2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE) | 979-8-3503-4501-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/CHASE60773.2024.00019

Systematically Assessing the Security Risks of AI/ML-enabled Connected Healthcare Systems

Mohammed Elnawawy[†], Mohammadreza Hallajiyan[†], Gargi Mitra[†], Shahrear Iqbal[*], Karthik Pattabiraman[†]
[†]University of British Columbia, [*]National Research Council Canada
Email: {mnawawy, hallaj, gargi}@ece.ubc.ca, shahrear.iqbal@nrc-cnrc.gc.ca, karthikp@ece.ubc.ca

***Abstract*—The adoption of machine-learning-enabled systems in the healthcare domain is on the rise. While the use of ML in healthcare has several benefits, it also expands the threat surface of medical systems. We show that the use of ML in medical systems, particularly connected systems that involve interfacing the ML engine with multiple peripheral devices, has security risks that might cause life-threatening damage to a patient's health in case of adversarial interventions. These new risks arise due to security vulnerabilities in the peripheral devices and communication channels. We present a case study where we demonstrate an attack on an ML-enabled blood glucose monitoring system by introducing adversarial data points during inference. We show that an adversary can achieve this by exploiting a known vulnerability in the Bluetooth communication channel connecting the glucose meter with the ML-enabled app. We further show that state-of-the-art risk assessment techniques are not adequate for identifying and assessing these new risks. Our study highlights the need for novel risk analysis methods for analyzing the security of AI-enabled connected health devices.**

***Index Terms*—Machine learning, FDA, medical system security, risk analysis, multi-vendor systems.**

I. Introduction

The use of Artificial Intelligence (AI), especially Machine Learning (ML) techniques, is becoming increasingly popular in the medical field. As of October 2022, the U.S. Food and Drug Administration (FDA) has approved 521 ML-enabled devices across 15 different medical disciplines (e.g., Cardiology, Ophthalmology, and Gastroenterology) [1]. However, the use of ML has expanded the threat surface of medical systems [2]–[16] making them more vulnerable to cyberattacks.

ML-enabled medical devices are used for performing critical activities such as remote patient monitoring, controlling surgical equipment, automatic drug administration, and preliminary/advanced disease diagnosis – tasks that require high accuracy and reliability [1]. If an adversary compromises such a device, it can force the ML engine to make incorrect predictions or decisions, which can have catastrophic consequences, such as wrong treatment leading to health complications.

An adversary can force an ML engine to generate incorrect predictions or decisions by injecting carefully crafted malicious data points either during training or inference. Preventing such attacks in ML-enabled medical devices is challenging. These ML-enabled devices are typically interconnected with other peripheral sensor devices that collect physiological data of patients, which are then processed by the ML engine. Therefore, it is not enough to secure the ML-enabled device, since adversaries can exploit vulnerabilities in the peripheral devices to inject malicious data points in the ML engine.

To protect the end-to-end system, one must systematically identify and assess the security risks [1] of the overall system due to vulnerabilities in peripheral devices. *To the best of our knowledge, there is no systematic technique for identifying and assessing the end-to-end risks of ML-enabled medical systems.*

Identification of risks in ML-enabled connected medical systems has two challenges. *First,* at deployment, the ML-enabled device is interfaced with several other peripheral devices, each of which may be manufactured by a different company. For instance, a user of the ML-enabled blood glucose monitoring app Dreamed Advisor Pro [17], needs to install the app on a smartphone, and then connect to it a smartwatch, a glucose meter, and an insulin pump, all of which would be manufactured by different companies, and may have their own security vulnerabilities. *Second,* each app user may use peripheral devices from different sets of manufacturers, leading to diverse vulnerabilities among different users of the same app. For instance, one user of the app might use a vulnerable smartphone, while another user might use a vulnerable glucose meter.

Furthermore, it is also challenging to assess the *severity* of these risks. This is because the severity of a risk posed by a vulnerable peripheral device might differ when assessed in the context of the individual device (in-silo assessment) versus when assessed in the context of the entire system. For instance, consider a user who connects the Dreamed Advisor Pro app to a glucose meter with a write-access vulnerability, and a smartphone with a read-access vulnerability. When assessed separately, the glucose meter would have a higher perceived risk than the smartphone. However, for an adversary who wants to inject adversarial glucose meter readings into the app, being able to read data from the smartphone (e.g., meal timings, latest insulin dose, carbohydrates taken) might be useful for crafting malicious data points that adhere to physiological constraints. Adhering to the physiological constraints is important for the adversary to get the malicious data points accepted as valid inputs by the ML engine. Therefore, we need to holistically consider the risks from the interplay of vulnerabilities in peripheral devices.

[1] We define risk as the probability of a security vulnerability getting exploited, and its potential impact or loss.

2832-2975/24/$31.00 ©2024 IEEE
DOI 10.1109/CHASE60773.2024.00019

In this paper, we perform a systematic analysis to highlight the security risks of end-to-end ML-enabled connected medical systems. Our analysis consists of three steps. *First*, we conduct a systematic exploration of the FDA-approved ML-enabled medical devices to understand the ML techniques that they use, and the damage that can be caused to a patient if the ML technique mispredicts their case. *Second*, we conduct an extensive review to *identify* possible ways in which adversaries can inject malicious data points into an ML-enabled medical device at deployment. This involves a cross-domain analysis, where we map known attacks on ML algorithms with known vulnerabilities in peripheral devices that would make the attacks practical. *Finally*, we perform a critical evaluation of state-of-the-art risk assessment frameworks used by the ML-enabled medical device manufacturing companies today. We identify the loopholes in these risk assessment strategies that might make manufacturers miss the risks arising due to vulnerabilities in connected peripheral devices.

Contributions. The main contributions of this paper are:

1) We perform a systematic cross-domain security analysis of commercial ML-enabled medical devices approved by the FDA (Section III), to highlight the security risks of connected health devices.
2) We then perform a case study on a realistic ML-enabled blood glucose management system (BGMS) (Section IV) to demonstrate an attack on the system where the adversary compromises a communication link in the system.
3) Finally, we perform an evaluation of state-of-the-art risk assessment techniques (Section V). We find that they are inadequate in identifying and analyzing the severity of security risks in ML-enabled medical systems, particularly the risks posed to the ML engine by vulnerable peripheral devices. We also highlight directions for improvement.

II. Motivation and Background

We highlight security risks in AI/ML-enabled medical systems due to vulnerabilities in their connected peripheral components, using the example of the BGMS. Following this, we demonstrate the generalizability of identified risks to any connected ML-enabled medical system.

A. Blood Glucose Management: Background

Diabetes is a chronic health condition that hinders the body's natural insulin production capability, leading to elevated blood glucose levels. It has detrimental effects on a patient's health, and sudden spikes or drops in blood glucose can be life-threatening. Blood glucose levels can be divided into three ranges, hypoglycemic ($<$ 70-80 mg/dL), normal (80 mg/dl - 125 mg/dl), and hyperglycemic ($>$ 125 mg/dL while fasting and $>$ 180 mg/dL two hours postprandial) [18], [19]. A consistently hyperglycemic patient (i.e., diabetic) requires insulin injections to normalize their glucose levels. In contrast, a hypoglycemic patient does not need insulin injections.

BGMS apps help diabetic patients monitor their blood glucose levels and administer insulin bolus whenever the glucose levels begin to rise at an abnormal rate. The patient can either manually inject the insulin, or use an automated insulin pump connected to and controlled by the BGMS app. It is crucial to calculate the insulin bolus dose accurately — an overdose can lead to a sharp drop in blood glucose, while an insufficient dose may not bring it down to the normal range.

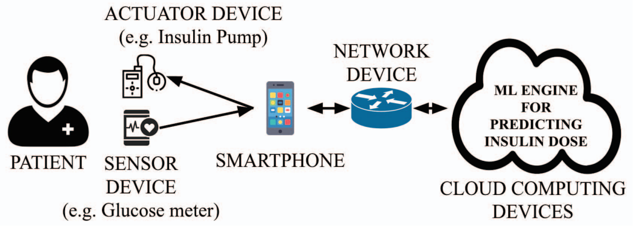

Fig. 1: A blood glucose management system that uses ML on the cloud, and interfaces with multiple peripheral devices.

B. An ML-enabled Blood Glucose Management System

We consider a commercial FDA-approved ML-enabled BGMS app, the Dreamed Advisor Pro [17]. This app assists diabetic patients in maintaining normal blood glucose levels by periodically recommending insulin bolus doses, personalized meal plans, and physical activities. However, since the specific ML technique used by the Dreamed Advisor Pro app is not publicly disclosed, we instead use a well-known, public ML-based blood glucose prediction technique [20].

BGMS apps suggest insulin doses based on the patient's predicted blood glucose level in the near future (next 30 or 60 mins) [20]. This prediction is done by an ML engine running at the back-end of the BGMS app, using recent physiological values of the patient, such as blood glucose measurements, insulin doses taken, meal timings, carbohydrate intake, etc. These values are either entered manually into the app by the patient or read from other peripheral sensor devices (e.g., continuous glucose monitors and smartwatches) connected to the patient's body. Besides sensor devices, the app can interface with actuator devices (e.g., insulin pumps) to execute actions suggested by the ML-enabled app. Together, the ML-enabled app and connected peripheral devices form the BGMS.

Figure 1 shows an end-to-end schematic representation of an ML-enabled BGMS. The glucose monitor records the blood glucose levels of the patient at regular time intervals and transmits them to the app over a Bluetooth communication channel. The app uses these values for data visualization and also sends them to a cloud server over the Internet for storage and processing by the AI/ML engine. Finally, the predicted insulin dose is either displayed on the app or sent to the automated insulin pump attached to the patient's body.

Security Risks. We consider a scenario where an adversary intends to cause harm to a targeted user of the BGMS app by forcing the BGMS to inject a high dose of insulin into the user's body when it is not supposed to, or vice versa. We further assume that the insulin pump used by the target is secure against known vulnerabilities [21] and that the ML engine runs on a secure cloud server. Under such circumstances, the adversary can still force the ML engine to mispredict the insulin dose by injecting carefully crafted adversarial data points into it via the peripheral devices. For instance, the adversary can modify the blood glucose values sent to the ML engine [10], by exploiting vulnerabilities in the third-party glucose monitor interfaced with the BGMS app.

The Dreamed Advisor Pro app is compatible with all non-continuous glucose meters with regulatory approval and nine different models of continuous glucose monitors manufactured by different companies [22]. Commercial glucose meters have been known to have firmware vulnerabilities [23] that would allow an adversary to change the blood glucose level readings that are sent by the glucose meter to the BGMS app. Alternatively, an adversary could also exploit known vulnerabilities in the Bluetooth communication protocol [21] to launch a man-in-the-middle attack and change the blood glucose level readings. Furthermore, these glucose monitors are also susceptible to physical attacks that can be carried out with electromagnetic radiation [24]. Additionally, vulnerabilities in devices and communication links in the end-to-end data processing pipeline, such as mobile devices and routers, can be exploited to manipulate blood glucose measurements.

C. Unique Security Risks in AI/ML-enabled Medical Systems

The security risks discussed in the context of the BGMS app apply to any ML-enabled medical device connected to multiple peripheral devices (often manufactured by other companies) at deployment. A security breach in any of these peripheral devices could enable an adversary to manipulate data sent to the ML engine, resulting in mispredictions regarding the patient's condition. These mispredictions pose a direct threat to the patient's health.

Manufacturers of ML-enabled devices face two challenges in anticipating and assessing the aforementioned security risks during design and manufacturing. *First,* these devices are built to be compatible with a diverse range of peripheral devices for the operational convenience of the consumers. This makes it difficult for the manufacturer to predict what peripheral devices the consumer would connect with the ML-enabled device at deployment, and what vulnerabilities those devices might have. The interplay of different vulnerabilities would enable an adversary to perform different types of attacks on the ML engine, which makes it challenging to analyze the risk via the in-silo testing performed today. *Second,* while ML-enabled connected systems are used in many domains such as smart homes and industrial control systems, performing risk analysis of medical systems is more difficult as physiological data is much more complex and varies widely across individuals [25]. Consequently, the impact of manipulating physiological data might be different for different patients. Manufacturers typically prioritize accuracy and failures in non-adversarial scenarios, but a comprehensive end-to-end security risk analysis should consider the impact of adversarial inputs on different patients and the feasibility of attacks.

For instance, in the BGMS attack in Section II-B, the adversary attempts to generate inaccurate insulin dosage predictions by introducing adversarial inputs to the ML engine. Most known attacks on ML engines [2]–[16] require the adversary to observe and manipulate at least a subset of input sensor values to alter the predicted insulin dose. The adversary's efficiency in crafting adversarial inputs increases with greater observability into different sensor values. In Section II-B, we assumed the adversary could only manipulate glucose level readings through vulnerabilities in the glucose meter or the Bluetooth link. However, if the adversary can observe commands sent to the insulin pump, they could craft adversarial glucose meter readings to yield a higher success rate with equal or fewer perturbations. While unauthorized read access to an individual insulin pump poses a low-level risk, in the context of the entire BGMS, it becomes a high-level risk.

III. Attacks on ML-Based Systems and Their Relevance in the Healthcare Domain

Motivated by the BGMS example in Section II, we systematically investigate a subset of FDA-approved AI/ML-enabled medical devices/software [1] to identify potential security risks at deployment. We perform this investigation in two steps. *First,* we identify the ML techniques used by each of the devices/software systems. We survey existing work in the domain of AI/ML security to understand what types of attacks may target these techniques (Section III-A). *Next,* for each of these devices/software, we examine the practicality of the attack scenarios identified in the previous step (SIII-B).

Selecting medical systems for our investigation. As of December 2022, the U.S. FDA has approved 521 ML-enabled medical devices [2] across 15 different physiological panels. However, there is no automated risk analysis technique today, and analyzing all the 521 devices manually would be arduous and time-consuming. Therefore, we selected a subset of these devices for manual analysis. We used the following selection criteria to ensure a fair representation of the set of devices.

1) We select at least one device from each physiological panel to study if risks due to vulnerabilities in peripheral devices are common across all medical domains;
2) Within the same physiological panel, we select devices that perform different types of diagnosis or treatments, to ensure coverage across different medical activities.
3) We select an equal number of two types of ML-enabled devices - software that can be installed on the consumer's pre-existing device, and software that is sold bundled with proprietary hardware. This would help us understand if one of these is more secure than the other;
4) We select devices that are used in hospitals and clinics under medical supervision, as well as devices that are used by patients at home without medical supervision. This would help us understand if the environment in which the medical device is deployed affects its security.

Additionally, to ensure that sufficient information is available for each selected device, e.g., the ML algorithm used,

[2] As per the terminology used by the FDA website, the term 'device' refers to both physical devices as well as software solutions.

the type of data processed, etc, we select 20 different devices across 13 of the 15 physiological panels, as shown in Table I. Unfortunately, due to insufficient information, we could not select any device from the Dental and Hematology panels.

A. Known Attacks on ML Algorithms Used by FDA-Approved Medical Devices

Table I presents our study of the ML algorithms used by the devices we selected for our evaluation. Our goal is to understand if there are known attacks against these ML algorithms that can be used by adversaries to make these ML engines mispredict the outcome. We also examine the types of tasks for which these ML algorithms are used, and the worst-case consequences of misprediction by the ML engine.

Survey Process. We performed the following steps to identify known attacks on the ML algorithms used by the devices.

Step 1. Identifying the ML algorithm and input features used by the device: We analyze device information available in the Premarket Notification summaries submitted by manufacturers to the FDA during the approval process. These summaries are available on the FDA website [1], and contain crucial details such as the ML algorithm and the type of data processed. However, for some devices, like GI Genius, the summaries lack specific information about the ML algorithm. In such cases, we explore the manufacturer's website and, if even that is inadequate, we estimate the ML algorithm based on the device's task and processed inputs. We look for known ML algorithms that perform the same task using similar input features. For example, for GI Genius, we found a relevant paper [26] that performs gastrointestinal lesion detection (the same task performed by GI Genius) with a high accuracy using Convolutional Neural Networks (CNN).

Step 2. Identifying known attacks on the ML algorithm: We search for known attacks in the literature that target the ML algorithms identified in Step 1. We focus on attacks described in research papers published in both conferences and journals. The discovery of such attacks does not definitively establish the vulnerability of the ML engine in the device under consideration. Rather, it identifies potential risks, emphasizing the need for systematic risk identification and mitigation.

Step 3. Estimating worst-case impact of mispredictions: We estimate the worst-case impact of mispredictions by these devices from our understanding of the device functionality and description provided by the manufacturer, either in the device summary or on their website. We deem the misprediction to be potentially fatal if the device is used for the treatment or diagnosis of a patient in medical emergencies, and a medical expert would not have enough time to assess the correctness of the device output. For instance, the NuVasive Pulse System is used by surgeons during spinal surgeries for continuously monitoring the neurophysiological status of the patient, and a misprediction by the device would be potentially fatal.

Insights. We obtained the following insights from the information that we gathered using the aforementioned process.

1) We observed that the majority of the ML algorithms are vulnerable to inference-time attacks, with a few susceptible to training-time attacks. Both of these pose major health risks for patients. However, executing inference-time attacks is comparatively easier for adversaries as they demand fewer adversarial inputs than training-time attacks. Most of the devices prone to training-time attacks are deployed in hospitals or diagnostic centers, where a shared set of peripheral devices is used to collect data from multiple patients. If adversaries successfully compromise these peripheral devices over an extended period, they can manipulate sufficient patient data to poison the training dataset [3]. This would affect a large number of patients. Examples include the Deep Learning Image Recognition Software, and the Oxehealth Vital Signs monitor.
2) Even when the devices are operated by medical practitioners, detecting a misprediction might be challenging for two reasons. *First,* physiological data exhibit significant variance even among patients with the same medical condition, owing to diverse underlying health conditions and demographic factors [27]. *Second,* many devices are used for infrequently performed diagnoses/medical procedures, or are used only for medical emergencies. Under such circumstances, the lack of the particular patient's historical physiological information makes it challenging for the medical practitioner to detect an anomaly. Examples are Cardiologs ECG Analysis Platform, GI Genius, ABMD Software, and NuVasive Pulse System.
3) Some devices (e.g., the One Drop Blood Glucose Monitoring System), are used by patients at home without continuous medical supervision. Detecting a misprediction from such devices would be much more challenging than devices that are directly operated by medical practitioners.
4) Many of these devices are used in clinics and hospitals for disease diagnosis, treatment, and patient monitoring. However, a few are used by patients at home. Hospitals and clinics would typically have a higher security budget than individual patients at home, and hence have better security. Consequently, designing a one-size-fits-all security solution for ML-enabled medical systems is challenging. Therefore, while designing security solutions for medical devices, the implementation effort and cost should be considered.
5) Some of the ML-enabled softwares are sold bundled with proprietary hardware (i.e., software-in-medical-device), while some can be installed by the user on any general-purpose computer (i.e., software-as-medical-device). The latter have a broader threat surface due to diverse combinations of hardware, software, and Operating System (OS) vulnerabilities across various general-purpose computer models, making the assessment of risk severity challenging.

B. Analyzing the Functionality and Vulnerability Landscape of FDA-approved (AI/ML)-Enabled Medical Devices

We investigate how an adversary can exploit the vulnerabilities identified in §III-A. Table I shows that all the identified ML attacks involve manipulating inputs during training

[3] Many systems undergo periodic re-training on recent physiological data.

Sl. No.	Device Name [1]	Physiological Panel	Device Functionality	User	Type of ML algo used	Type of data processed	Known attacks (Attack type)	Potential impact of misprediction
1	CardioLogs ECG Analysis Platform†	Cardiovascular	Cardiac arrhythmia detector	Medical practitioners	Deep Neural Network (DNN)	Image	Chen et al. [2] Ⓘ	Wrong treatment (Fatal)
2	Oxehealth Vital Signs†	Cardiovascular	Camera-based monitor for heart, pulse, and respiratory rate	Medical practitioners	Hybrid convolutional Long short term memory networks (LSTM)	Video	Albattah et al. [3] (T,I)	Wrong treatment
3	GI Genius‡	Gastroenterology/ Urology	Gastro-intestinal lesion detection	Medical practitioners	Convolutional neural networks (CNN) *	Video	Amin et al. [28] Ⓣ	Wrong diagnosis
4	SOZO‡	Gastroenterology/ Urology	Body fluid analyzer for assessing protein-calorie malnutrition	Medical practitioners	CNN *	Numeric	Byra et al. [29] Ⓘ	Wrong diagnosis
5	WellDoc BlueStar†	General hospital	Diabetes management	Medical practitioners, patients	Darknet-53 CNN	Numeric	Lal et al. [4] Ⓘ	Wrong diagnosis
6	d-Nav System†	General hospital	Insulin dose predictor	Medical practitioners, patients	Multi-layer perception (MLP) and LSTM	Numeric	Zhou et al. [30] Ⓘ	Wrong treatment (Fatal)
7	MBT-CA System‡	Microbiology	Spectometry	Medical practitioners	DNN *	Numeric	Meiseles et al. [5] Ⓘ	Wrong diagnosis (Fatal)
8	KIDScore D3†	Obstetrics & Gynaecology	Embryo image assessment	Medical practitioners	Decentralized federated learning	Image	Nguyen et al. [31] Ⓟ	Wrong diagnosis
9	NuVasive Pulse System‡	Orthopedic	Neurological monitoring	Medical practitioners	CNN *	Image	Joel et al. [6] Ⓘ	Mistake in surgery (Fatal)
10	ABMD Software†	Radiology	Bone densitometer	Medical practitioners	Inception-v3 and Densenet-121 *	Image	Bortsova et al. [7] Ⓘ	Wrong diagnosis
11	Deep Learning Image Reconstruction†	Radiology	X-ray reconstruction	Medical practitioners	ResNet-18	Image	Menon et al. [8] Ⓣ Paul et al. [32] Ⓘ	Wrong diagnosis
12	Air Next‡	Anesthesiology	Spirometer	Medical practitioners	CatBoost ResNet-50 *	Image	Vargas et al. [9] Ⓘ	Wrong diagnosis
13	One Drop Blood Glucose Monitoring System‡	Clinical Chemistry	Diabetes management	Patients	MLP	Numeric	Levy-Loboda et al. [10] Ⓘ	Wrong treatment (Fatal)
14	OTIS 2.1 and THiA Optical Coherence Tomography System‡	General and Plastic Surgery	Human tissue imaging	Medical practitioners	Support Vector Machines (SVM)	Image	Ma et al. [16] Ⓘ	Wrong diagnosis
15	EarliPoint System‡	Neurology	Diagnosis of Pediatric Autism Spectrum Disorder	Medical practitioners	Graph Neural Network (GNN)	Image	Chen et al. [11] Ⓣ	Wrong diagnosis
16	BrainScope TBI‡	Neurology	Brain injury assessment	Medical practitioners	CNN + Recurrent neural networks (RNN)	Numeric	Yu et al. [12] Ⓘ	Wrong treatment (Fatal)
17	IDx-DR v2.3†	Ophthalmic	Diabetic Retinopathy Detection	Medical practitioners	Federated learning	Image	Nielsen et al. [13] Ⓘ	Wrong diagnosis (loss of vision)
18	Iris Intelligent Retinal Imaging System†	Ophthalmic	Storage, management and display of retinal images	Medical practitioners	DNN	Image	Mangaokar et al. [14] Ⓘ	Wrong diagnosis (loss of vision)
19	Paige Prostate†	Pathology	Cancer diagnosis	Medical practitioners	CNN	Numeric	Hu et al. [15] Ⓣ	Wrong treatment (Fatal)
20	Tissue of Origin Test Kit‡	Pathology	Malignant Tumor diagnosis	Medical practitioners	SVM	Image	Ma et al. [16] Ⓘ	Wrong treatment (Fatal)

TABLE I: A study of different FDA-Approved ML-enabled medical devices and their security vulnerabilities
†: Software as medical device, ‡: Software in medical device, *: Best-guessed ML algorithm,
Ⓣ: Training-time attack, Ⓘ: Inference-time attack, Ⓟ: Privacy attack

or inference. For each ML-enabled device, we search for compatible peripheral devices and communication channels that would allow adversaries to introduce malicious data into the ML engine. We also assess the adequacy (or the lack thereof) of manufacturers' risk assessments, as mentioned in their Premarket Notifications, to determine their effectiveness in preventing such security risks. Table II presents this study.

Survey Process. To understand the vulnerability landscape of each device, we performed the following two steps.

Step 1. Identifying peripheral devices and communication media compatible with the ML-enabled device: We identify compatible peripheral sensor devices, communication media, and operating system from its Premarket Notification summary [1] and information on the manufacturer's website. One or more of these can be a potential point of attack.

Step 2. Identifying vulnerable peripherals that can be exploited for attacking the ML engine: For each potential attack point, we look for known attacks and vulnerabilities by searching research papers and vulnerability databases [33], [34]. We list at least one vulnerability that would allow an adversary to eavesdrop or inject malicious data into the ML engine, enabling them to execute the attacks identified in §III-A. This list of vulnerabilities is not comprehensive. We highlight at least one vulnerability to motivate the risk analysis technique.

Insights. We summarize the insights from this study below.

1) We found known vulnerabilities in the peripheral devices compatible with several ML-enabled devices. While most vulnerabilities affect only a small group of devices, a few vulnerabilities affect all devices of a certain type. For instance, the Conexus telemetry protocol vulnerability [35] only affects the ECG monitors from Medtronic. However, another attack [36] affects all infrared-sensitive cameras.
2) Some vulnerabilities (e.g., [35] for the Cardiologs ECG Analysis Platform, and [36] for Oxehealth Vital Signs) require the adversary to execute the attack locally as they have to be within the Bluetooth [35] communication range, or within the range of infrared light emission [36]. Such attacks can be executed by insiders or by breaching the physical security of the hospital or the patient's home.
3) Many of the vulnerabilities can be exploited remotely (e.g., [37] for Oxehealth and [38] for the IDx-DR) over the Internet. Since connectivity to the Internet is mandatory for these devices, preventing remote attacks is challenging.
4) In some cases, identifying the attack path is challenging. For example, the IDx-DR software relies on inputs from the Topcon NW200 Fundus camera. Although we found no known vulnerability in the camera, it comes bundled with a computer running Windows 7 by default, which has known vulnerabilities [38]. These Windows 7 vulnerabilities could enable an adversary to inject malicious inputs into the ML engine. While updates for Windows 7 may address such vulnerabilities, medical devices typically do not receive routine security updates. In a specific case [38], the vendor even decided not to release a patch, assuming most users would upgrade to Windows 10.
5) We did not find any known vulnerability in the peripheral devices for some systems (e.g., SOZO, WellDoc Bluestar, and Air Next). However, many of these systems use Bluetooth, Internet communications, and web services. Adversaries can exploit vulnerabilities [21] in these communication channels for injecting adversarial inputs.
6) We found that many of the ML-enabled device manufacturers (e.g., the NuVasive Pulse System) do not perform any security evaluation, and only focus on accuracy and safe operating conditions (e.g., protecting the devices from electrical hazards). Even the manufacturers who consider security, rarely consider the peripheral devices. For instance, the developers of the IDx-DR software evaluate the software for various security risks, but not its peripheral device, the Topcon NW200 Fundus Camera. However, security evaluation of the software alone is insufficient. This is because an adversary might execute the inference-time attack [13] shown in Table I by exploiting the vulnerability in the camera [38] to install malware that manipulates the images that are sent to the input of the ML engine.

IV. Case Study

We present a case study to demonstrate the security risks in the ML-enabled BGMS described in Section II-B. We show a practical attack on the BGMS in which the attacker exploits the vulnerabilities in connected devices to negatively affect the predictions of the ML-enabled decision-making component [4].

A. Attack Description

Adversarial Goal. The attacker aims to endanger a targeted patient's life by causing the ML model to misdiagnose the patient's condition, thereby leading to an incorrect insulin dose suggestion. While minor prediction errors are benign, a substantial error could have life-threatening consequences. In this case study, we consider an attacker aiming to make the model predict a high blood glucose level (hyperglycemia) when the patient actually has a low (hypoglycemia) or normal blood glucose level. If the attacker succeeds, the BGMS would erroneously recommend more insulin, causing the patient's glucose level to drop significantly below normal. The implications range from incorrect diagnosis (e.g., in the WellDoc BlueStar system) to incorrect treatment (e.g., in d-Nav and One Drop BGMS systems), potentially leading to fatal outcomes.

Adversarial Capabilities. We assume the attacker has reasonable and realistic capabilities, wherein they can only tamper with the CGM measurements. Manipulating the manually entered finger-based glucose readings, carbohydrate intake, and bolus dose is beyond the attacker's capabilities. However, the attacker can compromise the smartphone [48] to gain read-only access to these values once they have been gathered from external sensors or after being manually entered by the patients in the mobile app. The attacker is oblivious to the structure and parameters of the underlying ML model (black box attack),

[4] No human subjects were used in our experiments. Instead, we rely on a publicly available anonymized dataset and a publicly available prediction model that resembles the original model in terms of functionality and features.

Sl. No.	Device Name	Risk assessment guideline followed	Known attacks and vulnerabilities in compatible peripheral sensor devices	Connected to the Internet or Bluetooth ?
1	CardioLogs ECG Analysis Platform†	Inadequate information - Acknowledges the need for cybersecurity of cloud-based software	Portable ECG Monitors - { [35]} Ⓛ	Cellular network, Bluetooth
2	Oxehealth Vital Signs†	Guidance for the Content of Premarket Submissions for Management of Cybersecurity in Medical Devices [39]	Infra-red sensitive cameras - [36] Ⓛ, { [37]} Ⓡ	Intranet / Internet
3	GI Genius‡	Moderate level of concern as defined in the "Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices." [39]	Endoscope cameras - { [40]} Ⓡ	Intranet / Internet
4	SOZO‡	None	No third-party peripheral device used	Bluetooth, Intranet/Internet
5	WellDoc BlueStar†	Guidance for the Content of Premarket Submissions for Management of Cybersecurity in Medical Devices [39]	No vulnerability identified in peripheral sensor devices	Bluetooth, Cloud Service API
6	d-Nav System†	Guidance for the Content of Premarket Submissions for Management of Cybersecurity in Medical Devices [39]	No vulnerability identified in peripheral sensor devices	Cloud Service API
7	MBT-CA System‡	None	No third-party peripheral device used	No
8	KIDScore D3†	None	No vulnerability identified in peripheral sensor devices	Intranet / Internet
9	NuVasive Pulse System‡	None	Infra-red sensitive cameras - [36] Ⓛ, { [37]} Ⓡ	Internet
10	ABMD Software†	None	No vulnerability identified in peripheral sensor devices	Unknown
11	Deep Learning Image Reconstruction†	None	X-ray machines - { [41]} Ⓡ	Unknown
12	Air Next‡	None	No third-party peripheral device used	Bluetooth, Internet
13	One Drop Blood Glucose Monitoring System	None	No third-party peripheral device used	Bluetooth, Internet
14	OTIS 2.1 and THiA Optical Coherence Tomography System	ANSI AAMI ISO 14971:2007/(R)2010 [42], IEC 62304:2006/A1:2015 [43]	No third-party peripheral device used	Unknown
15	EarliPoint System‡	None	Webcams installed on personal computers - [44] Ⓡ	Internet
16	BrainScope TBI‡	None	No third-party peripheral devices used	Internet
17	IDx-DR v2.3†	Considers security concerns related to data confidentiality, integrity, availability, denial of service attacks and malware. Risks related to the failure of various software components and their potential impact on patient reports were also adequately addressed [45].	This device uses the Topcon NW200 Fundus camera, which comes packaged with a PC running Windows 7 OS. The Windows 7 OS has known vulnerabilities. - { [38]} Ⓡ	Internet
18	Iris Intelligent Retinal Imaging System†	Ensures HIPAA [46] compliance	Retinal cameras such as Topcon NW200 - Same vulnerable peripherals as in the case of IDx-DR v2.3	Internet
19	Paige Prostate†	Considers software security as per "Content of Premarket Submissions for Management of Cybersecurity in Medical Devices: Guidance for Industry and Food and Drug Administration Staff". Also encrypts the communication between the device and servers.	Medical scanners - { [47]} Ⓛ	Internet
20	Tissue of Origin Test Kit‡	None	No third-party peripheral device used	Internet

TABLE II: Known vulnerabilities in peripheral devices and communication media compatible with FDA-approved ML-enabled medical devices. Ⓛ: Locally exploitable only, Ⓡ: Remotely exploitable

and does not have access to the training set. The attacker can attack the Bluetooth communication stack via known exploits [21] to intercept and manipulate the CGM measurements. This is because the FDA-approved diabetes management devices (e.g., One Drop and WellDoc BlueStar) use Bluetooth communication to transmit the collected measurements.

Attack Strategy. The attacker aims to misdiagnose the patient as hyperglycemic by pushing predicted blood glucose levels toward the hyperglycemic range. This involves modifying hypoglycemic or normal blood glucose levels to values exceeding 125 mg/dL (hyperglycemic while fasting) or 180 mg/dL (hyperglycemic postprandial). To achieve this, the attacker manipulates CGM readings for a specific duration, causing the BGMS app to misdiagnose the patient's blood glucose level. Determining the minimum time duration and extent of manipulation requires careful consideration.

B. Experimental Setup

We first present the ML model used in the BGMS setup, followed by a description of the dataset used for our experiments. Next, we describe the Universal Robustness Evaluation Toolkit (URET) [49], used for generating the adversarial inputs.

Targeted ML model. Since the specific ML algorithm used in the Dreamed Advisor Pro app (described in Section II-B) is confidential, we approximated it using a time-series prediction model developed by Rubin-Falcone et al. [20]. This model uses a bidirectional long short-term memory (LSTM) recurrent neural network (RNN) architecture, and uses root mean square error (RMSE) and mean absolute error (MAE) to evaluate the prediction accuracy. Intuitively, both RMSE and MAE indicate the difference between predicted and actual glucose levels. The higher the difference, the worse the prediction. Further, our chosen target model uses a neural network similar to the FDA-approved d-Nav System [50] in Tables I and II (i.e., LSTM).

Rubine-Falcone et al. [20] built two models - (i) a personalized model for each patient trained on the patient's data, and (ii) an aggregate model trained on the data of all patients. Their average RMSEs are 18.2 and 31.7, and average MAEs are 12.8 and 23.6 on the 30 and 60-minute horizons, respectively.

Dataset. To demonstrate the impact of adversarial inputs on the predictions of the targeted ML model, we use the 2020 OhioT1DM dataset [51]. This was used by the target model developers [20] for evaluating its accuracy. The dataset comprises physiological measurements of six Type-1 diabetic patients. The main features are CGM blood glucose measurements, finger-based measurements, basal insulin, bolus dose, carbohydrate intake, heart rate, sleeping patterns and acceleration, besides other physiological, and self-reported life-event features. The dataset spans eight weeks and consists of ≈10000 samples for training, and 2500 for testing, both recorded at approximately 5-minute intervals per patient.

Universal Robustness Evaluation Toolkit (URET). In Table I, we found three diabetes management devices (5,6, and 13) that are vulnerable to inference-time attacks. Hence, we decided to launch an *evasion attack* against our target model, which manipulates data points at inference time by introducing slight perturbations to their original values to evade detection by the ML model and cause misclassification [52]. We use URET [49], a general-purpose framework for generating adversarial inputs for evasion attacks. Unlike most evasion attack techniques that focus on the image classification domain [32], [53]–[55], URET can generate adversarial inputs regardless of the data types or task domain.

URET takes in a benign input instance and a set of pre-defined input transformations and attempts to find a sequence of transformations (e.g., increment or replace glucose level) resulting in an adversarial input that is both semantically and functionally correct. The URET framework is compatible with a wide variety of data types and domains and allows the user to specify which feature values to manipulate, and bounds and constraints on the feature values. As URET allows specifying constraints on the feature values, we can ensure the generated adversarial samples abide by victims' physiological limits.

Since we assume the attacker can only modify CGM values, we specify the CGM feature indices in URET's configuration file as the only modifiable feature. To ensure that adversarial CGM values respect the physiological levels, we constrain them to be between 125 and 499 mg/dL for fasting levels, since a hyperglycemic glucose level in a fasting state should exceed 125 mg/dL, and between 180 and 499 mg/dL for postprandial levels, since a hyperglycemic glucose level in a postprandial state should exceed 180 mg/dL (499 mg/dL is the highest reported glucose level in the OhioT1DM dataset). Further, since there are missing physiological measurements at specific timestamps in the dataset, especially CGM measurements, we specify the relationship between CGM glucose and a feature called "missing" as a dependency. This ensures the ML model does not interpret the missing CGM values as zero values.

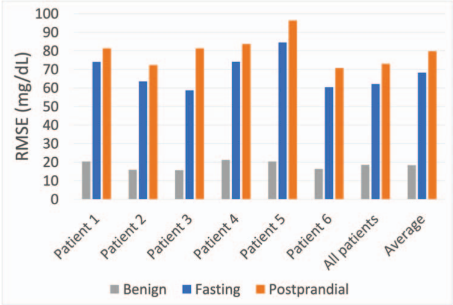

Fig. 2: RMSE of the benign model, the model attacked with fasting hyperglycemic blood glucose levels, and with postprandial hyperglycemic blood glucose levels.

In addition, the attacker can observe other feature values (e.g., bolus dose), which in turn play a role in adversarial data generation, since the attacker needs read access to those features to use the loss function to rank the adversarial input.

C. Results

While the target models use RMSE and MAE to evaluate the prediction accuracy, we consider the RMSE results only, as the two metrics show similar behaviors. Figure 2 shows the performance of personalized patient models and the aggregate model trained on all patients' data ('All patients') before and after the URET attacks while fasting and postprandial. The maximum RMSE value for the benign models is around 21 mg/dL, while the average RMSE across all 7 benign models is around 18.3 mg/dL. The figure also shows the RMSE values after the attack for both fasting (>125 mg/dL) and postprandial (>180 mg/dL) hyperglycemic glucose levels. Thus, the RMSE values increase regardless of the used threshold since glucose values are driven further away from the actual values.

Moreover, postprandial RMSE values are consistently higher than fasting RMSE values due to the use of a higher threshold for the lower bound of the adversarial input (i.e., 180 mg/dL instead of 125 mg/dL). The hike in RMSE values between the benign and the attacked models in both cases shows a significant difference between the actual and predicted blood glucose levels, implying diminished accuracy of the ML model. This would cause a potentially fatal insulin overdose.

Figures 3 and 4 show the percentage of instances that are misdiagnosed as hyperglycemic while actually being normal and hypoglycemic, respectively, for both the fasting and postprandial attack scenarios. A higher misclassification percentage demonstrates more susceptibility to the respective attack scenario, while a lower percentage implies more difficulty in crafting successful attacks. We make three observations from the two figures. *First,* URET achieves considerably

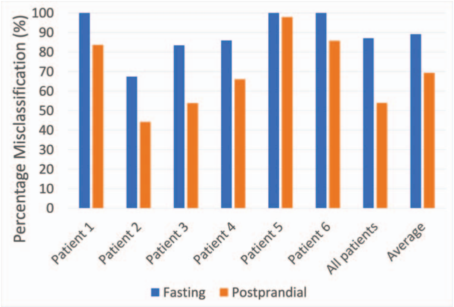

Fig. 3: Percentage of originally normal glucose instances that are misclassified as hyperglycemic.

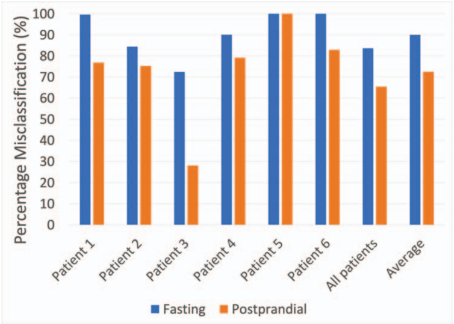

Fig. 4: Percentage of originally hypoglycemic glucose instances that are misclassified as hyperglycemic.

high attack success rates, reaching up to 100% in some cases. URET achieves comparable attack success rates in both normal-to-hyperglycemic and hypoglycemic-to-hyperglycemic scenarios on average, demonstrating the robustness of the attack generated by URET across different initial glucose levels. *Second,* the attack success rate is consistently higher during fasting compared to postprandial states, indicating that attacking a fasting patient is relatively easier for the adversary. This is because a smaller amount of perturbation is performed when increasing the CGM glucose from the original value to fasting hyperglycemic levels as opposed to postprandial hyperglycemic levels. This confirms that URET is more successful when the perturbation margin is smaller. *Third,* patients exhibit varying resilience to the URET attacks. For example, Figure 3 shows that the success rate of misdiagnosing a patient from normal to hyperglycemic is the lowest for patient 2 (67.4% fasting, and 44.2% postprandial) and is the highest for patient 5 (100.0% fasting, and 97.9% postprandial) suggesting that it is more challenging for URET to attack patient 2 compared to patient 5. Similarly, Figure 4 shows that the attack success rate of misdiagnosing a patient from hypoglycemic to hyperglycemic is the lowest with patient 3 (72.4% fasting, and 28.0% postprandial) and is the highest with patient 5 (100.0% fasting, and 100.0% postprandial) suggesting that patient 5 is more vulnerable compared to patient 3. We hypothesize that this is because the rich history of some patients leads to a more robust prediction model that is more difficult to attack.

Summary. We demonstrated an attack where an attacker compromises the Bluetooth communication between the CGM peripheral device and the BGMS mobile app, enabling them to manipulate the measured glucose levels. The experimental results show that the attacker can cause the ML model to misdiagnose the patient's medical condition as hyperglycemic. In the best-case, this leads to a wrong diagnosis, and in the worst-case , it has fatal consequences for the patient.

D. Discussion: Attack Practicality and Broader Impact

The assumptions made in the proposed attack are practical and align with previously demonstrated attacks [21]. While no real attack on a BGMS has been reported, security breaches in other medical devices, such as pacemakers, have been reported in the recent past [56]. The severity of these breaches is evident from the widespread recall of multiple pacemaker models and the substantial financial loss incurred by the manufacturer [57]. These incidents also raised concerns about targeted attacks on high-profile individuals using such devices. For example, in 2013, former US Vice President Dick Cheney revealed that he had the wireless capabilities of his implanted pacemaker deactivated due to fears that an adversary could cause a cardiac arrest by sending a malicious signal to his pacemaker [58].

V. Risk Assessment Techniques Evaluation

We assess the state-of-the-art risk assessment techniques, investigating both their strengths and limitations. We discuss the benefits of these methods in risk analysis processes while pointing out their shortcomings using the BGMS attack example. Finally, we highlight the need for a new risk assessment framework for ML-enabled connected medical systems.

A. A Survey of Existing Techniques

We consider five broad categories of risk assessment methods. However, in the context of ML-enabled connected medical devices, these techniques often lack in three key aspects: (1) impact on affected users (more users affected, higher risk), (2) ease of vulnerability detection (easier detection, lower risk), and (3) ease of post-exploitation mitigation, including available remediation levels and responsible entities. Table III presents the extent to which existing risk assessment methods incorporate these factors. We describe the methods below.

DREAD. DREAD [59] is a *risk rating system* built by Microsoft, primarily for evaluating risks posed to conventional web applications based on their damage potential, reproducibility, exploitability, affected users, and discoverability. However, the DREAD system cannot be used for ML-enabled connected medical systems. This is because DREAD does not detail how security risks in individual connected components may pose a threat to the overall system.

STRIDE. STRIDE [59] by Microsoft is another qualitative model designed to *identify and categorize threats* in web applications. It addresses six attack types: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and

Method	Damage Potential	Exploitability	Affected Users	Detectability	Ease of Mitigation
DREAD	✓	✓	✓	✓	✗
STRIDE	✓	✗	✗	✗	✗
FMEA	✓	✗	✗	✓	✗
CVSS	✓	✓	✗	✗	✗
Other Works [61], [62]	✓	✓	✗	✗	✗

TABLE III: Comparison of factors considered in existing risk analysis frameworks

Elevation of Privilege, and offers corresponding countermeasures. However, using STRIDE for an end-to-end risk assessment in ML-based medical systems is challenging because it does not account for adversaries exploiting peripheral devices.

FMEA. Failure Modes and Effects Analysis (FMEA) [60] is a foundational analytical technique to *detect and mitigate* potential risks. This method involves a detailed examination of system components to identify potential causes of failure and their impact on system stability. However, it fails to provide end-to-end risk assessment for connected ML-based medical systems because it primarily focuses on individual component failures, overlooking broader systemic implications and how vulnerabilities propagate throughout the entire pipeline.

CVSS. The Common Vulnerability Scoring System (CVSS) [63] is an open risk scoring framework designed by FIRST.Org, Inc. to *capture the severity level* of software vulnerabilities. CVSS comprises three metric groups: Base, Temporal, and Environmental, capturing different vulnerability characteristics. While it is not a risk assessment model [63], it helps prioritize threats across system components through context-specific choices made by the risk management team. However, it lacks consistency in prioritizing metrics, notably in ML-enabled medical devices where availability loss could be life-threatening, unlike in web applications where it might lead to user dissatisfaction or financial repercussions [64].

FDA-approved Security Standards for Medical Devices. The FDA has established a cybersecurity guideline to help the industry identify cybersecurity risks in medical devices [39]. However, their primary focus is on risks associated with communication between medical devices and IT networks. Consequently, addressing vulnerabilities in AI/ML-enabled medical devices that do not interface with an IT network remains a challenge.

Academic Research Efforts and Industry. There has been a rich literature on performing risk assessment within the healthcare sector [61], [62]. While companies typically conduct preliminary risk evaluations for their ML models, they do not analyze the security risks associated with deploying the models in a connected healthcare environment [65]. Rather, their concern is ensuring the ML model is trained on an unbiased dataset, and evaluated using a diverse dataset. Similarly, while there are companies that claim to offer penetration testing services for medical devices, their testing is conducted in-silo rather than on the end-to-end connected system.

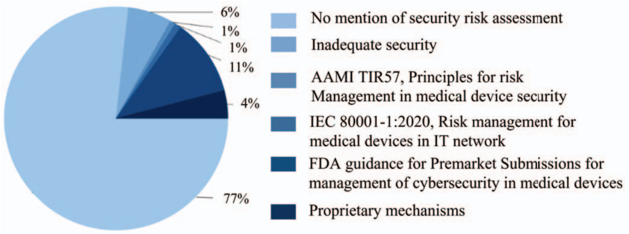

Fig. 5: Security risk assessment techniques by manufacturers of FDA-approved ML-enabled medical systems based on [39]

B. Limitations of Existing Techniques

We use the ML-enabled BGMS as an example to highlight the shortcomings of existing techniques in adequately considering all the factors required for an ideal risk analysis.

DREAD. DREAD fails to assess the actual potential damage if a peripheral device like the glucose meter or smartphone is compromised. In such cases, the attacker can not only access the data but also to manipulate it, thereby compromising both the integrity and confidentiality of the system. However, DREAD cannot differentiate between them accurately.

STRIDE. An attacker can manipulate data in several ways before it is fed into the ML engine. Nevertheless, according to the STRIDE model, all these methods are categorized under Tampering, without any additional insights into how these manipulations were performed.

FMEA. FMEA could potentially assess the risks associated with individual peripheral devices in the BGMS, like glucose meters or smartphones. However, FMEA does not offer insights into the risks associated with the propagation of this vulnerability within the system or its potential impact on the ML model's predictions.

CVSS. When a vulnerability impacts the availability of the BGMS, the CVSS assigns the same risk level as it would for a similar incident in other domains like web applications. This approach is not appropriate, as an availability issue in the BGMS could potentially lead to irreversible harm to patients.

FDA-approved Security Standards for Medical Devices. Encountering physical attacks like electromagnetic interference [24] is independent of the connectivity between the glucose meter and the IT network. Hence, the FDA-approved security standards also fall short of an ideal risk analysis.

C. Need for a New Risk Assessment Framework

Our investigation into the types of security risk assessment performed by manufacturers of ML-enabled medical devices is summarized in Figure 5.We find that over 80% of these manufacturers either do not provide information about the assessment in their documentation, or employ inadequate assessment methods. Another 5% use proprietary mechanisms, making it challenging to assess the adequacy of their approach. The remaining 12% utilize existing risk assessment techniques, which, as discussed, are insufficient for risk assessment of ML-enabled connected medical systems. Therefore, developing an efficient risk assessment technique for ML-enabled medical devices remains an open challenge.

VI. Conclusion and Future Research Directions

We conducted a detailed study of security risks associated with modern AI/ML-enabled medical devices, stemming from vulnerabilities in connected peripheral devices. We conducted a systematic security analysis of FDA-approved commercial AI/ML-enabled devices. Our analyses reveal vulnerabilities of these devices to existing adversarial attacks, raising concerns about the suitability of using such safety-critical devices on real-world patients. To validate our analysis, we executed a realistic adversarial attack on an ML-enabled blood glucose monitoring system, identifying security risks in the process. Additionally, we studied state-of-the-art risk assessment frameworks, underscoring their limitations in identifying security risks in connected ML-enabled medical systems and highlighting the need for a new framework.

Our work opens up three interesting future work directions – **(1)** Automated risk identification: Automating the risk identification process at scale would benefit device manufacturers as well as the security research community. This would require identifying relevant documents on the web and parsing a huge volume of unstructured documents, while at the same time being able to relate various ML concepts. The automated tool also needs to be interfaced with the state-of-the-art vulnerability databases and repositories of peer-reviewed research works so that it can even identify emerging threats in AI and medical devices; , **(2)** Building personalized spatial and temporal risk profiles per patient: Our case study shows that attacks on ML-enabled medical systems cause more damage to certain patients than others. Moreover, a patient is not equally vulnerable at all points of time. An interesting research problem is to study patients' data in more detail to develop customized spatial and temporal risk profiles for every patient; and, **(3)** Efficient risk mitigation techniques: This involves designing attack-resilient ML models, determining accountable entity and enforcing accountability in risk mitigation, accounting for the costs and deployment scenario.

We have made our code and datasets publicly available at: https://gm-repo.github.io/Security-MEDAI/

Acknowledgements

This project is supported by collaborative research funding from the National Research Council of Canada's Digital Health and Geospatial Analytics Program, UBC Four Year Fellowships (FYF), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] U.S. FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices, Last accessed: Nov 30, 2023.

[2] Huangxun Chen, Chenyu Huang, Qianyi Huang, Qian Zhang, and Wei Wang. Ecgadv: Generating adversarial electrocardiogram to misguide arrhythmia classification system. In *AAAI Conference on Artificial Intelligence*, volume 34, pages 3446–3453, 2020.

[3] Albatul Albattah and Murad A Rassam. Detection of Adversarial Attacks against the Hybrid Convolutional Long Short-Term Memory Deep Learning Technique for Healthcare Monitoring Applications. *Applied Sciences*, 13(11):6807, 2023.

[4] Sheeba Lal, Saeed Ur Rehman, Jamal Hussain Shah, Talha Meraj, Hafiz Tayyab Rauf, Robertas Damaševičius, Mazin Abed Mohammed, and Karrar Hameed Abdulkareem. Adversarial attack and defence through adversarial training and feature fusion for diabetic retinopathy recognition. *Sensors*, 21(11):3922, 2021.

[5] Amiel Meiseles, Ishai Rosenberg, Yair Motro, Lior Rokach, and Jacob Moran-Gilad. Adversarial Vulnerability of Deep Learning Models in Analyzing Next Generation Sequencing Data. In *IEEE BIBM*, pages 464–468, 2020.

[6] Marina Z Joel, Sachin Umrao, Enoch Chang, Rachel Choi, Daniel Yang, James Duncan, Antonio Omuro, Roy Herbst, Harlan M Krumholz, Sanjay Aneja, et al. Adversarial attack vulnerability of deep learning models for oncologic images. *MedRxiv*, 2021.

[7] Gerda Bortsova, Cristina González-Gonzalo, Suzanne C Wetstein, Florian Dubost, Ioannis Katramados, Laurens Hogeweg, Bart Liefers, Bram van Ginneken, Josien PW Pluim, Mitko Veta, et al. Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. *Medical Image Analysis*, 73:102141, 2021.

[8] Karthika Menon, V Khushi Bohra, Lakshana Murugan, Kavya Jaganathan, and Chamundeswari Arumugam. COVID-19 Diagnosis from Chest X-Ray Images Using Convolutional Neural Networks and Effects of Data Poisoning. In *ICCSA*, pages 508–521, 2021.

[9] Danilo Vasconcellos Vargas and Jiawei Su. Understanding the one-pixel attack: Propagation maps and locality analysis. In *CEUR Workshop Proceedings*, volume 2640, 2020.

[10] Tamar Levy-Loboda, Eitam Sheetrit, Idit F Liberty, Alon Haim, and Nir Nissim. Personalized insulin dose manipulation attack and its detection using interval-based temporal patterns and machine learning algorithms. *Journal of Biomedical Informatics*, 132:104129, 2022.

[11] Yuzhong Chen, Jiadong Yan, Mingxin Jiang, Tuo Zhang, Zhongbo Zhao, Weihua Zhao, Jian Zheng, Dezhong Yao, Rong Zhang, Keith M Kendrick, et al. Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[12] Jianfeng Yu, Kai Qiu, Pengju Wang, Caixia Su, Yufeng Fan, and Yongfeng Cao. Perturbing BEAMs: EEG adversarial attack to deep learning models for epilepsy diagnosing. *BMC Medical Informatics and Decision Making*, 23(1):115, 2023.

[13] Christopher Nielsen, Anup Tuladhar, and Nils D Forkert. Investigating the Vulnerability of Federated Learning-Based Diabetic Retinopathy Grade Classification to Gradient Inversion Attacks. In *International Workshop on Ophthalmic Medical Image Analysis*, pages 183–192, 2022.

[14] Neal Mangaokar, Jiameng Pu, Parantapa Bhattacharya, Chandan K Reddy, and Bimal Viswanath. Jekyll: Attacking medical image diagnostics using deep generative models. In *IEEE EuroS&P*, pages 139–157, 2020.

[15] Lei Hu, Da-Wei Zhou, Xiang-Yu Guo, Wen-Hao Xu, Li-Ming Wei, and Jun-Gong Zhao. Adversarial training for prostate cancer classification using magnetic resonance imaging. *Quantitative Imaging in Medicine and Surgery*, 12(6):3276, 2022.

[16] Xingjun Ma, Yuhao Niu, Lin Gu, Yisen Wang, Yitian Zhao, James Bailey, and Feng Lu. Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition*, 110:107332, 2021.

[17] U.S. FDA. DreaMed Advisor Pro. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN170043, Last accessed: Nov 30, 2023.

[18] Michael Kahn. Diabetes. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5T59G.

[19] MIchelle Mouri and Madhu Badireddy. Hyperglycemia. URL: https://www.ncbi.nlm.nih.gov/books/NBK430900/, Last accessed: Nov 30, 2023.

[20] Harry Rubin-Falcone, Ian Fox, and Jenna Wiens. Deep Residual Time-Series Forecasting: Application to Blood Glucose Prediction. In *KDH@ ECAI*, pages 105–109, 2020.

[21] Kasper Rasmussen. BLURtooth: Exploiting Cross- Transport Key Derivation in Bluetooth Classic and Bluetooth Low Energy. In *AsiaCCS*, 2022.

[22] Dreamed Diabetes Ltd. DreaMed Advisor Pro: Manual For Personal Use iOS. URL: https://dreamed-diabetes.com/wp-content/uploads/2019/06/Dreamed-Advisor-Pro-iOS-Patient-IFU.pdf, Last accessed: Nov 30, 2023.

[23] Guillaume Dupont, Daniel Ricardo dos Santos, Elisa Costante, Jerry Den Hartog, and Sandro Etalle. A matter of life and death: analyzing the security of healthcare networks. In *SEC 2020: ICT Systems Security and Privacy Protection*, pages 355–369, 2020.

[24] SMJ Mortazavi, M Gholampour, M Haghani, G Mortazavi, and AR Mortazavi. Electromagnetic radiofrequency radiation emitted from GSM mobile phones decreases the accuracy of home blood glucose monitors. *Journal of Biomedical Physics & Engineering*, 4(3):111, 2014.

[25] Tim Hulsen, Saumya S Jamuar, Alan R Moody, Jason H Karnes, Orsolya Varga, Stine Hedensted, Roberto Spreafico, David A Hafler, and Eoin F McKinney. From big data to precision medicine. *Frontiers in medicine*, 6:34, 2019.

[26] Spiros V Georgakopoulos, Dimitris K Iakovidis, Michael Vasilakakis, Vassilis P Plagianakos, and Anastasios Koulaouzidis. Weakly-supervised convolutional learning for detection of inflammatory gastrointestinal lesions. In *IEEE IST*, pages 510–514, 2016.

[27] Amalia M Issa. Personalized medicine and the practice of medicine in the 21st century. *McGill Journal of Medicine: MJM*, 10(1):53, 2007.

[28] Muhammad Shahid Amin, Jamal Hussain Shah, Mussarat Yasmin, Ghulam Jillani Ansari, Muhamamd Attique Khan, Usman Tariq, Ye Jin Kim, and Byoungchol Chang. A two-stream fusion assisted deep learning framework for stomach diseases classification. *CMC-Comput. Mater. Contin*, 73:4423–4439, 2022.

[29] Michal Byra, Grzegorz Styczynski, Cezary Szmigielski, Piotr Kalinowski, Lukasz Michalowski, Rafal Paluszkiewicz, Bogna Ziarkiewicz-Wroblewska, Krzysztof Zieniewicz, and Andrzej Nowicki. Adversarial attacks on deep learning models for fatty liver disease classification by modification of ultrasound image reconstruction method. In *IEEE IUS*, pages 1–4, 2020.

[30] Xugui Zhou, Maxfield Kouzel, and Homa Alemzadeh. Robustness testing of data and knowledge driven anomaly detection in cyber-physical systems. In *IEEE/IFIP DSN-W*, pages 44–51, 2022.

[31] TV Nguyen, MA Dakka, SM Diakiw, MD VerMilyea, M Perugini, JMM Hall, and D Perugini. A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. *Scientific Reports*, 12(1):8888, 2022.

[32] Rahul Paul, Matthew Schabath, Robert Gillies, Lawrence Hall, and Dmitry Goldgof. Mitigating adversarial attacks on medical image understanding systems. In *IEEE ISBI*, pages 1517–1521, 2020.

[33] Mitre. Common Vulnerabilities and Exposures. URL: https://cve.mitre.org/, Last accessed: Nov 30, 2023.

[34] Information Technology Laboratory, USA. National Vulnerability Database. URL: https://nvd.nist.gov/vuln/search, Last accessed: Nov 30, 2023.

[35] Mitre. Conexus Telemetry Protocol vulnerability. URL: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2019-6538, Last accessed: Nov 30, 2023.

[36] Wei Wang, Yao Yao, Xin Liu, Xiang Li, Pei Hao, and Ting Zhu. I can see the light: Attacks on autonomous vehicles using invisible lights. In *ACM SIGSAC CCS*, pages 1930–1944, 2021.

[37] Mitre. Sony IPELA E Series Camera vulnerability (1). URL: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2018-3938, Last accessed: Nov 30, 2023.

[38] Mitre. Windows 7 vulnerability (2). URL: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2019-5921, Last accessed: Nov 30, 2023.

[39] U.S. FDA. Content of Premarket Submissions for Management of Cybersecurity in Medical Devices. URL: https://www.fda.gov/media/86174/download, Last accessed: Nov 30, 2023.

[40] Mitre. Shekar Endoscope vulnerability (1). URL: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2017-10722, Last accessed: Nov 30, 2023.

[41] Mitre. GE Healthcare Discovery vulnerability. URL: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-7232, Last accessed: Nov 30, 2023.

[42] AAMI. ANSI/AAMI/ISO 14971: 2007/(R) 2010, Medical devices—Application of risk management to medical devices.

[43] International Electrotechnical Commission et al. IEC 62304: 2006/A1: 2015. *Medical device software-Software life-cycle processes*, 2015.

[44] Mitre. H264WebCam vulnerability. URL: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2010-2349, Last accessed: Nov 30, 2023.

[45] U.S. FDA. IDx-DR v2.3. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K213037, Last accessed: Nov 30, 2023.

[46] George J Annas. HIPAA regulations: a new era of medical-record privacy? *New England Journal of Medicine*, 348:1486, 2003.

[47] Mitre. Philips MRI 1.5T and MRI 3T vulnerability (1). URL: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2021-26262, Last accessed: Nov 30, 2023.

[48] Guillermo Suarez-Tangil, Juan E Tapiador, and Pedro Peris-Lopez. Compartmentation policies for android apps: A combinatorial optimization approach. In *NSS 2015*, pages 63–77, 2015.

[49] Kevin Eykholt, Taesung Lee, Douglas Schales, Jiyong Jang, and Ian Molloy. URET: Universal Robustness Evaluation Toolkit (for Evasion). In *USENIX Security 23*, pages 3817–3833, 2023.

[50] U.S. FDA. d-Nav System. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K181916, Last accessed: Nov 30, 2023.

[51] Cindy Marling and Razvan Bunescu. The OhioT1DM dataset for blood glucose level prediction: Update 2020. In *CEUR workshop proceedings*, volume 2675, page 71. NIH Public Access, 2020.

[52] Battista Biggio, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Šrndić, Pavel Laskov, Giorgio Giacinto, and Fabio Roli. Evasion attacks against machine learning at test time. In *ECML PKDD 2013, Proceedings, Part III 13*, pages 387–402, 2013.

[53] Samuel G Finlayson, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane. Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289, 2019.

[54] Samuel G Finlayson, Hyung Won Chung, Isaac S Kohane, and Andrew L Beam. Adversarial attacks against medical deep learning systems. *arXiv preprint arXiv:1804.05296*, 2018.

[55] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *IEEE SP*, pages 39–57, 2017.

[56] Alex Hern. Hacking risk leads to recall of 500,000 pacemakers due to patient death fears. URL: https://www.theguardian.com/technology/2017/aug/31/hacking-risk-recall-pacemakers-patient-death-fears-fda-firmware-update, Last accessed: Nov 30, 2023.

[57] Leigh Day. Lawyers investigate claims following further recall of 'Assurity' and 'Endurity' Abbott Laboratories pacemakers. URL: https://www.lexology.com/library/detail.aspx?g=08a4668c-cea0-4f60-b348-394ae5209ebc, Last accessed: Nov 30, 2023.

[58] Sanjay Gupta. Dick Cheney's heart. URL: https://www.cbsnews.com/news/dick-cheneys-heart/, Last accessed: Nov 30, 2023.

[59] J.D. Meier and Microsoft Corporation. *Improving Web Application Security: Threats and Countermeasures*. Patterns & practices. Microsoft press, 2003.

[60] Hu-Chen Liu, Li-Jun Zhang, Ye-Jia Ping, and Liang Wang. Failure mode and effects analysis for proactive healthcare risk evaluation: a systematic literature review. *Journal of evaluation in clinical practice*, 26(4):1320–1337, 2020.

[61] Tom Mahler, Yuval Elovici, and Yuval Shahar. A new methodology for information security risk assessment for medical devices and its evaluation. *arXiv preprint arXiv:2002.06938*, 2020.

[62] Tahreem Yaqoob, Haider Abbas, and Narmeen Shafqat. Integrated security, safety, and privacy risk assessment framework for medical devices. *IEEE journal of biomedical and health informatics*, 24(6):1752–1761, 2019.

[63] R Abraham, D Arora, M Coles, M Eckert, M Heitman, A Manion, S Moore, S Romanowsky, K Scarfone, J Stuppi, et al. Common vulnerability scoring system v3. 0: Specification document. *First*, 2015.

[64] Jonathan Spring, Eric Hatleback, Allen Householder, Art Manion, and Deana Shick. Time to Change the CVSS? *IEEE S&P*, 19(2):74–78, 2021.

[65] Eric Wu, Kevin Wu, Roxana Daneshjou, David Ouyang, Daniel E Ho, and James Zou. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nature Medicine*, 27(4):582–584, 2021.