Review

Mapping Machine Learning–Driven Cybersecurity Solutions in Health Care: Scoping Literature Review

Kunal Rajput[1], MBChB; Sharukh Zuberi[2], MBBS, BSc; Mireille Elhajj[3,4,5], PhD; Washington Ochieng[3,5], BSc, MSc, PhD; Ara Darzi[1], MBChB, MD; Saira Ghafur[1], MBChB, MSc

[1]Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, England, United Kingdom
[2]Department of Surgery, The Royal Free Hospital, London, England, United Kingdom
[3]Department of Civil and Environmental Engineering, Faculty of Engineering, Imperial College London, London, England, United Kingdom
[4]Astra-Terra Limited, London, England, United Kingdom
[5]Centre for Active Resilience and Security, Imperial College London, London, England, United Kingdom

Corresponding Author:

Kunal Rajput, MBChB
Department of Surgery and Cancer
Faculty of Medicine, Imperial College London
Faculty Building, South Kensington Campus
London, England SW7 2AZ
United Kingdom
Phone: 44 02075895111
Email: k.rajput24@imperial.ac.uk

Abstract

Background: Health care systems face escalating cyberattacks, including the UK Synnovis ransomware attack, which halted pathology services for 14 weeks; the Ascension Health breach affecting 5.6 million patients; and the Change Healthcare breach costing US $2.5 billion. Conventional cybersecurity measures in health care remain reactive and inadequate against evolving threats. Machine learning (ML) offers adaptive, predictive, real-time cyber defense; yet, there is limited clarity on how ML tools are applied across cybersecurity domains, their real-world effectiveness, and where gaps remain.

Objective: This study aims to map ML applications in health care cybersecurity against the National Institute of Standards and Technology Cybersecurity Framework version 2.0, summarize ML performance, and identify research gaps and implementation considerations.

Methods: A systematic search of Ovid MEDLINE, Embase, and Scopus was conducted on July 30, 2025, for studies between 2019 and 2025. Eligible studies applied ML-based approaches to organizational-level cybersecurity in health care settings, with outcomes related to data privacy or cybersecurity strengthening. Studies on smart devices, blockchain, or those lacking empirical data were excluded. Title and abstract and full-text screening were conducted independently by 2 (KR and SZ) reviewers following the Arksey and O'Malley framework and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, with discrepancies resolved by consensus. Data were synthesized narratively and mapped against the 6 National Institute of Standards and Technology Cybersecurity Framework version 2.0 functions (Identify, Protect, Detect, Respond, Recover, and Govern).

Results: From 10,348 articles identified, 45 studies across 18 countries were included, applying 80 ML models. Most studies addressed "Protect" (n=22, 48.9%), encompassing federated learning, homomorphic encryption, and deidentification pipelines. "Detect" (n=13, 28.9%) covered intrusion detection and anomaly-based threat detection. "Identify" (n=8, 17.8%) addressed risk assessment and vulnerability prediction. Only 2 studies addressed "Respond," and none addressed "Recover" or "Govern." Classical ML, deep learning, and natural language processing predominated, with intrusion detection being the most common application (n=29). Despite strong controlled performance, only one study demonstrated real-world deployment; most rely on synthetic or outdated benchmark datasets and inconsistent reporting metrics.

Conclusions: To our knowledge, this is the first scoping review to map ML-driven cybersecurity solutions against all six National Institute of Standards and Technology Cybersecurity Framework version 2.0 functions, offering a structured, policy-relevant evidence base. ML applications remain concentrated in preincident functions, with gaps in response, recovery, and governance. This highlights systematic blind spots and priorities for researchers and implementers. The evidence supports

a phased implementation approach: beginning with detection systems, where evidence is most established and integration is most feasible, progressing to privacy-preserving architectures, for which the literature currently offers no guidance. Implementation requires sustained investment in infrastructure, representative datasets, explainable AI, and real-world validation. Limitations include language bias, methodological heterogeneity, and exclusion of medical devices and proprietary solutions.

Trial Registration: OSF Registries osf.io/4cjmu; https://osf.io/4cjmu/overview

J Med Internet Res 2026;28:e93950; doi: 10.2196/93950

Keywords: machine learning; artificial intelligence; cybersecurity; healthcare systems; intrusion detection; data protection; digital health

Introduction

Background

The digitization of health care through networked health information systems, electronic health record (EHR) infrastructure, cloud storage, and telemedicine has transformed care delivery but simultaneously expanded the attack surface, making the sector increasingly vulnerable to sophisticated cyberattacks [1]. Recent high-profile incidents illustrate the scale of disruption. The Synnovis ransomware attack in Southeast London disrupted pathology services across 5 major hospitals and 194 primary care practices, resulting in a 14-wk suspension of blood product processing and delayed clinical diagnostics [2]. The Ascension Health ransomware attack affected 5.6 million patients' records and led to widespread disruption of clinical services, including delayed treatments and appointments [3]. Change Healthcare, a subsidiary of UnitedHealth Group, incurred economic losses estimated at nearly US $2.5 billion due to billing and prescription shutdowns, highlighting the direct financial impact of cyber incidents [4]. These incidents demonstrate that cyberattacks in health care can have critical impacts across multiple layers, including patient care, financial stability, and operational continuity.

Conventional cybersecurity tools, such as firewalls, antivirus software, intrusion detection systems, and access controls, rely on static rules and known threat signatures to identify and block attacks [5]. Over time, as health care systems have become increasingly digitalized, these tools have evolved to include more sophisticated features, such as real-time monitoring, automated patch management, and basic anomaly detection. However, their fundamentally reactive design limits their ability to address novel or rapidly evolving cyber threats. In contrast, machine learning (ML)-based systems have shown potential to offer a more adaptive, predictive, and real-time approach to cyber defense. These models can detect anomalies without relying on predefined attack signatures and can automate responses, providing scalable and timely defense. This capability is especially important in health care, where even brief outages can threaten patient safety [6].

AI represents a double-edged sword in cybersecurity. On one edge, malicious actors are increasingly using AI to orchestrate sophisticated cyberattacks capable of propagating quickly across interconnected health care systems and disrupting critical services [7,8]. Emerging threat intelligence reports suggest that AI tools have been used as autonomous penetration testing orchestrators, achieving high levels of autonomous execution across the cyberattack lifecycle [8]. Almost three-quarters of phishing attacks now leverage AI technology, with some estimates suggesting that the majority of phishing emails now incorporate AI in some form, whether for text generation, personalization, or obfuscation [9]. Conversely, AI also offers powerful defensive capabilities, enabling health care systems to adopt equally advanced AI-driven defense mechanisms that can detect anomalies in real time and automate responses to emerging threats [10,11]. This technological arms race, where both attackers and defenders leverage AI, underscores the critical need for health care organizations to not only implement AI-based security solutions but also establish robust governance frameworks and regulatory oversight to ensure these systems are deployed responsibly, monitored continuously, and protected against adversarial exploitation.

To date, AI adoption in health care has been predominantly concentrated on clinical domains such as medical imaging interpretation, pathology, and predictive analytics, with a strong emphasis on improving diagnostic accuracy and workflow efficiency [12]. ML adoption in health care cybersecurity remains limited due to the evolving complexity of cyber threats, fragmented regulation, and reliance on human expertise for real-time decisions [13]. Unlike diagnostics, where clear regulatory pathways exist, cybersecurity lacks a unified framework for ML integration. This review focuses specifically on ML, a subset of AI in which systems learn patterns from data, encompassing classical algorithms, deep learning (DL), and natural language processing (NLP). While Internet of Medical Things and smart device security represent a related concern, this review focuses on organizational-level health care systems, where ML-based cybersecurity solutions are most developed.

The UK's Cyber Security Strategy for Health and Adult Social Care (2023-2030) highlights the use of AI for intelligent threat detection and system resilience [14]. Similarly, the US National Cybersecurity Strategy (2023) advocates for AI-driven tools to enable real-time anomaly detection and adaptive defense. These policies reflect a growing shift toward proactive, AI-enabled cybersecurity strategies [15].

The National Institute of Standards and Technology Cybersecurity Framework (NIST-CSF) 2.0, released in 2024, reinforces this shift toward proactive and adaptive cyber defense [16]. The six pillars "Identify," "Protect," "Detect,"

"Respond," "Recover," and "Govern" together emphasize strategic oversight, continuous improvement, and integration of emerging technologies such as ML. For health care systems, NIST-CSF offers a flexible structure to align cybersecurity efforts with operational demands, providing a domain framework for the adoption of intelligent, AI-enabled tools for real-time detection and response.

Several reviews have examined the application of AI and ML to cybersecurity more broadly. Wiafe et al [17] conducted a systematic mapping of AI for cybersecurity across general computing environments, cataloging techniques across domains including intrusion detection, malware analysis, and network security. Ali et al [18] provided a systematic review of AI and ML techniques for cybersecurity, and a subsequent review by the same group focused specifically on DL methods for malware and intrusion detection [19]. Dhanushkodi and Thejas [20] examined AI-enabled threat detection across general security contexts, highlighting advances in anomaly-based and signature-free detection. While these reviews provide valuable insights into the broader AI-cybersecurity landscape, they are not specific to health care and do not account for the sector's unique constraints such as the sensitivity of patient data, the criticality of service continuity, legacy infrastructure, and the complex regulatory environment governing health systems. Furthermore, existing reviews tend to focus on particular technical domains, such as intrusion detection or malware classification, without offering a comprehensive mapping of how the full breadth of ML tools aligns with an established cybersecurity governance framework applicable to health care [21,22].

No prior scoping review has systematically mapped ML applications in health care cybersecurity against the NIST CSF 2.0, which captures all phases of the cyber resilience lifecycle. This review addresses that gap by providing the first governance-aligned, health care–specific synthesis of ML cybersecurity evidence organized against NIST CSF 2.0, with the explicit aim of informing organizational implementation priorities and identifying where the evidence base is critically absent.

Objectives

This scoping review aims to systematically map ML applications to health care cybersecurity tools. The specific objectives are to:

1. Identify and categorize ML techniques developed in health care cybersecurity
2. Map ML applications across the six NIST CSF 2.0 functions (Identify, Protect, Detect, Respond, Recover, and Govern)
3. Characterize the health care environments and systems addressed by these applications
4. Analyze the methodological approaches to cyber resilience and the geographic distribution of research
5. Identify implementation considerations and highlight research gaps to inform future ML cybersecurity development in health care settings.

Methods

The structure of this scoping review was guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Checklist 1). In addition, search reporting adhered to the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) guidelines [23]. A completed PRISMA-S checklist is provided in Checklist 2. The review was registered with Open Science Framework (OSF Registries 4cjmu).

Studies were selected based on the Population, Intervention, Comparison, Outcome, Study design criteria. Eligible studies were those involving organization-level cybersecurity in a health care setting, including hospitals, clinics, and health networks, where an ML-based approach was applied to enhance cyber resilience through predictive analytics, anomaly detection, and automated response. Outcomes of interest included assessments of data privacy, protection of health care data, or strengthening of cybersecurity practices. Studies primarily focused on smart or wireless health care devices or blockchain-based cybersecurity were excluded, as were nonpeer-reviewed sources including gray literature, conference abstracts, books, editorials, commentaries, case reports, review articles, and theoretical or technical concept papers without empirical evaluation.

A total of 3 electronic databases were searched: Ovid MEDLINE, Embase Classic+ Embase, and Scopus on July 30, 2025. MEDLINE and Embase Classic+ Embase were searched simultaneously via the Ovid platform as a multidatabase search; Scopus was searched separately via its native interface. Ovid MEDLINE and Embase provide comprehensive coverage of the peer-reviewed health care and biomedical literature, while Scopus offers broad multidisciplinary indexing encompassing computer science, engineering, and cybersecurity.

The search strategy was developed de novo in consultation with an expert librarian at Imperial College London and combined both medical subject headings terms for MEDLINE- and Emtree-controlled vocabulary terms for Embase with free-text terms. Boolean operators (AND, OR) were used to combine concept blocks; truncation using the asterisk wildcard and adjacency operators (adj1, adj2) were applied. Keywords included variations of the following: "cybersecurity," "cyber resilience," "healthcare," "health system," "artificial intelligence," "machine learning," "data privacy," "vulnerability," "threat," and "risk." Searches were limited to English language and human studies between 2019 and 2025. Additional studies were identified through reference list screening from included articles. No websites, tables of contents, or print sources were purposefully browsed, and no authors, experts, or manufacturers were contacted to identify additional studies. Gray literature, study registries, and other sources were not searched, and formal peer review of the search strategy was not conducted. Following reviewer and editor feedback, the search strategy was revised and rerun during the revision process; however, the search end date of

July 30, 2025, was retained as the fixed time horizon for the review. No email alerts were used. Full search strategies for all 3 databases are available in Multimedia Appendix 1.

Study selection was conducted in two phases using the Covidence platform (developed by Veritas Health Innovation). A total of two independent reviewers (KR and SZ) conducted titles or abstracts screening and full-text screening. Any discrepancies were resolved through discussion. In keeping with scoping review methodology, no formal critical appraisal or risk of bias assessment was conducted. Data extraction elements were defined prospectively and refined following preliminary searches to provide a comprehensive overview of ML applications in health care cybersecurity and to capture the breadth of study designs, health care settings, ML techniques, and NIST framework classifications. A standardized data extraction form was developed in Microsoft Excel and validated by research team members with expertise in health care cybersecurity and AI (SG and WO). Data extraction was completed by one author (KR) and independently checked by a second reviewer (SZ), with discrepancies resolved through discussion. The extracted information included study characteristics (lead author, year of publication, study design, and country of authorship), AI model type, training data source, performance metrics (eg, accuracy, precision, recall, F_1-score, and area under the receiver operating characteristic), and reported ML-based cyber resilience domain, such as anomaly detection, data privacy protection, intrusion detection, threat analysis, and risk assessment. Performance metrics were extracted as reported by study authors; given heterogeneity in metrics used, validation approaches, and dataset types across studies, cross-study comparison of performance figures was not reported.

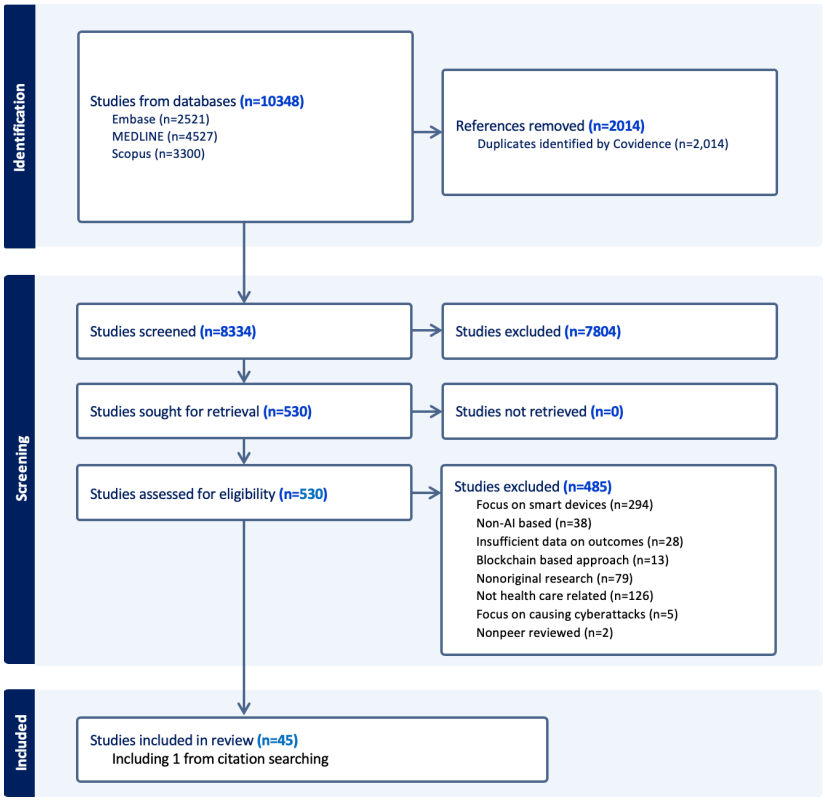
Studies were systematically categorized and mapped against the NIST CSF 2.0 functions as the organizing structure, classifying evidence across the six core functions: Identify, Protect, Detect, Respond, Recover, and Govern. Classification was completed by one author (KR) and independently checked by a second reviewer (SZ), with discrepancies resolved through discussion. Each study was assigned to a single NIST CSF 2.0 function based on its primary cybersecurity objective. Within each thematic domain, studies were compared by ML technique, health care environment, cybersecurity application, study design, and key findings. Results are presented using a combination of narrative description and visual representations.

Results

Study Characteristics

There were 10,348 articles identified from the database search. After removing 2014 duplicates, the remaining records were screened by title and abstract by two independent reviewers (KR and SZ), resulting in 530 articles selected for full-text review. Figure 1 summarizes the PRISMA-ScR flowchart of the study.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses for Scoping Reviews) flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of peer-reviewed studies examining machine learning applications in health care cybersecurity, drawn from systematic searches of Ovid MEDLINE, Embase, and Scopus conducted on July 30, 2025 (search period: 2019-2025).



Following independent full-text review, 45 [24-68] global studies (including 1 from citation search) were finalized for analysis. These were conducted across 18 countries and applied 80 ML models to diverse datasets in health care cybersecurity. Table 1 summarizes the study characteristics of the included studies. China (7/45, 15.6%) and India (7/45, 15.6%) had the highest representation among included studies. The studies were conducted across diverse health care environments, most commonly health care systems (17/45, 37.8%), followed by electronic health records (7/45, 15.6%), medical data (6/45, 13.3%), health care networks (5/45, 11.1%), and others. ML models were trained on diverse sources, including real-world health care data, cybersecurity or network datasets, and synthetic or benchmark datasets. Included studies varied considerably in dataset size and type, encompassing network traffic samples, patient records, email corpora, NLP token sets, and cyberattack event logs, reflecting the heterogeneous data environments across health care cybersecurity research.

Table 1. Characteristics of 45 peer-reviewed studies included in a scoping review of machine learning applications in health care cybersecurity, identified through systematic searches of Ovid MEDLINE, Embase, and Scopus on July 30, 2025 (2019-2025), categorized by study design, country of first author, and health care environment.

Category	Number of studies (N=45), n (%)	References
Study type		
Quasi-experimental study	41 (91.1)	Silvestri et al [24], Islam et al [25], Li and Wang [26], Yi et al [27], Yusof et al [28], Liu et al [29], Sheu et al [30], Nguyen et al [31], Kumar et al [32], Almousa et al [33], S and S [34], Gao et al [35], Altalla' et al [36], Mesfer Aldossary [37], Akter et al [38], Jonnagaddala et al [39], Samiayya et al [40], Tabassum et al [41], Hurst et al [42], Saranya et al [43], Öztürk et al [44], Wazid et al [45], Sengan et al [46], Thanh et al [47], Alanazi [48], Ahmed and Shakir [49], Halman and Alenazi [50], Asif et al [51], Qiao et al [52], Fernández Maimó et al [53], Abidi et al [54], Huang et al [58], An et al [59], Bo et al [60], Akter et al [62], Ganguli et al [63], Alzubi et al [64], Cabrero-Holgueras et al [65], Mohammadi et al [66], Goldschmidt et al [67], Gwon et al [68].
Observational study	4 (8.9)	Islam et al [25], Ünözkan et al [55], Dolezel et al [56], Gupta et al [57].
Country of first author		
China	7 (15.6)	Li et al [26], Yi et al [27], Gao et al [35], Qiao et al [52], Huang et al [58], An et al [59], Bo et al [60].
India	7 (15.6)	S and S [34], Samiayya et al [40], Saranya et al [43], Wazid et al [45], Sengan et al [46], Gupta et al [57], Kaur et al [61].
Saudi Arabia	5 (11.1)	Almousa and Uliyan [33], Mesfer Aldossary [37], Alanazi [48], Halman and Alenazi [50], Abidi et al [54].
Australia	5 (11.1)	Liu et al [29], Nguyen et al [31], Akter et al [38], Jonnagaddala et al [39], Akter et al [62].
United Kingdom	3 (6.7)	Islam et al [25], Tabassum et al [41], Hurst et al [42].
Turkey	2 (4.4)	Öztürk et al [44], Wazid et al [45].
United States	2 (4.4)	Dolezel et al [56], Ganguli et al [63].
Jordan	2 (4.4)	[36,64] Altalla' et al [36], Alzubi et al [64].
Spain	2 (4.4)	Fernández Maimó et al [53], Cabrero-Holgueras [65].
Other countries	10 (22.2)	Silvestri et al [24], Yusof et al [28], Sheu et al [30], Kumar et al [32], Thanh et al [47], Ahmed and Shakir [49], Asif et al [51], Mohammadi et al [66], Goldschmidt et al [67], Gwon et al [68].
Health care environment		
Health care systems	17 (37.8)	Akter et al [38], Jonnagaddala et al, Öztürk et al [44], Wazid et al [45], Sengan et al [46], Alanazi [48], Ahmed and Shakir [49], Halman and Alenazi [50], Asif et al [51], Abidi et al [54], Ünözkan et al [55], Dolezel et al [56], An et al [59], Kaur et al [61], Akter et al [62], Ganguli et al [63], Mohammadi et al [66].
Electronic health records	7 (15.6)	Liu et al [29], Kumar et al [32], Tabassum et al [41], Hurst et al [42], Gupta et al [57], Goldschmidt et al [67], Gwon et al [68]. Jonnagaddala et al [39].
Medical data	6 (13.3)	Sheu et al [30], Nguyen et al [31], Gao et al [35], Altalla' et al [36], Alzubi et al [64], Cabrero-Holgueras et al [65].
Health care networks	5 (11.1)	Li et al [26], Samiayya et al [40], Saranya et al [43], Thanh et al [47], Qiao et al [52].
Medical images and imaging systems	4 (8.9)	S and S [34], Mesfer Aldossary [37], Huang et al [58], Bo et al [60].
Cyber-physical systems	3 (6.7)	Silvestri et al [24], Yi et al [27], Fernández et al [53].
Health care infrastructure	1 (2.2)	Yusof et al [28].
Employee email	1 (2.2)	Almousa et al [33].
Supply chain	1 (2.2)	Islam et al [25].

Figure 2 presents the annual publication volumes. Notably, between 2019 and 2021, the number of publications examining ML and cybersecurity in health care remained consistently low. However, 2022 marked a turning point, with

a sharp increase in publications. This upward trend persisted through 2024.

The distribution of ML techniques across cybersecurity applications is summarized in Figure 3. Classical ML (n=33) and DL (n=27) demonstrated the widest application breadth, with both predominantly used for privacy prevention (n=14) and intrusion detection (n=15). Overall, intrusion detection represented the most common application across all techniques (n=29).

Figure 2. Annual publication trend of 45 peer-reviewed studies examining machine learning applications in health care cybersecurity, identified through systematic searches of Ovid MEDLINE, Embase, and Scopus (search period: 2019-2025; search date: July 30, 2025), illustrating year-on-year growth in research output from 2019 to 2025.

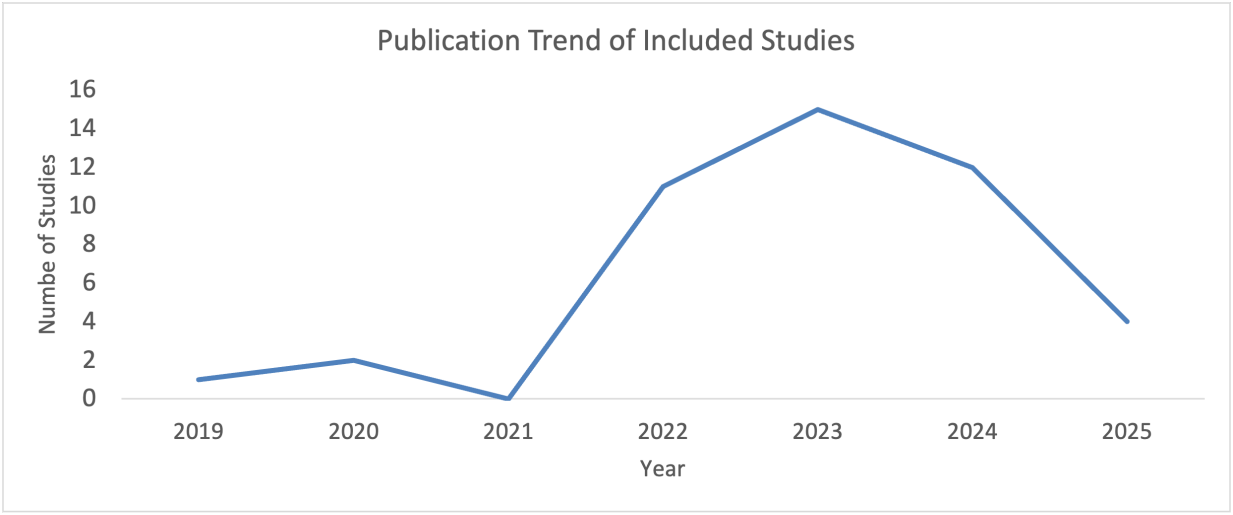
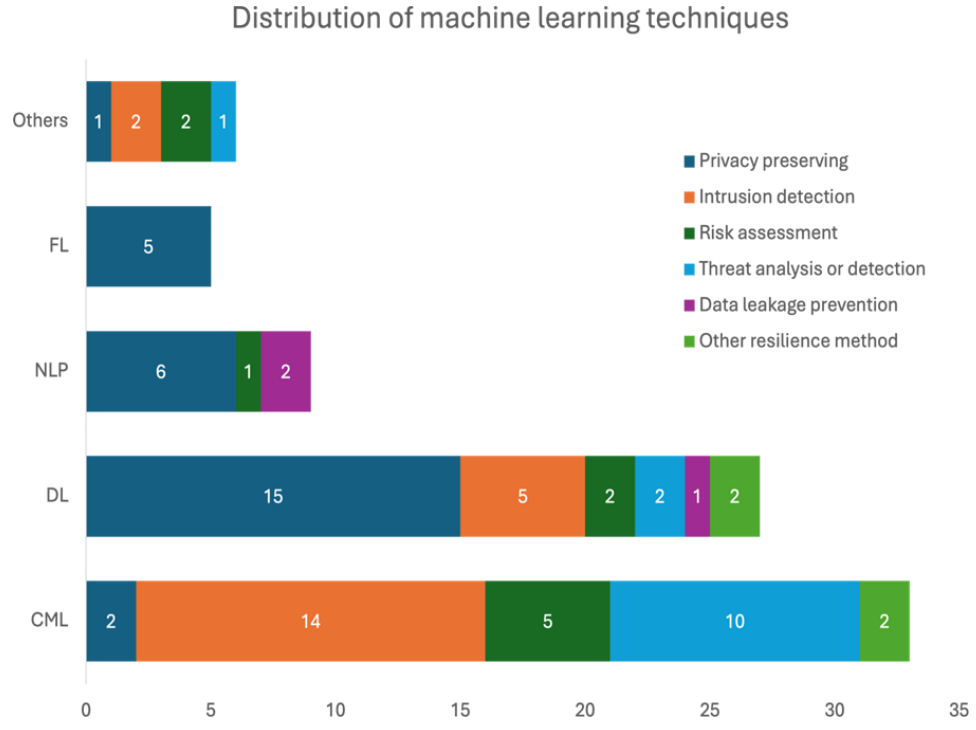


Figure 3. Distribution of machine learning technique categories (classical machine learning, deep learning, natural language processing, and federated learning) across cybersecurity application domains (privacy-preserving, intrusion detection, risk assessment, threat analysis, and data leakage prevention) in 45 peer-reviewed studies of health care cybersecurity (2019-2025). Multiple ML models were applied in most studies; therefore, the total number of ML techniques (n=80) exceeds the number of reviewed papers (n=45). CML: classic machine learning; DL: deep learning; FL: federated learning; NLP: natural language processing.



Distribution of ML Applications Across NIST Cybersecurity Framework Functions

The distribution reveals a pronounced concentration in preincident functions, with Protect, Detect, and Identify collectively accounting for 43 [23-52, 55-67] of the 45 included studies (95.6%). The "Protect" function (22/45, 48.9%) emerged as the most prominent area of ML application, with applications centered on privacy-preserving technologies, data leakage prevention, sensitive information identification, physical isolation mechanisms, asynchronous transfer protocols, and antispoofing measures. Notable among these were privacy-preserving ML techniques, including federated learning approaches. The "Detect" function (13/45, 28.9%) represented the second major application area, encompassing intrusion detection and mitigation, anomaly-based threat detection, spear phishing attack detection, and impact mitigation. The "Identify" function (8/45, 17.8%) showed moderate representation, with studies focusing on risk assessment and prioritization, predictive analytics, threat prediction and analysis, vulnerability detection, and predictive analysis models. These applications demonstrated a proactive approach to cybersecurity through anticipatory capabilities.

The "Respond" function received limited attention, with only two specific applications identified: ransomware spread mitigation and threat detection and blocking. The "Recover" and "Govern" functions received no representation from the reviewed literature (Figure 4). Classical ML and DL dominated across all represented functions, concentrated within Protect and Detect; no ML technique was applied to the Govern or Recover functions (Figure 5). Detailed findings for the studies included in each function are presented in the following sections. Multimedia Appendix 2 [24-68] provides a table that summarizes the included studies with the lead author, year of publication, and main author country of authorship.

Figure 4. Distribution of 45 peer-reviewed studies examining machine learning–based cyber resilience tools in healthcare settings (2019-2025) across the six core functions of the National Institute of Standards and Technology Cybersecurity Framework, version 2.0 (NIST CSF 2.0) [16]: Identify, Protect, Detect, Respond, Recover, and Govern. Studies were identified through systematic searches of Ovid MEDLINE, Embase, and Scopus conducted on July 30, 2025. NIST: National Institute of Standards and Technology.

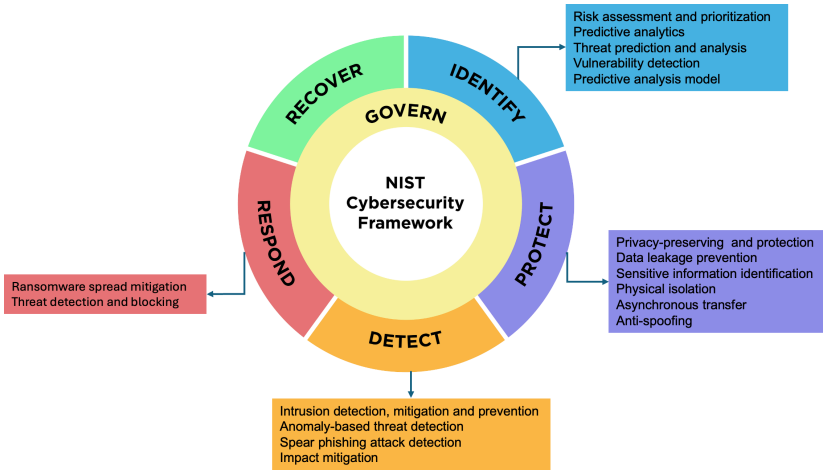
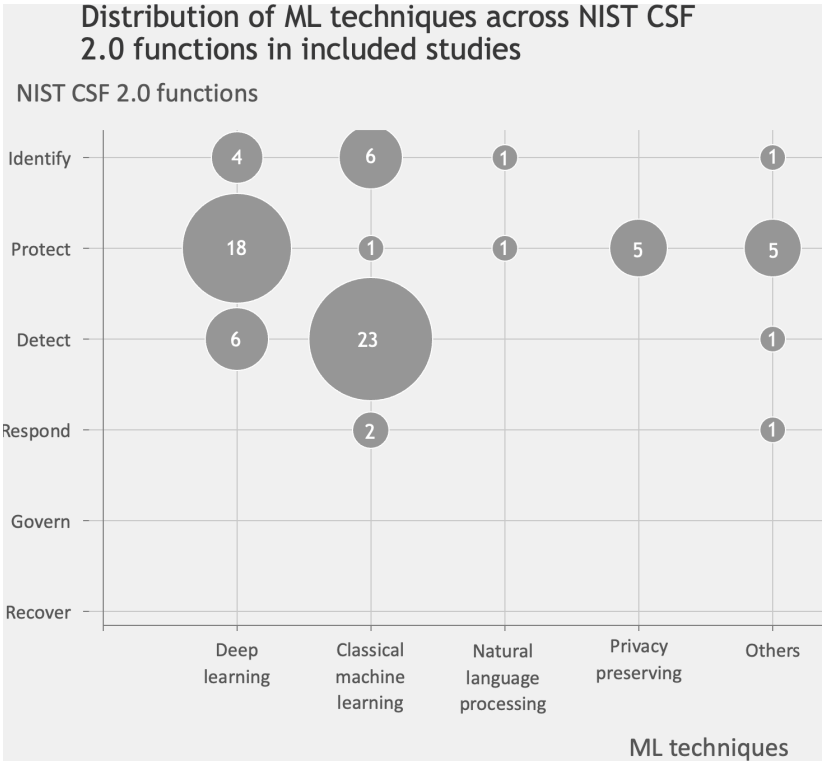


Figure 5. Bubble chart mapping machine learning technique categories against NIST CSF 2.0 (National Institute of Standards and Technology Cybersecurity framework 2.0) functions, with bubble size proportional to the number of included studies (range: 1-23). Classical machine learning applied to threat detection (n=23) and deep learning applied to protective controls (n=18) dominated the evidence base. No studies addressed the Govern or Recover functions, representing critical gaps in the application of machine learning to health care cybersecurity governance and resilience. CST: Cybersecurity framework; ML: Machine learning; NIST: National Institute of Standards and Technology [16].



Protect Function

Of the 45 studies, 22 studies [29-39,57-60,62-68] focused on the "Protect" function, collectively encompassing 295,201 patient records, 1,073,101 public data samples, 38,514 discharge summaries, and 100 test instances. The main focus was on privacy-preserving and data protection techniques (n=19), with some studies addressing physical or network isolation and antispoofing measures.

Techniques clustered around cryptographic and decentralized learning approaches include computation on encrypted data [32] using homomorphic encryption (HE) variants, attribute-based schemes, and elliptic-curve methods [29,39,57]. Deidentification pipelines such as Open De-Identification and hybrid NLP or ML approaches that remove or replace sensitive health information with realistic surrogates. Generative models such as generative adversarial networks (GANs), deepGAN [34,37], and local differentially private GANs [68] that produce synthetic medical data unlinked to individuals. Additionally, federated learning (FL), which enables collaborative model training across decentralized data sources without centralizing raw data, is further enhanced by differential-privacy extensions [31,66].

HE enabled computation on encrypted data (including packed [65] and fully HE variants [30] for medical images); FL and decentralized or relay frameworks preserved data sovereignty while supporting collaborative model training [38,59,60]. Ensemble techniques targeted data leakage and integrity (NLP and artificial neural network) [67] and automated phishing detection (Multinomial Naïve Bayes, support vector machine [SVM], convolutional neural network [CNN], long short-term memory [LSTM]) [33]. Network flow watermarking techniques have been used for real-time privacy and data leakage detection methods in digital health care systems, with particular application in medical image protection [58]. ChatGPT-3.5 and GPT-4 (OpenAI) were tested for clinical note deidentification and synthetic data generation [36]. Unlike locally deployed open-weight models, these are application programming interface–based systems without publicly available weights; testing on clinical text therefore involves transmitting potentially identifiable data to third-party infrastructure, a data governance exposure that is absent when using locally hosted alternatives.

Reported outcomes were generally strong, with FL studies commonly reporting accuracies in the 90%-99% range, under controlled or simulated conditions [31,38,59,62,66]. Open De-Identification and NLP pipelines achieved high precision and F_1-scores (eg, Open De-Identification: precision 95.4%, F_1-score 92.2%) [29,39,57,67,68]. Several GAN or large language model approaches reported accuracies over 82% up to ≈99% [35,37,60,63]. Large-scale phishing detection achieved ≈99.4% accuracy across 525,754 emails, as reported by study authors [33]. Notably, an Open De-Identification pipeline was reported to be deployed in real time at a tertiary hospital and demonstrated more than 90% performance on privacy tasks using real EHR datasets [29]. Reported outcomes suggest strong technical performance for privacy tasks; however, evaluations were heterogeneous, metrics were

inconsistently reported, and few studies included external or operational validation.

Detect Function

There were 13 studies belonging to the "Detect" function. These include a total of 1,070,777 EHR entries, 41,403 attack events, and 777,983 public data records [40-52]. Cyber resilience techniques include intrusion detection (n=8), intrusion detection with mitigation or prevention (n=2), and other single-instance methods such as anomaly, insider-threat, spear phishing, and impact-mitigation detection.

Classical ML techniques such as SVM, decision tree (DT), random forest (RF), K-nearest neighbor (KNN), Naïve Bayes, logistic regression, and gradient-boosted learners such as extreme gradient boosting (XGBoost) or adaptive boosting [42,44,49,50] formed the backbone of many studies and were used to detect EHR data misuse and to classify different attack types. Ensemble and hybrid schemes combined multiple classical learners (eg, a combination of SVM, KNN, DT, or RF ensembles) and reported that ensemble approaches delivered competitive accuracy with added stability across attack types in simulated evaluations [45]. DL Architectures (CNNs, LSTMs, gated recurrent units) and bespoke neural networks such as ImmuneNet and hybrid deep frameworks predominated where high-dimensional or sequential inputs were required (network flows, email text, and time-series vitals) [40,47,48]. These models frequently produced near-perfect benchmark scores (many reports more than 99% accuracy on public datasets under controlled testing conditions) at the cost of greater computational complexity. Unsupervised and semisupervised methods (isolation forest, local outlier factor, and other anomaly detectors) were used either standalone for novelty detection or as prelabeling steps to bootstrap supervised training; studies that combined unsupervised labeling with downstream classifiers showed strong sensitivity for contextual anomalies [41,45]. Studies also explored specialized multimodal fusion architectures integrating CNN, LSTM, or gated recurrent units feature streams for richer representation detection rate on the publicly available dataset, and bio-inspired or swarm-intelligence methods (eg, swarm-based network watermarking) for robust, low-overhead detection and real-time data-leakage protection [52].

Identify Function

There were eight studies in the "Identify" function, including 352 cyberattack events, 51 expert annotations, 514,220 NLP tokens, 1000 execution paths, 65,536 network packets, and 825 county-level data points [24-28,55,56,61]. Studies focused primarily on risk assessment (n=5) and predictive analysis (n=3).

Studies demonstrated ML applications across cybersecurity risk assessment and vulnerability prediction in health care settings. Predictive models for hospital data breaches and security vulnerability assessment and their potential to evolve into usable exploits were developed from classical learners (KNN, SVM, DT, RF) [25,55,56]. Probabilistic approaches such as Bayesian networks [28] and fuzzy or knowledge-based systems (adaptive neuro-fuzzy inference system) [61] were applied to security risk assessment during health care web application development, reporting strong performance on the datasets evaluated. Deep neural networks (stacked autoencoders and deep neural networks) developed a hospital computer network information security risk assessment [26,28]. An intelligent vulnerability detector for healthcare cyber-physical systems used recurrent neural network/LSTM architectures to identify vulnerabilities [27]. NLP pipelines were applied to large text corpora to extract threat and vulnerability intelligence [26]. However, these studies were simulation-based, used heterogeneous datasets, and inconsistently reported evaluation metrics, limiting cross-study comparability. Where provided, point estimates included KNN ≈87% (prediction accuracy), a top NLP study reporting 99.8% accuracy (with precision 96.6% and F_1-score 87.5%), a Bayesian network reporting 75% accuracy (recall 73%), and an SVM model achieving 83.1% accuracy (precision 65.1%, recall 58.3%, and F_1-score 61.5%).

Respond Function

A total of Two studies belonged to the "Respond" function of the NIST-CSF 2.0. One study explored ransomware spread mitigation involving 150,537 ransomware datasets, and the other focused on intrusion mitigation using 16,000 patient vital signs [53,54]. A real-time ransomware detection system used SVM for anomaly detection and Naïve Bayes for classification. Integrating software-defined networking and network function virtualization, it isolated and replaced infected systems, mitigating attacks in under 30 seconds, faster than the typical 1-minute spread [53]. The second study proposed a crossover-based multilayer perceptron model–detected attacks on EHR and promptly alerted health care professionals to prevent data leakage. Experimental results on synthetic and real-world datasets showed the crossover-based multilayer perceptron model outperformed existing methods using DL and CNN, with 97% accuracy, 93% precision, 92% recall, and 92% F_1-score, while also exhibiting lower computational complexity [54].

Discussion

Principal Findings

This review maps ML-driven cybersecurity applications in health care against the NIST CSF 2.0 to identify where evidence clusters and gaps persist. In doing so, it provides the first structured, health care–specific synthesis of ML cybersecurity evidence aligned to this framework. Classical algorithms, DL, NLP, and FL were the predominant technique categories, applied across diverse health care environments internationally, with research output accelerating sharply from 2022 [69-71]. The evidence base is heavily concentrated in preincident functions, with Protect, Detect, and Identify collectively accounting for the substantial majority of included studies, while Respond was underrepresented; Recover and Govern remain entirely absent. Despite high technical accuracy under controlled

conditions, real-world deployment was almost entirely absent, with barriers around dataset generalizability, explainability, and regulatory readiness persisting across the literature. These findings have direct implications for how health care organizations might prioritize ML adoption.

Overall Contribution and Landscape of Included Studies

Prior reviews by Wiafe et al [17], Ali et al [18], and Dhanushkodi and Thejas [20] examined ML and AI for cybersecurity across general computing environments, cataloging techniques such as intrusion detection, malware analysis, and anomaly detection, but none were specific to health care or mapped findings against an established governance framework [17-20]. The global landscape of medical AI has evolved rapidly, with China emerging as a leading contributor in health care AI research, surpassing the United States in the number of publications over the past two years [72]. Nonetheless, this review also likely underestimates China's contributions because of the language limitations in the search strategy.

This growth in research output reflects a scientific response to an expanding threat landscape, further accelerated by the COVID-19 pandemic, which drove rapid digitization of mission-critical health care services [73,74]. The shift to interconnected health infrastructure, telehealth, and widespread adoption of connected medical devices expanded the attack surface. The escalation of cyberattacks from 2022 onward, fueled by the high value of patient data for identity theft and insurance fraud, further intensified the urgency for ML-driven solutions [75,76]. Health care organizations have been identified as prime targets for financially motivated attackers, partly because the criticality of patient care services creates pressure to restore operations rapidly [77].

The dominance of the "Protect" function is consistent with the heightened regulatory and ethical scrutiny surrounding patient data in health care, where data protection obligations create strong institutional incentives for this class of solution [78]. The prominence of the Detect function likely reflects the technical tractability of intrusion detection as a well-defined, benchmarkable problem with established datasets [79]. By contrast, the near-absence of Respond and the complete absence of Recover and Govern studies suggest that the research literature has not yet engaged substantively with the organizational dimensions of cyber resilience through ML approaches [2,16,69].

Across the broader cybersecurity literature, governance tasks such as risk monitoring, automated compliance, and auditability have received comparatively limited attention relative to detection and protection [80,81]. On the other hand, studies in other critical infrastructures have demonstrated automated root cause analysis and self-healing mechanisms achieving zero-downtime recovery [82], highlighting the translational opportunity for health care. Governance and risk management are therefore essential to guide the responsible deployment of ML models and organizational resilience [16].

Performance of ML Techniques

DL and hybrid models consistently demonstrated strong detection performance in controlled evaluations, with CNN and LSTM architectures predominating, consistent with the systematic review by Ali et al [18], though results derive predominantly from synthetic or benchmark datasets and lack operational validation. Similarly, classical learners (RF, DT, and SVM) remained competitive in controlled evaluations, though real-world deployment performance was not assessed in the included studies [40,41,47-49]. Supervised ML algorithms, particularly ensemble SVMs and ANNs, have dominated academic literature. Similar findings were reported by Mukkamalla et al [83]. Likewise, the broader mapping study by Wiafe et al [17] identified rising SVM usage between 2013 and 2018, attributed to their suitability for small, high-dimensional, and nonlinear datasets.

Ensemble deidentification approaches, HE combined with other ML models, and FL were the most frequently adopted techniques. FL allows different institutions to train ML models on their local data and share only model parameters with a central server or other nodes, rather than sensitive raw data [31,66,84]. FL has been proposed as particularly suited to dynamic health care networks because it enables real-time updates and scales across institutions without raw data sharing. However, challenges such as institutional data heterogeneity, communication costs, model aggregation vulnerabilities, and potential biases limit real-world reliability [85]. These aggregation vulnerabilities include gradient reconstruction attacks, in which shared model updates are exploited to recover original training data, and model poisoning, in which malicious local updates corrupt the global model [86,87]. Combining FL with HE strengthens the privacy-preserving architecture, as model updates remain encrypted during transfer [88].

Proactive risk assessment applications demonstrated that both data-driven and knowledge-based ML approaches can identify vulnerabilities and predict breaches in health care settings, though all evaluations were simulation-based and none approached operational readiness [25,26,61]. In health care network environments, where the vast majority of traffic is benign, class imbalance means that high accuracy can be achieved even by models that fail to detect most attacks [89]. Metrics such as precision, recall, and F_1-score provide a more complete picture, but these too were inconsistently reported across the included studies [81]. The asymmetric cost of errors is particularly consequential in clinical settings: false negatives carry direct patient safety implications, while high false positive rates risk alert fatigue among already stretched clinical and information technology teams, potentially causing genuine threats to be overlooked [90]. The absence of standardized, clinically contextualized evaluation frameworks across the included studies therefore limits the extent to which reported performance figures can be used to judge real-world effectiveness [91].

Barriers to Real-World Applicability

Many studies relied on outdated or non–health care network datasets (CICIDS-2017, NSLKDDCup99, Kaggle), with the oldest exceeding over 20 years old [92]. Methodological inconsistencies, including nonstandardized metrics, varying preprocessing techniques, undocumented hyperparameter tuning, and an absence of benchmarking against established cybersecurity tools, further undermine reproducibility [81,89].

These patterns have substantive implications for the maturity and reliability of the evidence base. First, the near-exclusive reliance on synthetic or benchmark datasets raises questions about generalizability, as models optimized on static, historical network traffic data are unlikely to perform reliably against the dynamic, heterogeneous threat environments of modern health care infrastructure [92]. Second, the absence of external or prospective validation means that reported performance figures cannot currently be taken as indicators of deployment readiness; high benchmark accuracy is a necessary but insufficient condition for clinical adoption without validation on representative, real-world data [89,91]. Third, heterogeneous evaluation practices, including inconsistent metric selection, varying dataset splits, and undocumented tuning, make cross-study comparison unreliable and limit the conclusions that health care organizations or policymakers can draw [81,89]. The evidence base is therefore best characterized as exploratory rather than confirmatory. It demonstrates that ML techniques can achieve strong results under controlled conditions but are not yet sufficient to establish deployment readiness in clinical environments.

On an institutional level, health care organizations lack the technical infrastructure and interdisciplinary expertise to implement and maintain ML-driven cybersecurity systems effectively [90,93]. This challenge is compounded by the fact that explainable AI, intended to make AI understandable, remains poorly defined, with terms like interpretability and transparency often used interchangeably [94]. Explainable AI has been proposed to build clinician trust, support regulatory approval, and improve communications between clinicians and cybersecurity teams [95]. Several studies noted the absence of explainability as a barrier to clinical trust, pointing to explainable AI as a potential avenue for future research, though its practical integration into health care cybersecurity remains undemonstrated in the included literature [30,48].

Ethical, regulatory, and infrastructural challenges hinder ML adoption in health care [96]. Access to sensitive patient data raises concerns around privacy, consent, and compliance with regulations [15,97,98]. Rapid development and simulation of ML models have outpaced regulatory frameworks, creating a gray area that increases compliance burdens and delays clinical deployment. Recent policy initiatives, including the National Health Service Innovation Accelerator [99] and the US AI Action Plan [15], signal growing governmental recognition of AI's role in health care cybersecurity and could provide implementation pathways for organizations navigating this landscape. Parallel efforts from the European Commission's hospital cybersecurity action plan and the WHO's AI for health framework similarly reflect an emerging international consensus on AI-enabled cyber resilience in health care [100,101].

Limitations of This Review

This review has several limitations. First, language and publication bias may have excluded non-English or nonindexed studies, and gray literature was not systematically searched. Second, the rapid evolution of ML and cyber threats means findings become outdated within a short timeframe. Third, heterogeneity in study designs, datasets, and evaluation metrics limited direct cross-study comparison. Fourth, the review excluded medical devices to focus on institutional cybersecurity, potentially omitting insights into device-specific risks, and proprietary AI-based cybersecurity products were not included due to limited methodological transparency, which may restrict understanding of current real-world implementations. Additionally, classification of studies against NIST CSF 2.0 functions was conducted by the review team and, while guided by each study's primary cybersecurity objective, remains inherently interpretive. Studies addressing multiple cybersecurity tasks may reasonably be mapped to more than one function, and alternative classifications by other reviewers are possible. Finally, the search was conducted till July 30, 2025, as prespecified in the registered protocol (OSF Registries 4cjmu); studies published after this date are not captured, and an updated review is recommended within two years given the pace of publication in this field.

Future Directions and Operational Implications

Health care organizations face the dual challenge of implementing ML-driven cybersecurity while managing legacy infrastructure, limited budgets, and regulatory uncertainty. Current national and international cybersecurity strategies consistently call for AI-driven detection, privacy-preserving architectures, and governance frameworks, yet this review demonstrates that the research base supporting the latter two remains markedly underdeveloped [14,15,99].

Based on the distribution of evidence identified in this review, a set of priorities for pilot testing can be outlined for health care organizations. These priorities are presented as hypotheses for consideration rather than a validated deployment model. The relative maturity of evidence in the "Detect" function, with intrusion detection representing the most studied application, represents a more tractable starting point for organizations considering ML pilot testing [40-42,44,45,47-50,52]. However, as Heine et al [91] highlight, despite excellent results on benchmark datasets, ML-based intrusion detection systems remain difficult to deploy in practice, with current simulation-based evaluation methodologies poorly reflecting real-world performance.

The growing volume of studies addressing privacy-preserving techniques, particularly FL and HE, indicates an emerging evidence base that may support a subsequent pilot-testing priority, particularly for organizations managing multisite data under strict privacy regulation, though

operational implementation evidence remains sparse [31,66,92]. The absence of any studies addressing the Govern or Recover functions represents the most significant gap identified by this review, underscoring that governance frameworks and recovery protocols for ML-driven cybersecurity in health care remain open research priorities rather than established practice.

The absence of real-world implementation evidence has direct resource implications for health care organizations considering ML adoption. These span AI procurement, infrastructure upgrades, staff training, and ongoing model monitoring, maintenance, and external validation. Pilot programs with careful cost tracking, rigorous documentation, and a primary focus on safety are needed to bridge the gap between experimental research and operational deployment. Without this foundation, policy ambition will continue to outpace the evidence needed to deploy ML-driven cybersecurity safely in clinical environments. The gap between policy ambition and operational evidence is unlikely to remain static. Anticipated advances in quantum computing represent a longer-term threat vector with the potential to render current cryptographic protections obsolete, reinforcing the case for sustained investment in health care–specific cybersecurity research and the development of adaptive defensive architectures [102-104].

Conclusions

By mapping ML-driven cybersecurity in health care against all 6 functions of the NIST CSF 2.0, this review offers a governance-aligned synthesis that complements prior reviews organized by attack type or ML architecture [17-20,105]. The NIST CSF 2.0 framework, commonly used to evaluate organizational cyber resilience, makes findings translatable into implementation priorities for organizations. Evidence clusters in Identify, Protect, and Detect; Respond, Recover, and Govern remain unaddressed; and real-world deployment was almost entirely absent. These later functions precisely determine whether a health care organization can contain, recover from, and build lasting organizational cyber resilience from an attack.

The relative maturity of intrusion detection and privacy-preserving methods offers a defensible entry point for ML adoption, but organizations pursuing end-to-end cyber resilience cannot yet draw on a validated evidence base for governance or recovery. Until sustained investment in governance research and real-world validation with representative datasets addresses these gaps, ML in health care cybersecurity will remain a promising experimental tool rather than a trusted operational safeguard.

Funding

The authors declared no financial support was received for this work.

Data Availability

This scoping review synthesizes existing published literature; no new primary data were collected. All data extraction and synthesis are presented in the manuscript tables and figures.

Authors' Contributions

Conceptualization, methodology, investigation, data curation, formal analysis, visualization, writing – original draft, writing – review & editing, project administration: KR.
SZ: Methodology, Investigation, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. SG: Conceptualization, Validation, Writing – review & editing, Supervision. ME: Writing – review & editing, Supervision. WO: Writing – review & editing, Supervision. AD: Writing – review & editing, Supervision. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed search strategy, inclusion and exclusion criteria.
[DOCX File (Microsoft Word File), 25 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Supplementary tables outlining data charting.
[DOCX File (Microsoft Word File), 88 KB-Multimedia Appendix 2]

Checklist 1

PRISMA-ScR Checklist.
[DOCX File (Microsoft Word File), 111 KB-Checklist 1]

Checklist 2

PRISMA-S checklist.
[DOCX File (Microsoft Word File), 19 KB-Checklist 2]

References

1. Coventry L, Branley D. Cybersecurity in healthcare: a narrative review of trends, threats and ways forward. Maturitas. Jul 2018;113:48-52. [doi: 10.1016/j.maturitas.2018.04.008] [Medline: 29903648]
2. Rajput K, Darzi A, Ghafur S. Overlooked and under-reported: the impact of cyberattacks on primary care in the UK National Health Service. Lancet Digit Health. Jul 2025;7(7):100879. [doi: 10.1016/j.landig.2025.100879] [Medline: 40414783]
3. Alder S. Ascension Ransomware Attack Affects 5.6 Million Patients. The HIPAA Journal. Aug 2024. URL: https://www.hipaajournal.com/ascension-cyberattack-2024 [Accessed 2026-07-23]
4. Understanding the Change Healthcare breach and its impact on security compliance. Hyperproof. URL: https://hyperproof.io/resource/understanding-the-change-healthcare-breach [Accessed 2026-06-29]
5. Aldosari B. Cybersecurity in healthcare: new threat to patient safety. Cureus. May 2025;17(5):e83614. [doi: 10.7759/cureus.83614] [Medline: 40486340]
6. Zeadally S, Adi E, Baig Z, Khan IA. Harnessing artificial intelligence capabilities to improve cybersecurity. IEEE Access. 2020;8:23817-23837. [doi: 10.1109/ACCESS.2020.2968045]
7. Guembe B, Azeta A, Misra S, Osamor VC, Fernandez-Sanz L, Pospelova V. The emerging threat of Ai-driven cyber attacks: a review. Appl Artif Intell. Dec 31, 2022;36(1):2037254. [doi: 10.1080/08839514.2022.2037254]
8. Disrupting the first reported AI-orchestrated cyber espionage campaign. Anthropic; 2025. URL: https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf [Accessed 2026-06-29]
9. AI cyber attack statistics 2025. Tech Advisors. 2025. URL: https://tech-adv.com/blog/ai-cyber-attack-statistics [Accessed 2026-06-29]
10. Solaiman BC, CohenI, editors. Cybersecurity of AI medical devices: risks, legislation, and challenges. In: Research Handbook on Health, AI and the Law. Edward Elgar Publishing Ltd; 2024. [doi: 10.4337/9781802205657.ch04]
11. Arefin S, Simcox M. AI-Driven Solutions for Safeguarding Healthcare Data: Innovations in Cybersecurity. IBR. 2024;17(6):74. [doi: 10.5539/ibr.v17n6p74]
12. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. Oct 2018;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
13. Gopalan SS, Raza A, Almobaideen W. IoT security in healthcare using AI: a survey. Presented at: 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA); Mar 16-18, 2021; Sharjah, United Arab Emirates. [doi: 10.1109/ICCSPA49915.2021.9385711]
14. A cyber resilient health and adult social care system in England: cyber security strategy to 2030. Gov.UK. 2023. URL: https://www.gov.uk/government/publications/cyber-security-strategy-for-health-and-social-care-2023-to-2030/a-cyber-resilient-health-and-adult-social-care-system-in-england-cyber-security-strategy-to-2030 [Accessed 2026-06-29]
15. Winning the race: America's AI action plan. Executive Office of the President of The United States. 2025. URL: https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf [Accessed 2026-07-11]
16. Pascoe C, Quinn S, Scarfone K. The NIST cybersecurity framework (CSF) 2.0. National Institute of Standards and Technology; 2024. [doi: 10.6028/NIST.CSWP.29]
17. Wiafe I, Koranteng FN, Obeng EN, Assyne N, Wiafe A, Gulliver SR. Artificial intelligence for cybersecurity: a systematic mapping of literature. IEEE Access. 2020;8:146598-146612. [doi: 10.1109/ACCESS.2020.3013145]
18. Ali R, Ali A, Iqbal F, Khattak AM. A systematic review of artificial intelligence and machine learning techniques for cyber security. In: Aleem S, editor. Presented at: International Conference on Big Data and Security; 2019. [doi: 10.1007/978-981-15-7530-3_44]
19. Ali R, Ali A, Iqbal F, Hussain M, Ullah F. Deep learning methods for malware and intrusion detection: a systematic literature review. Secur Commun Netw. Oct 10, 2022;2022(1):1-31. [doi: 10.1155/2022/2959222]
20. Dhanushkodi K, Thejas S. AI enabled threat detection: leveraging artificial intelligence for advanced security and cyber threat mitigation. IEEE Access. 2024;12:173127-173136. [doi: 10.1109/ACCESS.2024.3493957]
21. Angelo DD, MehtaT. AI provides an Rx for cybersecurity in healthcare. Paloalto. URL: https://www.paloaltonetworks.com/blog/2024/07/ai-provides-an-rx-for-cybersecurity-in-healthcare [Accessed 2026-06-29]
22. Pendyala SK. Strengthening healthcare cybersecurity: leveraging multi-cloud and AI solutions. J Comput Sci Appl Inf Technol. 2025;10(1):1-8. URL: https://symbiosisonlinepublishing.com/computer-science-technology/volume10-issue1-jcsait.php [doi: 10.15226/2474-9257/10/1/00163]
23. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. Jan 26, 2021;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]

24. Silvestri S, Islam S, Papastergiou S, Tzagkarakis C, Ciampi M. A machine learning approach for the NLP-based analysis of cyber threats and vulnerabilities of the healthcare ecosystem. Sensors (Basel). Jan 6, 2023;23(2):651. [doi: 10.3390/s23020651] [Medline: 36679446]
25. Islam S, Abba A, Ismail U, Mouratidis H, Papastergiou S. Vulnerability prediction for secure healthcare supply chain service delivery. Integr Comput Aided Eng. 2022;29(4):389-409. [doi: 10.3233/ICA-220689]
26. Li G, Dong Z, Wang Y. Information security of hospital computer network based on SAE deep neural network. Appl Math Nonlinear Sci. Jan 1, 2024;9(1). [doi: 10.2478/amns.2023.1.00466]
27. Yi X, Wu J, Li G, Bashir AK, Li J, AliA. Recurrent semantic learning-driven fast binary vulnerability detection in healthcare cyber physical systems. IEEE Trans Netw Sci Eng. 2023;10(5):2537-2550. [doi: 10.1109/TNSE.2022.3199990]
28. Yusof MHM, Zin AM, Satar NSM. Behavioral intrusion prediction model on Bayesian network over healthcare infrastructure. Comput Mater Contin. 2022;72(2):2445-2466. [doi: 10.32604/cmc.2022.023571]
29. Liu J, Gupta S, Chen A, et al. OpenDeID pipeline for unstructured electronic health record text notes based on rules and transformers: deidentification algorithm development and validation study. J Med Internet Res. Dec 6, 2023;25:e48145. [doi: 10.2196/48145] [Medline: 38055317]
30. Sheu RK, Lin YC, Pardeshi MS, et al. Adaptive autonomous protocol for secured remote healthcare using fully homomorphic encryption (AutoPro-RHC). Sensors (Basel). Oct 16, 2023;23(20):8504. [doi: 10.3390/s23208504] [Medline: 37896596]
31. Nguyen TV, Dakka MA, Diakiw SM, et al. A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. Sci Rep. May 25, 2022;12(1):8888. [doi: 10.1038/s41598-022-12833-x] [Medline: 35614106]
32. Kumar S, Wajeed MA, Kunabeva R, et al. Novel method for safeguarding personal health record in cloud connection using deep learning models. Comput Intell Neurosci. 2022;2022:3564436. [doi: 10.1155/2022/3564436] [Medline: 35345805]
33. Almousa BN, Uliyan DM. Anti-spoofing in medical employee's email using machine learning uclassify algorithm. Int J Adv Comput Sci Appl. 2023;14(7). [doi: 10.14569/IJACSA.2023.0140727]
34. S G, S G. Securing medical image privacy in cloud using deep learning network. J Cloud Comput. 2023;12(1). [doi: 10.1186/s13677-023-00422-w]
35. Gao X, Mi W, Feng X. Personal health data protection and intelligent healthcare applications under generative adversarial network. Sci Rep. May 13, 2025;15(1):16558. [doi: 10.1038/s41598-025-01575-1] [Medline: 40360631]
36. Altalla' B, Abdalla S, Altamimi A, et al. Evaluating GPT models for clinical note de-identification. Sci Rep. Jan 31, 2025;15(1):3852. [doi: 10.1038/s41598-025-86890-3] [Medline: 39890969]
37. Mesfer Aldossary S. DeepGan-privacy preserving of healthcare system using DL. Intell Autom Soft Comput. 2023;37(2):2199-2212. [doi: 10.32604/iasc.2023.038243]
38. Akter M, Moustafa N, Turnbull B. SPEI-FL: serverless privacy edge intelligence-enabled federated learning in smart healthcare systems. Cogn Comput. Sep 2024;16(5):2626-2641. [doi: 10.1007/s12559-024-10310-3]
39. Jonnagaddala J, Chen A, Batongbacal S, Nekkantti C. The OpenDeID corpus for patient de-identification. Sci Rep. Oct 7, 2021;11(1):19973. [doi: 10.1038/s41598-021-99554-9] [Medline: 34620985]
40. Samiayya D, Vincent P, Srinivasan K, Chang CY, Ganesh H, Kumaar MA. A hybrid framework for intrusion detection in healthcare systems using deep learning. Front Public Health. 2021;9:824898. [doi: 10.3389/fpubh.2021.824898] [Medline: 35096763]
41. Tabassum M, Mahmood S, Bukhari A, Alshemaimri B, Daud A, Khalique F. Anomaly-based threat detection in smart health using machine learning. BMC Med Inform Decis Mak. Nov 19, 2024;24(1):347. [doi: 10.1186/s12911-024-02760-4] [Medline: 39563355]
42. Hurst W, Tekinerdogan B, Alskaif T, Boddy A, Shone N. Securing electronic health records against insider-threats: a supervised machine learning approach. Smart Health. Dec 2022;26:100354. [doi: 10.1016/j.smhl.2022.100354]
43. Saranya N, Abbinavu T, Kumar PA, Gnanasekar R, Prasanth PG, Subha R. Retracted: RTIDS: a robust transformer-based approach for intrusion detection system. Presented at: 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS); Mar 14-15, 2024:1461-1464; Coimbatore, India. [doi: 10.1109/ICACCS60874.2024.10717109]
44. Öztürk T, Turgut Z, Akgün G, Köse C. Machine learning-based intrusion detection for SCADA systems in healthcare. Netw Model Anal Health Inform Bioinforma. Dec 2022;11(1). [doi: 10.1007/s13721-022-00390-2]
45. Wazid M, Singh J, Das AK, Rodrigues JJPC. An ensemble-based machine learning-envisioned intrusion detection in Industry 5.0-driven healthcare applications. IEEE Trans Consumer Electron. 2024;70(1):1903-1912. [doi: 10.1109/TCE.2023.3318850]

46. Sengan S, Khalaf OI, Sharma DK, Hamad AA. Secured and privacy-based IDS for healthcare systems on e-medical data using machine learning approach. Int J Reliab Qual E Healthc. Jul 2022;11(3):1-11. [doi: 10.4018/IJRQEH.289175]
47. Thanh Nguyen P, Dang Bich Huynh V, Dang Vo K, Thanh Phan P, Elhoseny M, Le DN. Deep learning based optimal multimodal fusion framework for intrusion detection systems for healthcare data. Comput Mater Contin. 2021;66(3):2555-2571. [doi: 10.32604/cmc.2021.012941]
48. Alanazi SA. ML-SPAs: fortifying healthcare cybersecurity leveraging varied machine learning approaches against spear phishing attacks. Comput Mater Contin. 2024;81(3):4049-4080. [doi: 10.32604/cmc.2024.057211]
49. Ahmed SS, Shakir HR. Enhancement of secure hospital healthcare monitoring system based–software defined network (SDN) with machine learning. Int J Inform Vis. 2024;8(4):2305. [doi: 10.62527/joiv.8.4.2425]
50. Halman LM, Alenazi MJF. MCAD: a machine learning based cyberattacks detector in software-defined networking (SDN) for healthcare systems. IEEE Access. 2023;11:37052-37067. [doi: 10.1109/ACCESS.2023.3266826]
51. Asif MW, Aqdus A, Amin R, Chaudhry SA, Alsubaei FS, Iqbal S. An efficient intrusion detection system using advanced machine learning techniques in SDN for healthcare system. IEEE J Biomed Health Inform. May 2026;30(5):3651-3664. [doi: 10.1109/JBHI.2025.3530563] [Medline: 40030900]
52. Qiao S, Guo Q, Shi F, et al. SIBW: a swarm intelligence-based network flow watermarking approach for privacy leakage detection in digital healthcare systems. IEEE J Biomed Health Inform. Mar 2026;30(3):1912-1924. [doi: 10.1109/JBHI.2025.3542561] [Medline: 40036416]
53. Fernández Maimó L, Huertas Celdrán A, Perales Gómez ÁL, García Clemente FJ, Weimer J, Lee I. Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments. Sensors (Basel). Mar 5, 2019;19(5):1114. [doi: 10.3390/s19051114] [Medline: 30841592]
54. Abidi MH, Alkhalefah H, Aboudaif MK. Enhancing healthcare data security and disease detection using crossover-based multilayer perceptron in smart healthcare systems. Comput Model Eng Sci. 2024;139(1):977-997. [doi: 10.32604/cmes.2023.044169]
55. Ünözkan H, Ertem M, Bendak S. Using attack graphs to defend healthcare systems from cyberattacks: a longitudinal empirical study. Netw Model Anal Health Inform Bioinform. 2022;11(1):52. [doi: 10.1007/s13721-022-00391-1] [Medline: 36408329]
56. Dolezel D, Beauvais B, Stigler Granados P, Fulton L, Kruse CS. Effects of internal and external factors on hospital data breaches: quantitative study. J Med Internet Res. Dec 21, 2023;25:e51471. [doi: 10.2196/51471] [Medline: 38127426]
57. Gupta S, Liu J, Wong ZSY, Jonnagaddala J. Preliminary evaluation of fine-tuning the OpenDeLD deidentification pipeline across multi-center corpora. Stud Health Technol Inform. Aug 22, 2024;316:719-723. [doi: 10.3233/SHTI240515] [Medline: 39176896]
58. Huang T, Xu J, Tu S, Han B. Robust zero-watermarking scheme based on a depthwise overparameterized VGG network in healthcare information security. Biomed Signal Process Control. Mar 2023;81:104478. [doi: 10.1016/j.bspc.2022.104478]
59. An Z, Zhang J, Jiang Z, Du J, Yin Z, Li C. FedMCC: federated multi-center clustering algorithm to improve privacy healthcare. Methods. Oct 2023;218:94-100. [doi: 10.1016/j.ymeth.2023.07.006] [Medline: 37507060]
60. Bo ZH, Guo Y, Lyu J, et al. Relay learning: a physically secure framework for clinical multi-site deep learning. NPJ Digit Med. Nov 4, 2023;6(1):204. [doi: 10.1038/s41746-023-00934-4] [Medline: 37925578]
61. Kaur J, Khan AI, Abushark YB, et al. Security risk assessment of healthcare web application through adaptive neuro-fuzzy inference system: a design perspective. Risk Manag Healthc Policy. 2020;13:355-371. [doi: 10.2147/RMHP.S233706] [Medline: 32425625]
62. Akter M, Moustafa N, Lynar T, Razzak I. Edge intelligence: federated learning-based privacy protection framework for smart healthcare systems. IEEE J Biomed Health Inform. Dec 2022;26(12):5805-5816. [doi: 10.1109/JBHI.2022.3192648] [Medline: 35857737]
63. Ganguli R, Lad R, Lin A, Yu X. Novel generative recurrent neural network framework to produce accurate, applicable, and deidentified synthetic medical data for patients with metastatic cancer. JCO Clin Cancer Inform. May 2023;7(7):e2200125. [doi: 10.1200/CCI.22.00125] [Medline: 37130342]
64. Alzubi JA, Alzubi OA, Beseiso M, Budati AK, Shankar K. Optimal multiple key-based homomorphic encryption with deep neural networks to secure medical data transmission and diagnosis. Expert Systems. May 2022;39(4). [doi: 10.1111/exsy.12879]
65. Cabrero-Holgueras J, Pastrana S. Towards realistic privacy-preserving deep learning over encrypted medical data. Front Cardiovasc Med. 2023;10:1117360. [doi: 10.3389/fcvm.2023.1117360] [Medline: 37187785]
66. Mohammadi N, Rezakhani A, Haj Seyyed Javadi H, asghari P. FLHB-AC: federated learning history-based access control using deep neural networks in healthcare system. J Inf Syst Telecommun. 2024;12(46):90-104. [doi: 10.61186/jist.44500.12.46.90]

67. Goldschmidt G, Zeiser FA, Righi RDR, Da Costa CA. ARTERIAL: a natural language processing model for prevention of information leakage from electronic health records. Presented at: 2023 XIII Brazilian Symposium on Computing Systems Engineering (SBESC); Nov 21-24, 2023:1-6; Porto Alegre, Brazil. [doi: 10.1109/SBESC60926.2023.10324212]
68. Gwon H, Ahn I, Kim Y, et al. LDP-GAN: generative adversarial networks with local differential privacy for patient medical records synthesis. Comput Biol Med. Jan 2024;168:107738. [doi: 10.1016/j.compbiomed.2023.107738] [Medline: 37995536]
69. Ghafur S, Kristensen S, Honeyford K, Martin G, Darzi A, Aylin P. A retrospective impact analysis of the WannaCry cyberattack on the NHS. NPJ Digit Med. 2019;2(1):98. [doi: 10.1038/s41746-019-0161-6] [Medline: 31602404]
70. Alder S. Hospital ransomware attack results in patient death. The HIPAA Journal. 2020. URL: https://www.hipaajournal.com/hospital-ransomware-attack-results-in-patient-death [Accessed 2026-07-11]
71. Ghafur S, Fontana G, Martin G, Grass E, Goodman J, Darzi A. Improving cyber security in the NHS. Imperial College London's Institute of Global Health Innovation; 2019. URL: https://www.imperial.ac.uk/media/imperial-college/institute-of-global-health-innovation/Cyber-report-2020.pdf [Accessed 2026-07-19]
72. Qiu Y, Ma W, Wang H, Tham YC, Wong TY. The landscape of medical AI in China. NEJM AI. Jun 26, 2025;2(7):AIp2401234. [doi: 10.1056/AIp2401234]
73. Wasserman L, Wasserman Y. Hospital cybersecurity risks and gaps: Review (for the non-cyber professional). Front Digit Health. 2022;4:862221. [doi: 10.3389/fdgth.2022.862221] [Medline: 36033634]
74. He Y, Aliyu A, Evans M, Luo C. Health care cybersecurity challenges and solutions under the climate of COVID-19: scoping review. J Med Internet Res. Apr 20, 2021;23(4):e21747. [doi: 10.2196/21747] [Medline: 33764885]
75. Chigada J, Madzinga R. Cyberattacks and threats during COVID-19: a systematic literature review. S Afr J Inf Manage. 2021;23(1):1-11. [doi: 10.4102/sajim.v23i1.1277]
76. Monteith S, Bauer M, Alda M, Geddes J, Whybrow PC, Glenn T. Increasing cybercrime since the pandemic: concerns for psychiatry. Curr Psychiatry Rep. Mar 3, 2021;23(4):18. [doi: 10.1007/s11920-021-01228-w] [Medline: 33660091]
77. Minnaar A, Herbig FJ. Cyberattacks and the cybercrime threat of ransomware to hospitals and healthcare services during the COVID-19 pandemic. Acta Criminol. 2021;34(3):155-185. URL: https://journals.co.za/doi/10.10520/ejc-crim_v34_n3_a10 [Accessed 2026-07-23]
78. Khalid N, Qayyum A, Bilal M, Al-Fuqaha A, Qadir J. Privacy-preserving artificial intelligence in healthcare: techniques and applications. Comput Biol Med. May 2023;158:106848. [doi: 10.1016/j.compbiomed.2023.106848] [Medline: 37044052]
79. Hady AA, Ghubaish A, Salman T, Unal D, Jain R. Intrusion detection system for healthcare systems using medical and network data: a comparison study. IEEE Access. 2020;8:106576-106584. [doi: 10.1109/ACCESS.2020.3000421]
80. Mohammed A. AI in cybersecurity: enhancing audits and compliance automation. SSRN. Preprint posted online on Feb 11, 2025. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5066097 [Accessed 2026-07-23]
81. Kaur R, Gabrijelčič D, Klobučar T. Artificial intelligence for cybersecurity: literature review and future research directions. Information Fusion. Sep 2023;97:101804. [doi: 10.1016/j.inffus.2023.101804]
82. Jha NN, Manwani P. Self-healing payment systems via AI-driven anomaly recovery: a zero-downtime framework for secure and reliable transactions. Int J Comput Eng Technol. 2025;16(2):200-212. [doi: 10.34218/IJCET_16_02_014]
83. Mukkamala S, Janoski G, Sung A, editors. Intrusion detection using neural networks and support vector machines. Presented at: Proceedings of the 2002 International Joint Conference on Neural Networks IJCNN'02 (Cat No02CH37290); May 12-17, 2002; Honolulu, HI. [doi: 10.1109/IJCNN.2002.1007774]
84. Ghimire B, Rawat DB. Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things. IEEE Internet Things J. 2022;9(11):8229-8249. [doi: 10.1109/JIOT.2022.3150363]
85. Alazab A, Khraisat A, Singh S, Jan T. Enhancing privacy-preserving intrusion detection through federated learning. Electronics (Basel). 2023;12(16):3382. [doi: 10.3390/electronics12163382]
86. Veit A, Estrin D, Hua Y, Shmatikov V, Bagdasaryan E. How to backdoor federated learning. Presented at: 23rd International Conference on Artificial Intelligence and Statistics; Aug 26-28, 2020; Palermo, Italy.
87. Zhu L, Liu Z, Han S. Deep leakage from gradients. Presented at: Proceedings of the 33rd International Conference on Neural Information Processing Systems; Dec 8-14, 2019; Vancouver, Canada.
88. Pulido-Gaytan B, Tchernykh A, Cortés-Mendoza JM, et al. Privacy-preserving neural networks with homomorphic encryption: challenges and opportunities. Peer-to-Peer Netw Appl. May 2021;14(3):1666-1691. [doi: 10.1007/s12083-021-01076-8]
89. Apruzzese G, Laskov P, Montes de Oca E, et al. The role of machine learning in cybersecurity. Digit Threats Res Pract. Mar 31, 2023;4(1):1-38. [doi: 10.1145/3545574]

90. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Futur Healthc J. Jul 2021;8(2):e188-e194. [doi: 10.7861/fhj.2021-0095]
91. Heine F, Laue T, Kleiner C. On the evaluation and deployment of machine learning approaches for intrusion detection. Presented at: 2020 IEEE International Conference on Big Data (Big Data); Dec 10-13, 2020; Atlanta, GA (virtual. [doi: 10.1109/BigData50022.2020.9378479]
92. Moustafa N, Slay J. UNSW-nb15: a comprehensive data set for network intrusion detection systems (UNSW-nb15 network data set). Presented at: 2015 Military Communications and Information Systems Conference; Nov 10-12, 2015; Canberra, Australia. [doi: 10.1109/MilCIS.2015.7348942]
93. Ameen AH, Mohammed MA, Rashid AN. Dimensions of artificial intelligence techniques, blockchain, and cyber security in the internet of medical things: opportunities, challenges, and future directions. J Intell Syst. Jan 12, 2023;32(1). [doi: 10.1515/jisys-2022-0267] [Medline: 20220267]
94. Reynaud S, Roxin A. Review of eXplainable artificial intelligence for cybersecurity systems. Discover Artif Intell. 2025;5(1):78. [doi: 10.1007/s44163-025-00318-5]
95. Zhang Z, Hamadi HA, Damiani E, Yeun CY, Taher F. Explainable artificial intelligence applications in cyber security: state-of-the-art in research. IEEE Access. 2022;10:93104-93139. [doi: 10.1109/ACCESS.2022.3204051]
96. Booven DV, Cheng-Bang C, Meenakshy M. Limitations of artificial intelligence in healthcare. In: Arora H, editor. Artificial Intelligence in Urologic Malignancies. Academic Press; 2025:231-246. [doi: 10.1016/B978-0-443-15504-8.00008-9]
97. Edwards L. The EU AI act: a summary of its significance and scope. Ada Lovelace Institute; 2021:1-25. URL: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf [Accessed 2026-06-23]
98. Artificial intelligence (regulation) bill [HL]. UK Parliament. URL: https://bills.parliament.uk/bills/3942 [Accessed 2026-06-29]
99. NHS innovation accelerator. NHS England. URL: https://www.england.nhs.uk/aac/what-we-do/how-can-the-aac-help-me/nhs-innovation-accelerator [Accessed 2026-06-29]
100. European action plan on the cybersecurity of hospitals and healthcare providers. European Commission. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/digital-health-and-care/european-action-plan-cybersecurity-hospitals-and-healthcare-providers_en [Accessed 2026-06-29]
101. Harnessing artificial intelligence for health. World Health Organization. URL: https://www.who.int/teams/digital-health-and-innovation/harnessing-artificial-intelligence-for-health [Accessed 2026-06-29]
102. Mosca M. Cybersecurity in an era with quantum computers: will we be ready? IEEE Secur Privacy. 2018;16(5):38-41. [doi: 10.1109/MSP.2018.3761723]
103. Ovabor K, Owolabi OO, Atkison T, et al. Quantum-driven predictive cybersecurity framework for safeguarding electronic health records (EHR) and enhancing patient data privacy in healthcare systems. World J Adv Res Rev. 2025;25(1):1015-1023. [doi: 10.30574/wjarr.2025.25.1.0124]
104. Gupta K, Saxena D, Rani P, et al. An intelligent quantum cyber-security framework for healthcare data management. IEEE Trans Automat Sci Eng. 2025;22:6884-6895. [doi: 10.1109/TASE.2024.3456209]
105. Nguyen TT, Reddi VJ. Deep reinforcement learning for cyber security. IEEE Trans Neural Netw Learning Syst. 2021;34(8):3779-3795. [doi: 10.1109/TNNLS.2021.3121870]

Abbreviations

CNN: convolutional neural network
DL: deep learning
DT: decision tree
EHR: electronic health record
FL: federated learning
GAN: generative adversarial network
HE: homomorphic encryption
KNN: K-nearest neighbor
LSTM: long short-term memory
ML: machine learning
NIST CSF 2.0: National Institute of Standards and Technology Cybersecurity Framework 2.0
NLP: natural language processing
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RF: random forest
SVM: support vector machine

XGBoost: extreme gradient boosting

Edited by Stefano Brini; peer-reviewed by Peter Taiwo, Tafheem Ahmad Wani; submitted 22.Feb.2026; final revised version received 18.Jun.2026; accepted 23.Jun.2026; published 27.Jul.2026

Please cite as:
Rajput K, Zuberi S, Elhajj M, Ochieng W, Darzi A, Ghafur S
Mapping Machine Learning–Driven Cybersecurity Solutions in Health Care: Scoping Literature Review
J Med Internet Res 2026;28:e93950
URL: https://www.jmir.org/2026/1/e93950
doi: 10.2196/93950