the authentication systems successfully; When FSR is greater than 70%, FDIA samples are very likely to be detected. Furthermore, both smoothed and non-smoothed attacks achieved similar performance.*

5.2 Evaluation of FDIA on Human-centric Sensing Applications

This section investigates the question **"Can false data injection attacks deceive human-centric sensing applications?"**. We evaluate FDIA samples on a human activity detection model based on accelerometer (ACC) data (Section 5.2.1) and a stress detection model based on blood volume pulse (BVP) data (Section 5.2.2).

5.2.1 FDIA on Human Activity Detection System. This section's evaluations are performed on the WISDM dataset.

Activity Detection System: For four activities - walking (A), jogging (B), taking-stairs (C), and kicking-balls (M), we developed binary activity detection (AD) classifiers that take a 10-s ACC signal as input. Following the state-of-the-art study DeepSense [65], we develop an integration of convolution (CNN) and Long Short-Term Memory (LSTM) network, named CNN-LSTM model as the AD models. Classifiers for activities A, B, C, and M have the detection accuracy of 74%, 87%, 72%, and 88%.

FDIA on Activity Detection System: We generate three types of FDIA samples: (1) insert jogging (B) data into walking activity (A) data (A/B case); (2) insert kicking-ball (M) data into walking activity (A) data (A/M case); (3)

insert jogging (B) data into taking-stairs activity (C) data (C/B case), with different FSRs. Both smoothed and non-smoothed FDIA variations were generated.

Generated attack samples are evaluated on the corresponding AD models. E.g., A/B attack samples are generated to misinform the AD system that users are jogging while originally walking. So, the samples are evaluated by AD models of activity A and B. Percentage of these A/B attack samples detected as respective activity by the walking (A) or jogging (B) detection models are shown in Table 2a and 2b. When FSR is 60% or higher, less than 13% of the A/B samples are detected as walking (A) by the AD model of A, and more than 50% are detected as jogging (B) activity by AD model of B. Similarly, we have the results for A/M and C/B cases. For 60% or higher FSR, A/M and C/B FDIA samples are highly effective in deceiving the AD system that the target user is kicking balls (M) and jogging (B). *This section's results establish that both smoothed and non-smoothed FDIA samples can successfully deceive activity detection models into inferring a targeted wrong activity.*

Table 2. Classification results of activity detection models on smoothed and non-smoothed FDIA samples with different FSRs for jogging in walking (A/B), kicking-balls in walking (A/M) and jogging in taking-stairs (C/B)

(a)

| FSR | Classification result on smoothed attack samples | | | | | |
| | A/B | | A/M | | C/B | |
	A	B	A	M	C	B
10%	58.34%	7.67%	58.75%	25.33%	61.99%	7.49%
20%	47.50%	8.96%	52.92%	37.33%	51.20%	6.90%
30%	44.58%	16.84%	46.25%	47.17%	45.15%	11.90%
40%	32.50%	22.67%	44.17%	56.33%	40.45%	20.61%
50%	18.34%	34.67%	32.09%	65.83%	35.73%	32.87%
60%	12.50%	50.50%	26.25%	75.83%	27.82%	44.61%
70%	6.67%	66.33%	19.17%	83.32%	21.10%	69.74%
80%	2.94%	81.37%	17.50%	86.18%	18.01%	79.75%
90%	4.17%	86.50%	25.42%	88.67%	10.67%	87.88%

(b)

| FSR | Classification result on non-smoothed attack samples | | | | | |
| | A/B | | A/M | | C/B | |
	A	B	A	M	C	B
10%	55.00%	7.99%	61.67%	24.33%	56.29%	8.51%
20%	44.58%	13.33%	53.34%	38.17%	45.02%	9.29%
30%	41.25%	20.00%	47.50%	48.83%	40.60%	16.51%
40%	34.17%	24.83%	45.42%	59.00%	36.74%	24.06%
50%	17.50%	36.83%	34.17%	66.67%	31.92%	34.20%
60%	8.59%	55.50%	25.83%	77.33%	25.18%	48.89%
70%	6.25%	73.33%	18.75%	82.67%	15.57%	73.02%
80%	2.50%	82.50%	17.50%	86.17%	14.39%	83.22%
90%	6.25%	86.50%	25.00%	88.50%	10.33%	88.06%

5.2.2 FDIA on Emotional Stress Detection System. These evaluations are performed on the WESAD dataset.

Emotional Stress Detection System: Following recent works [24, 46] which use CNN-LSTM algorithm to detect mental stress based on physiological signal, we developed a binary stress detection CNN-LSTM classifier that takes a 20-s BVP signal at each 20-s interval. The classifier achieves an F1-score of 79%, a 75% accuracy on stress detection, and a 86% TNR, meaning the accuracy of detecting non-stress signals is 86%.

FDIA on Stress Detection System: We generate FDIA samples by inserting stressed signal portions into the target users' non-stressed BVP signal, with FSRs ranging from 10% to 90%. Table 3 shows the BVP-based stress detection model's classification results on the generated attack samples. Both smoothed and non-smoothed attacks perform similarly in deceiving the stress detection model. With FSR 30 - 40% attacks, about 30% attack samples are detected as stressed. With the increase of FSR, about 65-76% of attack samples are detected as stressed. *This section's evaluation demonstrates that FDIA (specifically on moderate to higher FSRs) effectively deceives the stress detection system by generating forgery 'stress' inferences while the target user is not stressed.*

5.2.3 Conclusion from FDIA Deception Evaluation. According to our evaluation, with 50-60% FSR, FDIA samples can deceive the authentication system with 32-45% false-sample-acceptance-rate, the activity detection models with 35-77% wrong-targeted-activity-inference rate, and the stress detection model with 42-55% misclassification rate. With lower FSR, FDIA samples perform highly in deceiving the authentication but poorly in deceiving the sensing applications. With higher FSR, FDIA samples perform poorly in deceiving the authentication system,

Table 3. Stress detection model's classification results on BVP attack samples with different FSRs

FSR	smoothed		non-smoothed	
	stressed	non-stressed	stressed	non-stressed
10%	19.35%	80.65%	18.00%	82.00%
20%	23.12%	76.88%	23.12%	76.88%
30%	27.12%	72.88%	27.59%	72.41%
40%	31.94%	68.06%	34.59%	65.41%
50%	41.59%	58.41%	43.35%	56.65%
60%	54.41%	45.59%	53.24%	46.76%
70%	65.59%	34.41%	65.94%	34.06%
80%	72.88%	27.12%	73.06%	26.94%
90%	76.76%	23.24%	76.06%	23.94%

hence cannot reach the sensing applications. Hence, 50-60% FSR is optimal for deceiving smart device sensing applications. Additionally, smoothing and non-smoothing attacks perform similarly.

6 METHODOLOGY: MAPPING SIAMESE-MIL FOR FDIA DETECTION

Siamese-MIL approach identifies if a signal comprises at least a pair of segments belonging to different individuals. It segments a signal sample into a set of all possible segment pairs (Section 6.2.1). A Siamese neural network (SNN) [7, 68] is trained to identify any segment pair that contains signals from different individuals. If the trained SNN identifies at least one mismatched segment pair, the respective signal sample is considered an FDIA sample (Section 6.2.2). SNN is trained through the MIL paradigm, where it is tailored to be highly effective in detecting matched (legit) segment pairs (i.e., high recall) where mismatched segment pair detection accuracy (true negative rate) can be lower. Additionally, it effectively avoids the impact of erroneous signal-labels during SNN training (Section 6.2.3). The attack detection framework and a background on SNN and MIL are discussed below.

6.1 Background Discussion

Siamese Neural Network (SNN). [7, 68] employs a unique structure to naturally compare a pair of inputs in terms of their semantic similarity or dissimilarity. This paper leverages SNN to distinguish input sensory signal samples of different individuals. (Detailed discussion in Appendix D)

Multiple Instance Learning (MIL). is a weakly supervised learning problem where, the input of a classifier is considered as a bag of instances, $B = \{x_1, x_2, ..., x_n\}$. Instances exhibit neither dependency nor ordering among each other. Each bag of instances B has an associated single binary label $Y \in \{0, 1\}$ known during training. However, individual labels of the instances within a bag remain unknown. The assumption of a MIL problem is:

$$Y = 0 \Leftrightarrow \exists x_j \in B, \ y_j = 0, \text{ and } Y = 1 \Leftrightarrow \forall x_j \in B, \ y_j = 1 \quad (1)$$

According to the MIL assumption, known labels are attached to the bags, where a positive bag has label $Y = 1$, and a negative bag has label $Y = 0$. A negative bag has at least one negative instance (i.e., $\exists x_j \in B, \ y_j = 0$), and a positive bag contains positive instances only (i.e., $\forall x_j \in B, \ y_j = 1$). This assumption generates an asymmetry from a learning perspective as all instances in a positive bag can be uniquely assigned a positive label, which cannot be done for a negative bag (which may contain both positive and negative instances). Thus, the relationship between bag label Y and instance label y_j is: $Y = min\{y_j\}$. The Siamese-MIL classifier model training approach that adapts MIL mechanism is discussed in the following section.

6.2 Siamese-MIL Approach

This section presents a novel MIL bag & instance generation mechanism for FDIA detection (Section 6.2.1), Siamese-MIL attack detection approach (Section 6.2.2), and the Siamese-MIL training approach (Section 6.2.3).

6.2.1 Siamese-MIL Bag and Instance Generation. In the FDIA, sensory data from other individuals is mixed with the legitimate user's data. We leverage this characteristic to detect the attack.

Siamese-MIL takes a W-s sample in the form of a MIL bag B to identify attacks. If the attack generated forged data is present in the W-s sample, the MIL bag label is negative (i.e., 0), and the label is positive (i.e., 1) otherwise. The novel contribution of the paper is how the bag instances are constructed. We segment the W-s sample into l segments x_k with overlap rate R, where $k = 1, 2, \ldots, l$, and the length of each x_k is 'V' seconds. We compose $_lC_2 = \frac{l!}{2!(l-2)!}$ (i.e., combination) pairs of small segments (x_i, x_j), $i \neq j$, ensuring that each small segment x_i is once paired with all other segments. These pairs are the instances of bag B representing the W-s sample.

According to the definition (Section 2.4), not all segments x_k in a W-s FDIA sensory sample will contain the same individual's data. In a bag B (representing W-s sample), if at least one segment pair (x_i, x_j) is such that they are not containing data from same individual, the instance-level label of that pair is 0, hence the bag label Y is 0, meaning the bag is containing injected attack synthetic data.

6.2.2 Siamese-MIL FDIA Detection Mechanism. Algorithm 1 demonstrates the FDIA detection approach. An SNN is trained to detect FDIA ('net' in Algorithm 1) that takes an instance (i.e., a (x_i, x_j) pair) as input and identifies if the small segments are from the same individual. If yes, the inferred instance-level similarity score is > 0.5 (line 6-10). The label of the bag is determined by the pair with the lowest similarity score. The lowest score > 0.5 means all instance-pairs are classified as positive (i.e., samples from same individual), inferring the bag as positive and the W-s sample is legitimate; otherwise, the algorithm indicates that there exists at least one negative instance-pair (i.e., samples from different individuals) in the bag, inferring the bag label as negative and the W-s sample is corrupted due to FDIA.

Algorithm 1 Siamese-MIL Attack Detection

Require: a Siamese network (*net*), a W-s sample (s), number of pairs to be extracted from s (l), small segment size (V-s), overlap rate (R)

1: Initialize $i = 0$, $minScore = 1$, $bag = \{\}$
2: **procedure** Bag Instance Generation($bag = \{\}$)
3:     Extract R overlapping 'V's signals x_k from s
4:     $bag = \forall(x_i, x_j)$, where $i \neq j$ and $(x_j, x_i) \notin bag$
5: **end procedure**
6: **for** each instance (x_i, x_j) in bag **do**
7:     $currentScore = net(x_i, x_j)$
8:     **if** $currentScore \leq minScore$ **then**
9:         $minScore = currentScore$
10:     **end if**
11: **end for**
12: **return** 1 if $minScore > 0.5$
13: **return** 0 otherwise

Rule of Three Approximation of MIL Paradigm. To further demonstrate the reliable and effective attack detection through Siamese-MIL, we make some approximations. Suppose our SNN has a true negative rate 'p'. We segment a corrupted W-s signal sample into l segments, where m segments contain corrupted other individual's data. We can consider a segment pair (x_i, x_j) impure if it contains different individuals' data, and there will be $G = {}_lC_2 - {}_mC_2 - {}_{(l-m)}C_2$ (if $m \geq 2$) or $G = l - 1$ (if $m = 1$) impure pairs in the bag. A negative bag is misclassified if and only if all impure instance-pairs are misclassified. Considering the evaluation of each instance-pair is mutually independent, the probability of misclassifying a negative bag is $(1 - p)^G$. According to the Rule of

Three [23], p has to be in the range of $[0, \frac{3}{G}]$ for negative bag misclassification with 95% confidence. Oppositely, if $p > \frac{3}{G}$, we have 95% confidence that the negative bags (attacked samples) will be classified correctly.

Consider the detection of the lowest FSR level 10%. We segment the W=10-s sensory signal into nine 2-s segments with 50% overlap. That means, we will have 36 instance-pairs. If 2 segments (most likely) contain the forged data, $G = 14$. Therefore, if the SNN has $p > \frac{3}{14} = 21.4\%$, the Siamese-MIL will detect the attack successfully with 95% confidence. Though it is an ideal approximation, it gives the insights about Siamese-MIL's effective attack detection capability. We evaluate the concept further in Section 7.1.

6.2.3 Siamese-MIL Training Approach. We define a loss function according to the MIL training paradigm, where the loss E_b is defined by Equation 2.

$$E_b = -\frac{1}{N}\sum_{i=1}^{N}(Y_i \log s_i^{min} + (1 - Y_i)\log(1 - s_i^{min})), \text{ where } s_i^{min} = \min\{s_i^1, s_i^2, ..., s_i^m\} \qquad (2)$$

Here, N is the training batch size (i.e., number of bags or W-s input windows in a batch), s_i^j is the similarity score of j-th instance-pair of the i-th bag, m is number of instances in a bag and Y_i is the label of i-th bag. The loss function (Equation 2) penalizes a bag B_i on the difference between bag label (Y_i) and the lowest instance-level score (i.e., similarity score discussed) in the bag.

In FDIA detection task, the sensory signal from a W-s detection window is weakly labeled, where only the bag-level label is available. Since the instance-pair-level labels are not available during training, supervised training approaches consider labels of all the instance-pairs of a negative bag as negative. Due to such erroneous instance-pair-level label assumption, the supervised learning approaches fail to achieve an optimal solution.

In Siamese-MIL training, the weights are updated according to the loss on the instance-pair whose corresponding similarity score is minimum among all the instance-pairs in the bag. If at least one instance-pair of a negative bag (i.e., $Y_i = 0$) has similarity score 0, the loss value on the concerned bag B_i is zero and the weights of the network will not be updated. Therefore, the Siamese-MIL training avoids weight updates due to positive instance-pairs in the negative training samples (or bags). For positive bag training, if all the instance-pairs are perfectly predicted as positive, then only the loss value on the concerned bag is zero and the weights of the network are not updated.

7 EVALUATION OF ATTACK DETECTION USING SIAMESE-MIL

This section investigates the question **" What is the performance of the Siamese-MIL detection approach against smart device sensor false data injection attacks?"**. We evaluate the attack detection performance of Siamese-MIL on smoothed gait-ACC FDIA samples used on authentication (Section 5.1), motion-ACC FDIA samples used on activity detection (Section 5.2.1) and BVP FDIA samples used on stress detection (Section 5.2.2) models discussed above. An equal number of attack samples were generated for each FSR, ranging from 10% to 90%. The effect of smoothing operation and different FSR variations of FDIA on the Siamese-MIL's performance will be discussed in Section 8. The performance of Siamese-MIL is compared to a baseline CNN-LSTM, following the papers [12, 20, 32, 65, 66] that address attacks and sensory signals detection. Detailed descriptions of the baseline and SNN models which are used in the evaluations are provided in Appendix A.

Following sections discuss the beneficial parameters of the Siamese-MIL approach and attack detection performance on gait-ACC (Section 7.1), motion-ACC (Section 7.2), and BVP (Section 7.3) samples.

7.1 Attack Detection Performance on Gait-ACC FDIA Samples from BB-MAS dataset

Beneficial Parameter Configurations: Siamese-MIL takes a $W = 10$-s 3-axis ACC sample (3×1000 dimension) as input to assess FDIA. We segment the sample into $V = 2$-s small segments with overlap rate $R = 50\%$ to compose the instance-pairs, then extract 36 instance-pairs (i.e. 9 small segments) from the sample. W, V, and R are hyper-parameters, and an ablation study on Siamese-MIL bag parameters is discussed in Appendix B.

Attack Detection Performance: Table 4a shows the attack detection performance of Siamese-MIL and baseline CNN-LSTM model on smoothed gait-ACC FDIA samples. Siamese-MIL significantly outperforms the baseline. The Siamese-MIL has a higher recall, implicating better legitimate signal assessment performance. *Notably, the Siamese-MIL has much better attack detection performance (11% higher precision and 17% higher TNR) than the baseline.* This is due to the MIL characteristic, where only one negative instance-pair is needed to identify a negative bag (i.e., attack sample). Thus, even if the SNN alone is not highly accurate in distinguishing two samples (i.e., instance-pairs) from different individuals, with enough instance-pairs within a bag, Siamese-MIL evaluation gets a higher chance of correctly identifying an attack sample (Section 6.2.2).

Insights on SNN's performance: We further evaluate the SNN's performance on differentiating instance-pairs from the same or different individuals. We generate pairs consisting of segments from a legitimate person (positive instance-pair) and pairs containing segments from different individuals (negative instance-pair). The SSN trained on smoothed gait-ACC data achieves a very high TPR of 98.95% and a lower TNR of 46.34%. The result confirms that SNN and Siamese-MIL are following the MIL and Rule of Three assumptions (discussed in Section 6.2.2). In the MIL framework, a positive bag is misclassified with even one misclassification of a positive instance-pair. Hence, a high TPR is required by the SNN to achieve high legit signal detection (i.e., high recall) performance by Siamese-MIL. Additionally, according to the Rule of Three assumption, above 21.4% TNR is needed to achieve high attack detection accuracy (i.e., for 36 instance-pairs MIL bags), where our SNN achieves a 46.34% TNR, resulting in high attack detection performance by the Siamese-MIL.

Table 4. Evaluation of CNN-LSTM and Siamese-MIL on smoothed FDIA samples from BB-MAS and WESAD dataset

(a) Evaluation on gait-ACC FDIA samples from BB-MAS dataset

Model	CNN-LSTM	Siamese-MIL
F1-score	82.41%	90.35%
Precision	77.45%	85.99%
Recall	89.22%	95.25%
TNR	72.16%	84.63%

(b) Evaluation on BVP FDIA samples from WESAD dataset

Model	CNN-LSTM	Siamese-MIL
F1-score	93.03%	98.32%
Precision	94.08%	97.58%
Recall	92.07%	99.10%
TNR	94.86%	97.48%

7.2 Attack Detection Performance on Motion-ACC FDIA Samples from WISDM dataset

Beneficial Parameter Configurations: Similar to the gait-ACC data evaluation, the hyper-parameters for the motion-ACC data evaluation are: $W = 10$-s, $V = 2$-s, and $R = 50\%$. Due to the lower sampling rate (20Hz), input data dimension is 3×200.

Attack Detection Performance: Besides A/B, A/M, C/B attack cases used in Section 5.2.1, we generate two more variations of FDIA samples: (1) insert taking-stairs data (C) into walking activity (A) data (A/C) case, and (2) insert walking data (A) into taking-stairs (C) data (C/A) case. Tables 5a and 5b show the attack detection performance of the baseline and Siamese-MIL. Siamese-MIL achieves 5.9%, 12.4%, 9.9%, 1.8% and 7.8% higher F1-scores in A/B, A/C, A/M, C/B, C/A attack cases. Notably, it achieves slightly better precision and TNR, and significantly higher recall than the baseline. *This evaluation indicates that, in general, the Siamese-MIL has a similar attack detection performance as CNN-LSTM, but it is significantly less likely to misclassify a legitimate input.*

7.3 Attack Detection Performance on BVP FDIA Samples from WESAD Dataset

Beneficial Parameter Configurations: Our approach takes a $W = 20$-s BVP sample (1×1280 dimension) as input. The input is divided into $V = 4$-s small segments (1×256 dimension) with overlap rate $R = 50\%$.

Table 5. Evaluation of CNN-LSTM and Siamese-MIL on smoothed motion-ACC FDIA samples from WISDM dataset

(a) CNN-LSTM evaluation

Activity	A/B	A/C	A/M	C/B	C/A
F1-score	87.18%	76.68%	80.72%	90.89%	80.52%
Precision	91.88%	82.77%	87.28%	89.84%	77.29%
Recall	83.80%	72.90%	76.67%	91.25%	84.11%
TNR	92.13%	84.20%	87.97%	93.12%	75.19%

(b) Siamese-MIL evaluation

Activity	A/B	A/C	A/M	C/B	C/A
F1-score	92.35%	86.18%	88.73%	92.52%	86.82%
Precision	92.50%	83.87%	88.33%	91.21%	83.13%
Recall	92.57%	89.53%	89.50%	93.96%	91.18%
TNR	92.36%	82.22%	87.85%	91.08%	81.48%

Attack Detection Performance: Table 4b shows the attack detection performance of Siamese-MIL and baseline on smoothed BVP FDIA samples. The Siamese-MIL outperforms the baseline and achieves a high recall and TNR, *demonstrating that Siamese-MIL is highly effective in detecting both the legitimate and attack BVP signals.*

In conclusion: According to Section 7's evaluation, Siamese-MIL is highly effective in detecting FDIA (high precision and TNR) and legitimate (high recall) gait-ACC, motion-ACC, and BVP signals and outperforms the baseline CNN-LSTM models significantly.

8 EVALUATION ON FDIA CHARACTERISTICS

The evaluations in Sections 5.1 and 5.2 demonstrate that FDIA with moderate FSR has a good chance to deceive authentication systems and FDIA (specifically on moderate to higher FSRs) effectively deceives human-centric sensing systems, therefore, this section investigates the question **"How does the FSR of FDIA samples and smoothing operation affect the Siamese-MIL's performance?"** by evaluating the Siamese-MIL's performance against FSR ranging from 10-90% and smoothing attack variations on all three datasets. Additionally, we evaluate Siamese-MIL's performance against attack samples when injected signals are from the same individual or different individuals on the BB-MAS and WESAD datasets.

Table 6. Attack detection rate (TNR) of Siamese-MIL on FDIA samples with different FSRs from BB-MAS and WESAD dataset

(a) Performance on gait-ACC FDIA samples from BB-MAS dataset

FSR	smoothed	non-smoothed
10%	80.19%	75.63%
20%	86.41%	84.38%
30%	88.91%	86.49%
40%	88.99%	85.32%
50%	87.58%	83.83%
60%	87.90%	86.96%
70%	88.28%	85.32%
80%	83.75%	81.88%
90%	67.81%	67.50%

(b) Performance on BVP FDIA samples from WESAD dataset

FSR	smoothed	non-smoothed
10%	97.06%	98.65%
20%	97.92%	98.77%
30%	97.68%	99.09%
40%	97.71%	98.39%
50%	97.85%	98.74%
60%	97.68%	98.65%
70%	98.06%	98.71%
80%	97.33%	98.44%
90%	97.44%	97.80%

Evaluation on FSR Variations. Tables 6a, 6b and 7 display Siamese-MIL's attack detection rate (TNR) with different variations on the three datasets. *According to our evaluation, Siamese-MIL achieves high attack detection rate (TNR) on 20 - 80% FSR, where performance drops slightly for 10% or 90% FSR attacks.*

The reason is, on 10% or 90% FSR attack samples, only a 1-s data segment (out of 10-s) is different than the rest. Since our Siamese-MIL instances are 2-s with 50% overlaps, at least one instance (out of 9) will contain a different signal, and the number of mismatched input instance-pairs is the lowest (8 out of 36). Compared to that, on 40% or 60% FSR attack samples, at least 4-s data segment (out of 10-s) is different, meaning at least 3 (out of 9)

Table 7. Siamese-MIL on smoothed and non-smoothed motion-ACC FDIA samples with different FSRs from WISDM dataset

(a) Detection rate (TNR) on smoothed data

FSR	A/B	A/C	A/M	C/B	C/A
10%	71.25%	75.33%	75.50%	71.25%	72.38%
20%	88.50%	83.42%	86.67%	87.16%	80.90%
30%	93.84%	83.92%	89.00%	92.87%	84.68%
40%	95.17%	85.42%	90.67%	94.73%	83.94%
50%	96.33%	84.84%	92.34%	94.62%	82.87%
60%	96.75%	86.58%	92.25%	96.12%	85.44%
70%	98.25%	86.92%	92.42%	95.14%	81.75%
80%	97.00%	83.58%	92.00%	96.31%	82.31%
90%	97.50%	71.50%	84.58%	95.62%	69.38%

(b) Detection rate (TNR) on non-smoothed data

FSR	A/B	A/C	A/M	C/B	C/A
10%	79.17%	66.09%	75.75%	74.34%	72.73%
20%	93.17%	78.00%	86.84%	89.84%	80.00%
30%	96.50%	78.50%	89.09%	92.71%	82.65%
40%	96.50%	81.09%	90.67%	94.19%	82.31%
50%	96.92%	78.92%	92.33%	94.47%	80.91%
60%	97.09%	81.01%	92.50%	96.39%	84.35%
70%	98.00%	82.92%	92.84%	94.87%	79.96%
80%	97.17%	79.75%	93.42%	95.94%	82.32%
90%	97.75%	63.17%	88.09%	95.47%	73.07%

instances will contain different signals than the rest. Hence, the number of mismatched input instance-pairs is a minimum of 18 (out of 36). Therefore, an SNN with lower mismatched pair detection accuracy would have a significantly high probability to correctly detect at least one mismatched instance-pair (from 18 instance-pairs) on the 40% or 60% FSR attack samples, compared to the 10% or 90% FSR attack samples (from 8 instance-pairs).

Notably, the Siamese-MIL achieves high attack detection performance against the BVP FDIA samples with all FSR variations. BVP signals during high and low emotional stress are distinctively different. Hence, even on high and low FSR, the forged and legit signal portions are easily distinguishable.

Insights from the Siamese-MIL's Performance on the Activity Misleading FIDA:. As shown in Table 7, on A/B and C/B attack samples, Siamese-MIL achieves 95-97% TNR on the 90% FSR samples where only 10% ACC data is from the legit user. This is caused by the higher ACC signal amplitude differences between activity B (i.e., jogging) and others. On 10% FSR attack, Siamese-MIL achieves relatively lower performance since a small high amplitude signal segment may even present in walking or taking-stairs activity signals, making the attack detection task difficult. On A/M attack samples, Siamese-MIL performs similarly. But in A/C and C/A attack cases, taking-stairs (C) activity consists of some walking and some climbing-steps, making them very similar to walking (A). Hence, on 10% or 90% FSR attacks, where only 10% of the signal is different from the rest, attack detection is challenging. Nevertheless, Siamese-MIL still achieves 70 - 75% TNR on the 10% or 90% FSR attack. Overall, *Siamese-MIL achieves a higher attack detection rate on FDIA samples where the legit and injected signal traits are highly dissimilar.*

Evaluation on Smoothing Operation. According to the Tables 6a, 6b and 7 results, Siamese-MIL performs similarly on smoothed and non-smoothed attack samples. Sensor data (i.e., ACC, BVP) contains noisy fluctuations similar to the forge-and-legit-signal-boundary-mismatches in the attack samples. So, even in the non-smoothed attack samples, classifiers cannot differentiate the signal fluctuations are due to noise or FDIA attack.

This section also **evaluates whether the injected signals from the same or different individuals affect the Siamese-MIL's performance.** We evaluated two kinds of attack samples: (1) injected signals are from the same person, and (2) injected signals are from multiple people, on BB-MAS and WESAD datasets. Siamese-MIL achieves similar performance on both variations (Table 8). It is due to the MIL paradigm - only one dissimilar instance-pair is needed to determine an attack. Though with the increase of injected signals' sources in an attack sample, the number of dissimilar instance-pairs (i.e., containing data from different individuals) increases, Siamese-MIL only achieves a marginally higher TNR against the multiple people injected attack samples. Hence, the number of sources of injected forged signals does not significantly affect the Siamese-MIL's performance.

In conclusion: According to the evaluation, low and high FSRs make FDIA detection harder. Smoothing operation or the number of sources of the injected forged signals in an attack sample do not significantly influence attack detection. However, Siamese-MIL performs consistently higher (on all FSRs) against the FDIA samples where the

Table 8. TNR of Siamese-MIL against attack samples when injected data is from a same and different individuals

	BB-MAS	WESAD
Same Person Attack	87.74%	99.80%
Different People Attack	89.04%	99.73%

Table 9. Run time and resource usage evaluation of Siamese-MIL on two phones, time is measured in milliseconds (ms).

	Time to make a bag (ms)	Time to infer a bag (ms)	Peak CPU usage	Peak RAM usage	Battery cost per day
HTC One M9	~0.20	~201.05	~40%	~45MB	~43mAh
Samsung SM-G920F	~0.21	~121.10	~40%	~50MB	~22mAh

legit and injected forged signal traits are highly dissimilar. Notably, the 10% FSR samples are not highly effective in deceiving the human-centric sensing applications, and the CASs detects 90% FSR samples with higher efficacy (Section 5), hence making the relatively lower performance of Siamese-MIL on low and high FSRs less impactful. *Appendix C further provides an FDIA mitigation strategy by integrating Siamese-MIL and CAS that is highly robust against all FSR (i.e., $10 - 90\%$) and smoothing variations of the FDIA.*

9 EXECUTION TIME AND RESOURCE USAGE ANALYSIS OF SIAMESE-MIL

To evaluate The Siamese-MIL's real-time executability and resource usage on smart devices, we deploy the model on two Android phones - HTC One M9 and Samsung SM-G920F, both are with Android 7.0. The program reads 10-s ACC signal (sampling rate 100 Hz) at a time, analyze and infer the class (attack or not) result. We evaluate the execution time of MIL bag generation procedure and Siamese-MIL's inference for a bag, peak CPU and RAM usage, and battery cost. Used SNN model is the one implemented in Section 7.1. We simulate that attack detection is performed every 10-s, therefore, there will be 8640 attack detection per day. Average parameters are shown in Table 9. According to the evaluation, Siamese-MIL model can make one FDIA detection within 250 milliseconds. Though peak CPU usage is about 40%, it won't occupy CPU resource for long. Both the phones have 3 GB RAM and batteries with capacity over 2500 mAh, so, the RAM and battery cost is only around 2%. The results suggest that Siamese-MIL is able to perform a real-time attack detection on resource constraint Android phones.

10 RELATED WORK

FDIA refers to an attack that compromises sensor readings to forge some events that are not happening, aiming to mislead the associated sensing system [2, 57]. FDIA was first introduced in the domain of smart power grid by Liu *et al.* [27]. But due to the growth of the Internet of Things (IoT) in the past decades, a large variety of sensors are widely used in smart devices to perform applications, including authentication, health monitoring, activity recognition, etc.; the threat of injection attacks has been realized in different fields and applications.

Gonzalez-Manzano *et al.* [17] studied the impact of FDIA on sensor-based continuous authentication for smartphones and showed that FDIA could cause a 13.2% error rate. It takes 2 to 8 mins to identify the attack, which gives enough time for the attacker to perform malicious activities. Additionally, FDIA has been evaluated in the context of deep learning-based predictive maintenance (PdM) system [32], microelectromechanical systems (MEMS) [60], and health-care monitoring [3] to cause wrong diagnosis. Also, Sikder *et al.* [56] have shown that FDIA can happen on a GPS device to infer wrong location data.

Advanced sensing data manipulation techniques have been discovered that make FDIA on sensors in smart devices even easier. Attackers can access the communication channels (e.g., Wi-Fi and Bluetooth) and use malware to modify the data being transmitted between modules. Ryan *et al.* [49] demonstrates eavesdropping and packet injection on BLE conversations between different smart device modules. Goyal *et al.* [18], and Rahman *et al.* [45] demonstrate vulnerabilities in sensor-data storage and transmission (through their Bluetooth or Wi-Fi-based) of

smart wearables, such as Fitbit, Garmin, and Jawbone sensors; Rahman *et al.* [44] have developed tools such as Fit Bite and GarMax to eavesdrop and modify fitness sensor data on smart wearable devices as well. BlueDoor [62] developed by Wang *et al.* can read and write sensor data on Bluetooth Low Energy (BLE) devices. Cayre *et al.* [10] describes an attack called InjectaBLE, which allows injecting malicious traffic into an existing BLE connection. Besides accessing communication channels, attackers may manipulate sensors remotely via malware. Mohamed *et al.* [33] present a framework called SMASheD, which can sniff and manipulate many of the Android's restricted sensors using a malicious app with only the INTERNET permission; Spy-sense [16] is a malicious app that exploits the active memory region of sensors and relays the collected information. It can delete or modify sensor data. Our discussion evidentiates that FDIA can be performed through accessing wireless communication channels and via malwares. Instead of focusing on prohibiting such sensory data manipulation, this paper develops an FDIA detection approach that will prohibit any attack sample from reaching human-centric sensing applications.

Though increasingly more studies are addressing the threat of FDIA in various scenarios, to the best of our knowledge, this paper is the first to address a targeted FDIA that can deceive continuous authentication systems (CASs) and misinform multiple human-centric sensing applications simultaneously.

11 STUDY SUMMARY AND DISCUSSION

The identified insights, observation, results and limitation of the presented study are discussed below:

Siamese-MIL follows the MIL paradigm and Rule of Three assumptions. According to the rule of three approximation, only a 21.4% or more TNR by the SNN is required to achieve high FDIA detection accuracy (for 36 instance-pairs MIL bags). Our evaluation in section 7.1 shows the developed SNN has a TNR of 46.34% and a TPR of 98.95%, resulting in a high attack detection performance by the Siamese-MIL. These results confirm that the Siamese network and Siamese- MIL framework follow the MIL paradigm and Rule of Three assumptions.

Low and high FSR make FDIA detection harder. The reason is, only the 10% signal portion of 10% and 90% FSR samples are different from the rest, and the SNN needs to be more accurate (on avg. 2+ times accurate for a 36 instance-pair MIL bag) to capture such mismatch. However, according to Sections 5.1 & 5.2, the 90% FSR samples are the easiest to be rejected by the CAS; and 10% FSR samples are not highly effective in deceiving human-centric sensing systems. Hence, the relatively lower performance of Siamese-MIL is less impactful.

Smoothing does not have a great influence on attack detection. According to Section 8, Siamese-MIL's performance on the smoothed and non-smoothed attack data are quite similar. Sensory data of smart devices comprises noisy fluctuations in the original signals similar to the attack sample's forge and legit signal boundary mismatches. Hence, even in the non-smoothed attack samples, detection classifiers cannot differentiate the signal fluctuations are due to noise or attack. Thus, the detection approaches perform similarly against both variations.

Effect of the signal trait. According to Section 8, Siamese-MIL achieves relatively lower TNR when the injected signal's trait is highly similar to the legitimate signal's trait. E.g., in A/C and C/A attack scenarios. Activity C (taking-stairs) contains some signals of A (walking) and climbing steps, making them very similar to A, so, Siamese-MIL only achieves a lower TNR of 81 - 88%.

FDIA's potential adverse effect and mitigation through Siamese-MIL. Tables 2 and 3 evaluations show that FDIA with 40% or higher FSRs can effectively deceive activity detection and stress detection models. Such deception may result in wrong follow-up interventions leading to health and monetary loss of the victim. Siamese-MIL and CAS integration (Table 11 in Appendix C) achieves a high FDIA detection performance against all presented attack variations, hence would be able to protect the smart-device users from such adverse effects.

12 CONCLUSION

The paper focuses on FDIA on smart device human-centric sensing applications. Our evaluation demonstrates that FDIA with 50-60% 'forged signal ratio (FSR)' can effectively deceive both the authentication and human-centric applications, generating a critically adverse effect on the victim. We presented a novel FDIA detection framework (Siamese-MIL) that generates a unique signal representation suitable for formulating the FDIA detection task as a MIL problem and integrates Siamese network and MIL train-test paradigm for effective attack detection. Our exhaustive evaluation on three datasets (BB-MAS [42], WISDM [64] and WESAD [53]), two modalities (i.e., accelerometer, and blood volume pulse) demonstrates the Siamese-MIL's generalizability and high efficacy against all variations of FDIA attacks. The Siamese-MIL FDIA detection approach is designed to extend the conventional authentication systems, prohibiting any attack signal to reach the human-centric applications. Such integration achieves a high attack detection accuracy on all possible attack variations.

ACKNOWLEDGMENTS

REFERENCES

[1] Osama Adel, Mostafa Soliman, and Walid Gomaa. 2021. Inertial Gait-based Person Authentication Using Siamese Networks. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–7.

[2] Mohiuddin Ahmed and Al-Sakib Khan Pathan. 2020. False data injection attack (FDIA): an overview and new metrics for fair evaluation of its countermeasure. *Complex Adaptive Systems Modeling* 8, 1 (2020), 1–14.

[3] Mohiuddin Ahmed and Abu SSM Barkat Ullah. 2017. False data injection attacks in healthcare. In *Australasian Conference on Data Mining*. Springer, 192–202.

[4] Nazneen Akhter, Hanumant Gite, Gulam Rabbani, and Karbhari Kale. 2015. Heart rate variability for biometric authentication using time-domain features. In *International Symposium on Security in Computing and Communication*. Springer, 168–175.

[5] Ala Alwan et al. 2011. *Global status report on noncommunicable diseases 2010.* World Health Organization.

[6] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. 2015. Physical human activity recognition using wearable sensors. *Sensors* 15, 12 (2015), 31314–31338.

[7] Jane Bromley, James W Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature verification using a "siamese" time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence* 7, 04 (1993), 669–688.

[8] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 1–33.

[9] Gavin C Cawley and Nicola LC Talbot. 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research* 11 (2010), 2079–2107.

[10] Romain Cayre, Florent Galtier, Guillaume Auriol, Vincent Nicomette, Mohamed Kaâniche, and Géraldine Marconato. 2021. InjectaBLE: Injecting malicious traffic into established Bluetooth Low Energy connections. In *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*.

[11] Mario Parreño Centeno, Yu Guan, and Aad van Moorsel. 2018. Mobile based continuous authentication using deep features. In *Proceedings of the 2nd International Workshop on Embedded and Mobile Deep Learning*. 19–24.

[12] Mohit Dua, Chhavi Jain, and Sushil Kumar. 2021. LSTM and CNN based ensemble approach for spoof detection task in automatic speaker verification systems. *Journal of Ambient Intelligence and Humanized Computing* (2021), 1–16.

[13] Pablo Fernandez-Lopez, Judith Liu-Jimenez, Kiyoshi Kiyokawa, Yang Wu, and Raul Sanchez-Reillo. 2019. Recurrent neural network for inertial gait user recognition in smartphones. *Sensors* 19, 18 (2019), 4054.

[14] Pablo Fernandez-Lopez, Judith Liu-Jimenez, Carlos Sanchez-Redondo, and Raul Sanchez-Reillo. 2016. Gait recognition using smartphone. In *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*. IEEE, 1–7.

[15] Victor Garcia and Joan Bruna. 2017. Few-shot learning with graph neural networks. *arXiv preprint arXiv:1711.04043* (2017).

[16] Thanassis Giannetsos and Tassos Dimitriou. 2013. Spy-sense: Spyware tool for executing stealthy exploits against sensor networks. In *Proceedings of the 2nd ACM workshop on Hot topics on wireless network security and privacy*. 7–12.

[17] Lorena Gonzalez-Manzano, Upal Mahbub, Jose M de Fuentes, and Rama Chellappa. 2020. Impact of injection attacks on sensor-based continuous authentication for smartphones. *Computer Communications* 163 (2020), 150–161.

[18] Rohit Goyal, Nicola Dragoni, and Angelo Spognardi. 2016. Mind the tracker you wear: a security analysis of wearable health trackers. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. 131–136.

[19] Alejandro S Guinea, Andrey Boytsov, Ludovic Mouline, and Yves Le Traon. 2018. Continuous identification in smart environments using wrist-worn inertial sensors. In *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 87–96.

[20] Md Hasan, Rafia Nishat Toma, Abdullah-Al Nahid, MManjurul Islam, Jong-Myon Kim, et al. 2019. Electricity theft detection in smart grid systems: A CNN-LSTM based approach. *Energies* 12, 17 (2019), 3310.

[21] Mohammed Mehedi Hassan, Md Zia Uddin, Amr Mohamed, and Ahmad Almogren. 2018. A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems* 81 (2018), 307–313.

[22] Anil K Jain, Arun Ross, and Sharath Pankanti. 2006. Biometrics: a tool for information security. *IEEE transactions on information forensics and security* 1, 2 (2006), 125–143.

[23] Borko D Jovanovic and Paul S Levy. 1997. A look at the rule of three. *The American Statistician* 51, 2 (1997), 137–139.

[24] Mingu Kang, Siho Shin, Jaehyo Jung, and Youn Tae Kim. 2021. Classification of Mental Stress Using CNN-LSTM Algorithms with Electrocardiogram Signals. *Journal of Healthcare Engineering* 2021 (2021).

[25] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. 1993. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 1-2 (1993), 76–81.

[26] Cheng-I Lai, Alberto Abad, Korin Richmond, Junichi Yamagishi, Najim Dehak, and Simon King. 2019. Attentive filtering networks for audio replay attack detection. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6316–6320.

[27] Yao Liu, Peng Ning, and Michael K Reiter. 2011. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)* 14, 1 (2011), 1–33.

[28] Upal Mahbub, Jukka Komulainen, Denzil Ferreira, and Rama Chellappa. 2019. Continuous authentication of smartphones based on application usage. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 1, 3 (2019), 165–180.

[29] Anindya Maiti, Ryan Heard, Mohd Sabra, and Murtuza Jadliwala. 2018. Towards inferring mechanical lock combinations using wrist-wearables as a side-channel. In *Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks*. 111–122.

[30] Anindya Maiti, Murtuza Jadliwala, Jibo He, and Igor Bilogrevic. 2018. Side-channel inference attacks on mobile keypads using smartwatches. *IEEE Transactions on Mobile Computing* 17, 9 (2018), 2180–2194.

[31] Cameron McCarthy, Nikhilesh Pradhan, Calum Redpath, and Andy Adler. 2016. Validation of the Empatica E4 wristband. In *2016 IEEE EMBS international student conference (ISC)*. IEEE, 1–4.

[32] Gautam Raj Mode, Prasad Calyam, and Khaza Anuarul Hoque. 2020. Impact of false data injection attacks on deep learning enabled predictive analytics. In *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 1–7.

[33] Manar Mohamed, Babins Shrestha, and Nitesh Saxena. 2016. Smashed: Sniffing and manipulating android sensor data for offensive purposes. *IEEE Transactions on Information Forensics and Security* 12, 4 (2016), 901–913.

[34] NJ Morrish, S-L Wang, LK Stevens, JH Fuller, and H Keen. 2001. Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* 44, 2 (2001), S14–S21.

[35] Arsalan Mosenia, Susmita Sur-Kolay, Anand Raghunathan, and Niraj K. Jha. 2017. CABA: Continuous Authentication Based on BioAura. *IEEE Trans. Comput.* 66, 5 (2017), 759–772. https://doi.org/10.1109/TC.2016.2622262

[36] Vahram Mouradian, Armen Poghosyan, and Levon Hovhannisyan. 2014. Continuous wearable health monitoring using novel PPG optical sensor and device. In *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 120–123.

[37] Abdulmajid Murad and Jae-Young Pyun. 2017. Deep recurrent neural networks for human activity recognition. *Sensors* 17, 11 (2017), 2556.

[38] Pratik Musale, Duin Baek, Nuwan Werellagama, Simon S Woo, and Bong Jun Choi. 2019. You walk, we authenticate: lightweight seamless authentication based on gait in wearable IoT systems. *IEEE Access* 7 (2019), 37883–37895.

[39] Tempestt Neal and Damon Woodard. 2019. Mobile Biometrics, Replay Attacks, and Behavior Profiling: An Empirical Analysis of Impostor Detection. In *2019 International Conference on Biometrics (ICB)*. IEEE, 1–8.

[40] Maryem Neyja, Shahid Mumtaz, Kazi Mohammed Saidul Huq, Sherif Adeshina Busari, Jonathan Rodriguez, and Zhenyu Zhou. 2017. An IoT-based e-health monitoring system using ECG signal. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 1–6.

[41] Keyurkumar Patel, Hu Han, Anil K Jain, and Greg Ott. 2015. Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks. In *2015 International Conference on Biometrics (ICB)*. IEEE, 98–105.

[42] Amith K. Belman; Li Wang; Sundaraja S. Iyengar; Pawel Sniatala; Robert Wright; Robert Dora; Jacob Baldwin; Zhanpeng Jin; Vir V. Phoha. 2019. SU-AIS BB-MAS (Syracuse University and Assured Information Security - Behavioral Biometrics Multi-device and multi-Activity data from Same users) Dataset. https://doi.org/10.21227/rpaz-0h66

[43] Hongyi Pu, Liang He, Chengcheng Zhao, David KY Yau, Peng Cheng, and Jiming Chen. 2020. Detecting replay attacks against industrial robots via power fingerprinting. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 285–297.

[44] Mahmudur Rahman, Bogdan Carbunar, and Umut Topkara. 2014. SensCrypt: A secure protocol for managing low power fitness trackers. In *2014 IEEE 22nd International Conference on Network Protocols*. IEEE, 191–196.

[45] Mahmudur Rahman, Bogdan Carbunar, and Umut Topkara. 2015. Secure management of low power fitness trackers. *IEEE Transactions on Mobile Computing* 15, 2 (2015), 447–459.

[46] Mohammad Naim Rastgoo, Bahareh Nakisa, Frederic Maire, Andry Rakotonirainy, and Vinod Chandran. 2019. Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications* 138 (2019), 112793.

[47] Gangireddy Narendra Kumar Reddy, M Sabarimalai Manikandan, and NVL Narasimha Murty. 2020. On-device integrated ppg quality assessment and sensor disconnection/saturation detection system for IoT health monitoring. *IEEE Transactions on Instrumentation and Measurement* 69, 9 (2020), 6351–6361.

[48] Gojka Roglic, Nigel Unwin, Peter H Bennett, Colin Mathers, Jaakko Tuomilehto, Satyajit Nag, Vincent Connolly, and Hilary King. 2005. The burden of mortality attributable to diabetes: realistic estimates for the year 2000. *Diabetes care* 28, 9 (2005), 2130–2135.

[49] Mike Ryan. 2013. Bluetooth: With low energy comes low security. In *7th {USENIX} Workshop on Offensive Technologies ({WOOT} 13)*.

[50] Prasan Kumar Sahoo, Hiren Kumar Thakkar, Wen-Yen Lin, Po-Cheng Chang, and Ming-Yih Lee. 2018. On the design of an efficient cardiac health monitoring system through combined analysis of ecg and scg signals. *Sensors* 18, 2 (2018), 379.

[51] Virginia Sandulescu, Sally Andrews, David Ellis, Nicola Bellotto, and Oscar Martinez Mozos. 2015. Stress detection using wearable physiological sensors. In *International work-conference on the interplay between natural and artificial computation*. Springer, 526–532.

[52] A. Sarkar, A. L. Abbott, and Z. Doerzaph. 2016. Biometric authentication using photoplethysmography signals. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. 1–7. https://doi.org/10.1109/BTAS.2016.7791193

[53] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 400–408.

[54] Jiacheng Shang and Jie Wu. 2019. A usable authentication system using wrist-worn photoplethysmography sensors on smartwatches. In *2019 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 1–9.

[55] Pekka Siirtola, Ella Peltonen, Heli Koskimäki, Henna Mönttinen, Juha Röning, and Susanna Pirttikangas. 2019. Wrist-worn Wearable Sensors to Understand Insides of the Human Body: Data Quality and Quantity. In *The 5th ACM Workshop on Wearable Systems and Applications*. 17–21.

[56] Amit Kumar Sikder, Hidayet Aksu, and A Selcuk Uluagac. 2017. 6thsense: A context-aware sensor-based attack detector for smart devices. In *26th {USENIX} Security Symposium ({USENIX} Security 17)*. 397–414.

[57] Amit Kumar Sikder, Giuseppe Petracca, Hidayet Aksu, Trent Jaeger, and A Selcuk Uluagac. 2018. A survey on sensor-based threats to internet-of-things (iot) devices and applications. *arXiv preprint arXiv:1802.02041* (2018).

[58] Jesús Solano, Christian Lopez, Esteban Rivera, Alejandra Castelblanco, Lizzy Tengana, and Martin Ochoa. 2020. SCRAP: Synthetically Composed Replay Attacks vs. Adversarial Machine Learning Attacks against Mouse-based Biometric Authentication. In *Proceedings of the 13th ACM Workshop on Artificial Intelligence and Security*. 37–47.

[59] Steven Taylor and Dana S Thordarson. 2002. Behavioural treatment of post-traumatic stress disorder associated with recovered memories. *Cognitive Behaviour Therapy* 31, 1 (2002), 8–17.

[60] Timothy Trippel, Ofir Weisse, Wenyuan Xu, Peter Honeyman, and Kevin Fu. 2017. WALNUT: Waging doubt on the integrity of MEMS accelerometers with acoustic injection attacks. In *2017 IEEE European symposium on security and privacy (EuroS&P)*. IEEE, 3–18.

[61] Chen Wang, Xiaonan Guo, Yan Wang, Yingying Chen, and Bo Liu. 2016. Friend or foe? Your wearable devices reveal your personal pin. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. 189–200.

[62] Jiliang Wang, Feng Hu, Ye Zhou, Yunhao Liu, Hanyi Zhang, and Zhe Liu. 2020. BlueDoor: breaking the secure information flow via BLE vulnerability. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*. 286–298.

[63] L Wang, L Li, X Zhou, S Pandya, and O Baser. 2016. A real-world evaluation of the clinical and economic burden of united states veteran patients with post-traumatic stress disorder. *Value in Health* 19, 7 (2016), A524.

[64] G. M. Weiss, K. Yoneda, and T. Hayajneh. 2019. Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living. *IEEE Access* 7 (2019), 133190–133202.

[65] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. 2017. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web*. 351–360.

[66] Ahmed Zekry, Ahmed Sayed, Mohamed Moussa, and Mohamed Elhabiby. 2021. Anomaly Detection using IoT Sensor-Assisted ConvLSTM Models for Connected Vehicles. In *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. IEEE, 1–6.

[67] Bo Zhao, Xinwei Sun, Yanwei Fu, Yuan Yao, and Yizhou Wang. 2018. Msplit lbi: Realizing feature selection and dense estimation simultaneously in few-shot and zero-shot learning. In *International conference on machine learning*. PMLR, 5912–5921.

[68] Wei Zheng, Le Yang, Robert J Genco, Jean Wactawski-Wende, Michael Buck, and Yijun Sun. 2019. SENSE: Siamese neural network for sequence embedding and alignment-free comparison. *Bioinformatics* 35, 11 (2019), 1820–1828.

A APPENDIX: PARAMETERS OF NEURAL NETWORK MODELS

This section introduces the structures of models used for each dataset in this study. The integration of convolution (CNN) and Long Short-Term Memory (LSTM) network is called CNN-LSTM model. Siamese neural network is abbreviated as SNN.

A.1 BB-MAS Dataset

Authentication Model. Each sub-network of the Siamese authentication model has two convolutional layers [256, 128] with (3,5) and (1,8) dimensional kernels. The output from convolutional layers is passed to three linear layers [1024, 512, 256] to generate the embedding. The L_1 distance between the two embeddings is computed and passed to a linear layer with Sigmoid function to calculate the similarity score.

CNN-LSTM FDIA Detection Model. The CNN implementation uses two convolutional layers [128, 64], and the LSTM implementation has one layer [256]. Three decision generation linear layers [128, 32, 2] take the LSTM ouput as input and infer the FDIA.

Siamese-MIL FDIA Detection Model. Siamese-MIL leverages a SNN to measure the similarity between signal segments. Each of the sub-networks of the SNN implementation has two convolution layers [256, 128], with (3, 3) and (1, 3) dimensional kernels, and followed by two linear layers [512, 256] and leaky relu activation function. The L_1 distance between the two generated encodings (1×256 dimensional) will be computed and passed to a dense layer with Sigmoid function to calculate the similarity score.

A.2 WISDM Dataset

Authentication Model. Each sub-network of the Siamese authentication model has two convolutional layers [128, 64] with (3,4) and (1,3) dimensional kernels. The output from convolutional layers is passed to two linear layers [512, 256] to generate the embedding. The L_1 distance between the two embeddings is computed and passed to a linear layer with Sigmoid function to calculate the similarity score.

Activity Detection Model. The activity detection models use CNN-LSTM structure. The CNN implementation consists of two convolutional layers [128, 64] with (3, 3) and (1, 3) dimensional kernels, and the LSTM has one layer [256] and is followed by three linear layers [64, 32, 2]. The decision is made through a Softmax layer.

CNN-LSTM FDIA Detection Model. The CNN-LSTM comprises two convolutional layers [128, 64], one layer [256] LSTM and three decision generation linear layers [64, 32, 2].

Siamese-MIL FDIA Detection Model. Each of the sub-networks of the SNN implementation has two convolutional layers with less kernels [128, 128], with (3, 2) and (1, 3) dimensional kernels. Two linear layers [512, 256] takes the output from the convolutional layers to generate encoding.

A.3 WESAD Dataset

Stress Detection Model. The stress detection model is implemented using CNN-LSTM framework. The CNN implementation consists of two convolutional layers [128, 128], with (1, 4) dimensional kernels, and the LSTM has one layer [256] and is followed by three linear layers [128, 32, 2] to make the inference.

CNN-LSTM FDIA Detection Model. The CNN-LSTM comprises two convolutional layers [128, 64], one layer [256] LSTM and three decision generation linear layers [128, 32, 2].

Siamese-MIL FDIA Detection Model. Each of the sub-networks of the Siamese-MIL's SNN model contains two convolutional layers [256, 128], with (1, 3) and (1, 4) dimensional kernels, and two linear layers [512, 256]. L_1 distance between two encodings is computed and then similarity score is calculated.

B APPENDIX: ABLATION STUDY OF MIL-BAG CONFIGURATIONS

We evaluate different MIL-bag representation parameters (Section 6.2.1) - MIL instance (i.e., small segment) size V, and overlap rate R. We evaluated V as the 10%, 20% and 30% of the input window length W, and R as 25% or 50%. The evaluations are performed on smoothed gait-ACC FDIA data from the BB-MAS dataset.

The evaluation on different instance sizes V and overlap rate R are shown in table 10. According to the evaluation results, the 50% overlap rate gives better performance, and 1-s and 2-s instance size V with $R = 50\%$ provides similarly high performance. When $R = 50\%$ and $V = 1$-s, there are 171 instance-pairs in a MIL bag representation, compared to 36 instance-pairs in a MIL bag when $R = 50\%$ and $V = 2$-s. That means, with $R = 50\%$ and $V = 1$-s hyper-parameters, the Siamese-MIL performs 4.75 times more computations (i.e., SNN instance-pairs comparisons) to generate similar attack detection performance (F1-score), compared to the $R = 50\%$ and $V = 2$-s hyper-parameter configuration Siamese-MIL. Since the Siamese-MIL attack detection approach needs to be real-time executable on the computationally constraint smart devices, we use $R = 50\%$ and $V = W * 20\%$ as the optimal hyper-parameter configuration.

Table 10. Average F1-scores with Siamese-MIL-bag parameter combinations on smoothed gait-ACC FDIA data.

V R	1-s	2-s	3-s
25%	89.96%	87.79%	84.18%
50%	90.45%	90.35%	89.07%

C APPENDIX: MITIGATION FOR THE ATTACK: AN EXTENSION OF AUTHENTICATION SYSTEM

Section 7 and 8 show that the developed Siamese-MIL method effectively detects FDIA samples in different scenarios. This section will discuss the mitigation strategy for the FDIA.

As discussed in Section 2, sensory signals are first verified by a continuous authentication system (CAS) before reaching any human-centric sensing application. Therefore, we can use the Siamese-MIL as a CAS extension and verify a signal's authenticity by voting. *If any of them considers a signal is an attack sample, it will be rejected.* To evaluate such integration's performance, we fuse the authentication systems developed in Section 5.1 with the Siamese-MIL models for BB-MAS and WISDM dataset. Evaluation of the mitigation strategy on different FDIA variations is shown in table 11a and 11b.

Table 11. Attack detection rate (TNR) of mitigation strategy on FDIA variations from BB-MAS and WISDM dataset

(a) Performance on gait-ACC FDIA samples from BB-MAS dataset

FSR	smoothed	non-smoothed
10%	82.72%	77.25%
20%	87.50%	85.94%
30%	90.34%	87.89%
40%	91.51%	87.21%
50%	91.50%	87.31%
60%	92.39%	90.53%
70%	92.00%	91.51%
80%	92.09%	90.63%
90%	88.84%	88.58%

(b) Performance on motion-ACC FDIA samples from WISDM dataset

FSR	A/B		A/C	
	smoothed	non-smoothed	smoothed	non-smoothed
10%	90.28%	91.10%	86.32%	87.15%
20%	95.86%	96.63%	91.48%	91.80%
30%	97.77%	98.14%	92.48%	93.07%
40%	98.32%	98.71%	93.19%	93.35%
50%	98.74%	98.83%	94.50%	94.13%
60%	99.01%	99.22%	94.78%	95.13%
70%	99.11%	99.11%	95.13%	94.78%
80%	99.29%	99.32%	94.94%	94.62%
90%	99.27%	99.34%	95.26%	94.23%

Compared to table 1a, 6a, 1b and 7, this approach mitigates the lower performance of Siamese-MIL on high FSR (90%) and the lower performance of authentication system on lower FSR (10-60%), achieving consistently
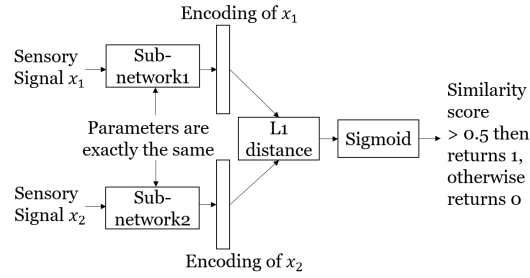
Fig. 2. Siamese neural network structure.

high TNR on all FDIA variations. Furthermore, similar to previous evaluations, the integrated system performed similarly to smoothed and non-smoothed attack samples. Therefore, with the integration of Siamese-MIL, the CAS is highly robust against all variations of the FDIA.

D APPENDIX: BACKGROUND DISCUSSION ON SIAMESE NEURAL NETWORK (SNN)

Siamese Neural Network (SNN) [7, 68] employs a unique structure to naturally compare a pair of inputs in terms of their semantic similarity or dissimilarity. Two identical sub-networks generate the embedding representations of the respective input instances. The sub-networks are joined together by a distance function that computes how close or far-apart the input pairs are in the embedding space. SNNs have been widely used in meta-learning [15, 67] due to their powerful discriminative capability that generalizes not just to new data but to entirely new classes of data from unknown distributions. Hence, SNNs are suitable for human-centric sensing attack detection tasks, where very few or no example of the target user's data is available.

Each individual has a unique behavioral or physiological pattern that is conveyed to their sensory signal information. Sensory signal from each individual can be categorized as a single class. This paper leverages SNN structure to to distinguish input sensory signal samples of different individuals (i.e., in FDIA detection). As shown in Figure 2, the sub-networks take two sensory-signal (i.e., ACC, BVP) input samples (x_1 and x_2) and generate encoding representations. The L_1 distance between the two encodings is computed, and the similarity score is obtained by passing the distance through a dense linear layer with a Sigmoid unit. A pair of signals is considered from the same individual if the similarity score > 0.5.