© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719759

AI-Enabled Cyber Resilience for Healthcare: Predictive Threat Detection and Risk Prioritization in Kingdom of Saudi Arabia Digital Hospitals

MUHAMMAD UMER MAJEED
ORCID: 0009-0001-7341-8147

Abstract- ***Kingdom of Saudi Arabia hospitals are moving rapidly toward integrated electronic records, connected clinical devices, cloud hosting, virtual care, and analytics-enabled operations. These capabilities improve access and continuity, yet they also create a larger attack surface in which ransomware, identity compromise, data leakage, device manipulation, and supply-chain disruption may affect patient safety as well as information security. This review develops an AI-enabled cyber-resilience model for Kingdom of Saudi Arabia digital hospitals, with emphasis on predictive threat detection and clinically informed risk prioritisation. The objectives were to synthesise recent evidence on healthcare cyber threats, examine how machine learning and explainable analytics can support security operations, align cyber-resilience practices with Kingdom of Saudi Arabia regulatory expectations, and propose a practical model suitable for high-availability clinical environments. A structured narrative review was undertaken using targeted searches of Scopus, Web of Science, PubMed, IEEE Xplore, and selected regulatory sources published from 2020 to 2025. Evidence was coded into five themes: digital-hospital attack surface, predictive detection, risk scoring, governance, and operational resilience. The synthesis indicates that AI can improve early detection when behavioural baselines, device telemetry, identity signals, vulnerability intelligence, and clinical criticality are fused in a governed security data layer. However, AI also introduces model drift, adversarial manipulation, privacy exposure, and automation bias. The proposed model therefore integrates zero-trust identity, monitored segmentation, AI-assisted security operations, clinician-aware triage, privacy controls, and disaster recovery into a single risk-prioritisation cycle. The review concludes that AI-enabled resilience should be treated as a socio-technical operating model rather than a tool purchase.***

Keywords: AI-Enabled Cybersecurity, Cyber Resilience, Kingdom of Saudi Arabia Hospitals, Predictive Threat Detection, Risk Prioritisation, Digital Health, Electronic Health Records, Internet of Medical Things

I. INTRODUCTION

Digital hospitals now depend on electronic medical records, networked diagnostic platforms, cloud services, remote monitoring, and connected biomedical equipment. In Kingdom of Saudi Arabia, this dependence is intensified by national health transformation objectives that emphasise digital access, interoperability, quality improvement, and analytics-driven service redesign [5,25]. A modern tertiary hospital may operate thousands of endpoints and clinical systems, including electronic prescribing, laboratory platforms, imaging archives, mobile nursing devices, smart beds, infusion systems, patient portals, and telehealth applications. Each asset contributes to clinical productivity, but each also expands the set of credentials, interfaces, datasets, and operational dependencies that adversaries can exploit.

Cybersecurity in this environment is no longer only a confidentiality problem. A successful intrusion can delay medication administration, disable scheduling, corrupt diagnostic images, interrupt remote monitoring, or force clinicians to revert to paper-based workflows during time-sensitive care. Studies of healthcare cybersecurity consistently show that digital health systems are vulnerable because they combine legacy devices, high operational urgency, heterogeneous vendors, scarce downtime for patching, and human workflows that cannot be stopped while investigations proceed [10,11,18]. Medical Internet of Things environments add further risk because many devices have constrained computing power, long replacement cycles, weak firmware management, and wireless exposure [12,16].

Artificial intelligence has become attractive because static controls struggle to recognise fast-changing attacks. Machine learning can model normal behaviour, correlate heterogeneous events, detect anomalies, prioritise alerts, and support faster decisions in security operations centres. Recent healthcare studies have demonstrated the value of deep learning, ensemble methods, reinforcement learning, and explainable models for intrusion detection in connected medical environments [15,21-23]. Nevertheless, AI is not inherently resilient. Models can inherit biased training data, overfit to laboratory datasets, miss rare attack paths, generate excessive false positives, or be manipulated by adversarial inputs [8,14,20]. For hospitals, a false negative may allow harm to unfold, while a false positive may interrupt clinical work unnecessarily.

This review argues that Kingdom of Saudi Arabia digital hospitals need a cyber-resilience framework that links AI detection to risk prioritisation, regulatory alignment, and clinical continuity. National cybersecurity and data protection requirements already create a strong baseline through mandatory controls, cloud security expectations, and health-data safeguards [1-4]. Yet compliance alone cannot provide predictive resilience unless hospitals translate requirements into measurable operating capabilities: asset visibility, identity assurance, telemetry quality, incident playbooks, recovery objectives, and executive risk ownership. The central question addressed here is how AI can be embedded in a hospital resilience model that detects emerging threats early, ranks them by clinical consequence, and guides proportionate response without compromising patient care.

II. AIM AND OBJECTIVES

The aim of this review is to develop a publishable, practice-oriented cyber-resilience model for Kingdom of Saudi Arabia digital hospitals that uses AI to improve predictive threat detection and risk prioritisation across electronic records, connected clinical systems, cloud platforms, and security operations. The first objective is to identify the main cyber-risk pathways affecting digital hospital availability, integrity, confidentiality, and patient safety. The second objective is to analyse AI techniques that can improve early detection, including anomaly detection, supervised classification, explainable models, behavioural analytics, and federated learning. The third objective is to define a clinically weighted risk-prioritisation approach that combines likelihood, exploitability, asset criticality, data sensitivity, patient-safety impact, and recovery dependency. The fourth objective is to align the model with Kingdom of Saudi Arabia cybersecurity, cloud, and personal-data protection expectations. The final objective is to propose governance and validation mechanisms that reduce false alarms, model drift, privacy exposure, and unsafe automation.

III. METHODOLOGY

A structured narrative review design was selected because the topic combines empirical studies, technical security models, standards, regulatory requirements, and hospital operating practices. Searches were conducted across Scopus, Web of Science, PubMed, IEEE Xplore, and Google Scholar for literature published between January 2020 and December 2025. Search terms included combinations of healthcare cybersecurity, cyber resilience, artificial intelligence, machine learning, intrusion detection, medical internet of things, electronic health record security, Kingdom of Saudi Arabia digital health, security operations centre, risk prioritisation, ransomware, cloud security, and clinical systems. Regulatory and practice documents from Kingdom of Saudi Arabia and international authorities were also reviewed where they informed hospital governance, risk management, data protection, or cybersecurity baselines [1-4,7,8,27-30].

The inclusion criteria prioritised sources that met at least one of four conditions: direct relevance to healthcare cybersecurity; discussion of AI-based detection or risk analytics; applicability to hospital operations or connected medical devices; or relevance to Kingdom of Saudi Arabia digital-health governance. Sources outside the publication window, purely commercial commentary, and papers lacking relevance to hospital cyber-resilience decisions were excluded. The evidence was coded in a matrix with

five columns: attack surface, detection method, risk-prioritisation variables, operational resilience requirements, and governance safeguards. A thematic synthesis was then used to produce an integrated model, supported by two tables and two graphical representations.

This methodology deliberately avoids treating cybersecurity maturity as a single score. Hospitals are high-dependency clinical ecosystems in which a laboratory interface, identity provider, nurse call server, pharmacy system, or imaging archive may have different clinical consequences even when the technical vulnerability appears similar. Therefore, the synthesis examined how cyber events propagate through patient-care pathways. Special attention was given to telemetry quality, explainability, privacy-preserving analytics, clinical escalation, and recovery readiness because these factors determine whether AI improves resilience or merely increases the alert burden [17,21,24].

IV. DIGITAL-HOSPITAL ATTACK SURFACE IN KINGDOM OF SAUDI ARABIA

Kingdom of Saudi Arabia hospitals operate within a fast-growing digital-health ecosystem. National transformation has increased attention to e-health services, digital readiness, interoperability, and analytics-enabled care [5,25]. The same transformation places hospitals among high-value targets because attackers can monetise health records, exploit downtime urgency, and use connected devices as entry points. Studies of healthcare risk describe a layered attack surface that begins with users and identities, extends through endpoints and medical devices, and reaches cloud workloads, databases, application programming interfaces, third-party providers, and backups [10-13].

The electronic record is the central dependency. If an electronic medical record or enterprise integration engine becomes unavailable, clinicians may lose access to allergies, medication history, previous investigations, orders, and discharge plans. Integrity attacks may be even more dangerous than downtime because altered laboratory results, modified prescriptions, or manipulated imaging metadata can produce incorrect clinical decisions. Connected clinical systems intensify these risks. Infusion pumps, monitoring stations, imaging modalities, laboratory analysers, and remote-care devices generate operational data that are clinically meaningful and security-sensitive. Their security weaknesses are often shaped by firmware age, proprietary protocols, vendor maintenance windows, and limited capacity for local agents [12,13].

Cloud adoption creates a second set of dependencies. Hospitals may use cloud environments for analytics, disaster recovery, telehealth, collaboration, or patient-facing services. Cloud security controls in Kingdom of Saudi Arabia require governance over shared responsibilities, tenant isolation, identity, encryption, logging, and continuous compliance [2]. AI-based monitoring can help detect abnormal workload activity, privilege escalation, or unusual data movement, but it cannot replace architecture. Poor segmentation, excessive permissions, weak service-account hygiene, and incomplete logging will degrade the performance of any detection model.

Human behaviour remains a prominent attack pathway. Phishing, credential reuse, social engineering, misdirected email, and unsafe use of removable media remain difficult to eliminate because hospitals run under pressure, with rotating staff, trainees, contractors, and urgent clinical tasks. Healthcare professionals may prioritise speed because patient care is immediate, while security controls are perceived as secondary or obstructive. The literature therefore supports interventions that combine awareness, usable authentication, workflow-sensitive access controls, and rapid reporting channels rather than relying on training alone [17,19].

V. AI FOR PREDICTIVE THREAT DETECTION

Predictive threat detection seeks to identify malicious activity before it becomes a clinical disruption. In a hospital, this requires more than classifying packets as benign or malicious. The system must understand device role, normal clinical rhythm, user behaviour, location, maintenance windows, and downstream dependency. For example, a large data transfer from

a research server may be normal during an approved project but suspicious from an anaesthesia workstation. A failed login burst may be routine after a password reset campaign but critical when combined with remote access from an unusual location.

Machine learning contributes by learning baselines and detecting deviations. Supervised algorithms can classify known attack patterns when trained on labelled datasets, while unsupervised and semi-supervised models can identify outliers in environments where new attacks are expected. Deep learning models, including convolutional, recurrent, and attention-based designs, have shown strong results for IoMT intrusion detection, especially when spatial and temporal traffic features are considered together [21-23]. Reinforcement learning can support adaptive response selection, although hospital use requires strict guardrails because automated containment can affect clinical availability [22].

Explainability is essential. A security analyst and a clinical technology manager must understand why an AI system ranked an alert as severe. Explainable models can reveal whether the decision was driven by unusual protocol use, impossible travel, privilege escalation, abnormal device traffic, rare process execution, or unexpected access to sensitive records [21,24]. This transparency supports trust, reduces alert fatigue, and enables targeted response. It also supports legal and governance accountability when decisions affect access to clinical systems or personal data [3,4,8].

Data fusion is another requirement. Single-source models are fragile. Network traffic may miss insider misuse, endpoint telemetry may be unavailable on legacy devices, and identity logs may not capture device manipulation. A resilient model should combine identity events, endpoint detections, network flows, domain-name activity, vulnerability data, asset inventory, email telemetry, data-loss signals, cloud logs, and biomedical-device context. Federated or privacy-preserving learning may be useful where hospitals want shared detection intelligence without centralising sensitive patient-level data [13,16]. However, such methods require governance over model updates, poisoning resistance, and validation across heterogeneous clinical environments.

AI can also support threat anticipation. External intelligence about exploited vulnerabilities, ransomware tactics, exposed services, and vendor advisories can be mapped to local assets. If an imaging server uses a vulnerable library, the system should increase its risk priority when exploitation is observed globally. This approach moves security operations from passive alerting toward predictive posture management, where the hospital prioritises actions before compromise occurs [7,14,18].

VI. RISK PRIORITISATION FOR CLINICAL OPERATIONS

Risk prioritisation is the bridge between AI detection and clinical resilience. Many hospitals receive more security alerts than they can investigate. Prioritisation must therefore answer a practical question: which event should receive attention first because it threatens patient care, sensitive data, regulatory exposure, or recovery capacity? Traditional scoring methods focus on technical severity and exploitability. In hospitals, that view is incomplete. A moderate vulnerability on a pharmacy integration server may be more urgent than a critical vulnerability on an isolated training workstation because the former directly affects medication workflow.

The proposed approach uses a clinically weighted risk score. The score combines seven dimensions: likelihood of malicious activity, exploitability, asset criticality, patient-safety impact, data sensitivity, lateral movement potential, and recovery dependency. AI contributes by estimating likelihood from behavioural evidence; vulnerability scanners inform exploitability; the configuration management database provides asset criticality; privacy registers identify data sensitivity; network maps indicate movement potential; and disaster-recovery records reveal recovery dependency. Security analysts retain decision authority, but the system presents a ranked queue with explanations and suggested response options.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719759

Table 1. AI-enabled cyber-resilience domains for Kingdom of Saudi Arabia digital hospitals.

Domain	Core purpose	Key inputs	AI contribution	Governance safeguard
Asset and identity trust	Know who and what is active	CMDB, IAM, vendor access, device roles	Entity baselines and privilege-risk alerts	Named owner, assurance level, review cycle
Security data fabric	Unify signals without unnecessary content exposure	Logs, flows, cloud events, device gateways	Normalisation, correlation, missing telemetry detection	Data minimisation and access control
Predictive analytics	Detect emerging attack behaviour early	Network, endpoint, vulnerability, email, identity signals	Anomaly detection, classification, scoring	Model validation, drift checks, explainability
Clinical risk prioritisation	Rank alerts by care consequence	Service dependency, data class, recovery dependency	Clinically weighted risk queue	Human approval for high-impact containment
Response and recovery	Preserve safe operations during incidents	Playbooks, backups, downtime procedures	Action recommendation and recovery readiness checks	Exercise evidence and after-action learning

Patient-safety impact should be explicit. Events involving electronic prescribing, medication dispensing, critical-care monitoring, emergency department triage, blood bank systems, radiology reporting, or operating-theatre scheduling should receive higher priority than events affecting administrative convenience. The same principle applies to integrity. A potential data alteration event in a diagnostic pathway deserves immediate clinical validation even if availability remains intact. Table 1 summarises the proposed domains for AI-enabled resilience, while Table 2 translates those domains into a risk-prioritisation matrix for Kingdom of Saudi Arabia digital hospitals.

Table 2. Clinically weighted risk-prioritisation matrix

.

Factor	High-score condition	Hospital example	Immediate trigger
Likelihood	Multiple correlated indicators across identity, endpoint and network	New admin login followed by unusual data movement	SOC validation within 15 minutes
Exploitability	Known active exploitation or internet-facing exposure	Unpatched portal or VPN linked to clinical network	Emergency patch or compensating control
Asset criticality	System supports time-sensitive clinical workflow	Electronic prescribing, radiology, laboratory or ICU monitoring	Clinical incident command notified
Data sensitivity	Large volume of identifiable health data or special-category records	Patient database, imaging archive, research dataset	Privacy officer and legal process engaged
Recovery dependency	Backup, failover, or manual workaround is uncertain	Single-site EMR component with untested restore	Recovery team validates restore readiness

Risk prioritisation must also account for response friction. Isolating a compromised device may be necessary, but doing so without clinical coordination can interrupt care. The mature response is not either security or care; it is a predefined decision pathway. For high-risk alerts, the security operations centre should contact clinical engineering, the relevant department lead, and incident command.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719759

Containment actions should follow approved playbooks that specify safe downtime procedures,

substitute devices, backup access, and communication templates. AI should accelerate these decisions, not make them invisibly.

VII. PROPOSED CYBER-RESILIENCE FRAMEWORK

Figure 1 presents the proposed architecture. The foundation is a trusted asset and identity layer. Every user, application, device, server, integration engine, and cloud workload should have an owner, role, dependency map, and assurance level. Zero-trust principles then require continuous verification, least privilege, strong authentication, device posture checks, segmentation, and monitored access paths. Kingdom of Saudi Arabia cybersecurity controls reinforce these practices by emphasising governance, asset management, access control, cryptography, operations, incident management, business continuity, and third-party security [1,2].

Figure 1. Minimalist AI-enabled cyber-resilience architecture for digital hospitals.

The second layer is the hospital security data fabric. This layer normalises logs and telemetry from identity providers, endpoints, network sensors, medical device gateways, cloud platforms, firewalls, vulnerability tools, email systems, and data-loss systems. Clinical context is added without exposing unnecessary patient content. For example, the model needs to know that a device supports intensive care monitoring, but it does not need full clinical notes to detect abnormal protocol behaviour. This separation supports data minimisation and health-data protection obligations [3,4].

The third layer contains AI analytics. Models are divided into four functions: anomaly detection, threat classification, risk scoring, and response recommendation. Each model has a documented purpose, approved training sources, validation metrics, drift monitoring, and owner. Explainability

is built into the analyst interface. The system should show what changed, why it matters, which assets are affected, which clinical services depend on them, and what actions are recommended. Model performance should be reviewed after incidents and exercises, not only during procurement.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719759

The fourth layer is operational resilience. This layer links AI findings to security operations, incident response, disaster recovery, and clinical continuity. A predictive alert that indicates ransomware staging should automatically check backup health, privileged-session activity, endpoint isolation readiness, and downtime communication channels. A suspected device manipulation alert should trigger clinical engineering review and validation of patient-facing outputs. A cloud misconfiguration alert should check exposure, encryption, identity permissions, and logging status. Figure 2 illustrates this operating cycle from detection to prioritisation, action, learning, and improvement.

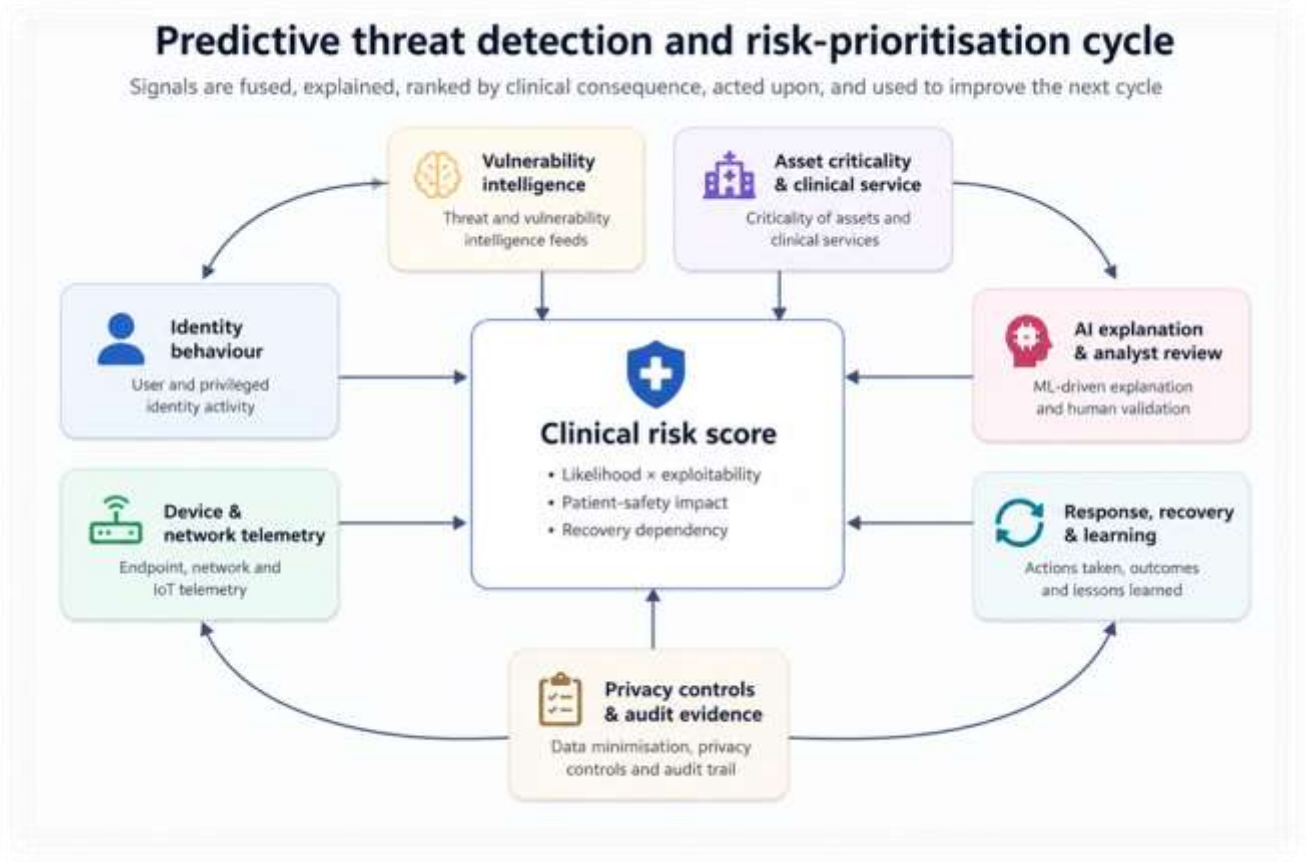

Figure 2. Predictive threat-detection and clinical risk-prioritisation workflow.

The final layer is governance. The board and executive leadership should receive risk indicators that reflect clinical consequence rather than raw alert volume. Useful indicators include percentage of critical assets with complete telemetry, mean time to validate high-risk alerts, percentage of AI alerts with explanations, unresolved high-risk vulnerabilities on clinical systems, backup recovery success rate, third-party remote access compliance, and number of clinical workflows covered by downtime playbooks. These measures align technical operations with accountable resilience.

VIII. KINGDOM OF SAUDI ARABIA COMPLIANCE AND OPERATING ALIGNMENT

The Kingdom of Saudi Arabia setting gives the proposed framework a distinctive governance profile. Digital hospitals must align predictive security with national cybersecurity controls, cloud obligations, and data-protection rules. The updated essential controls create a baseline for governance, asset

management, identity, access control, cryptography, operations, incident management, business continuity, and third-party oversight [1]. Cloud controls add requirements for shared responsibility, tenant governance, security monitoring, compliance evidence, and secure use of hosted services [2]. Personal-data law and its implementing rules require organisational, technical, and administrative measures for health data, including protection against unauthorised access, misuse, disclosure, loss, and unlawful processing [3,4]. These requirements mean that AI security analytics cannot be treated as experimental monitoring outside governance; they must be part of accountable hospital risk management.

A practical compliance mapping begins by linking each AI function to a control objective. Asset-discovery models support asset management. Behavioural identity analytics support access control and privileged-account monitoring. Data-exfiltration models support privacy safeguards and data-loss prevention. Cloud posture analytics support secure configuration and continuous compliance. Anomaly detection for biomedical gateways supports critical-system protection. Incident prediction and recovery checks support business continuity. This mapping enables the security operations centre to produce evidence for audits without reducing resilience to checkbox activity. The goal is not to create more documentation, but to prove that high-risk clinical assets are monitored, alerts are explainable, response decisions are recorded, and recovery capabilities are tested.

Governance should also define acceptable AI use. Hospitals should specify which data categories may be processed for detection, how long security telemetry is retained, who can access staff-behaviour analytics, and how false-positive investigations are closed. Role-based access to the security data fabric should prevent analysts from viewing patient content unless a formally authorised investigation requires it. Model-development teams should receive de-identified or minimised data wherever possible. When synthetic data are used for testing, hospitals should verify that they preserve relevant security patterns without exposing personal information. These practices help reconcile predictive analytics with privacy, clinical trust, and legal accountability [3,4,8].

Compliance alignment is especially important for third parties. Kingdom of Saudi Arabia hospitals depend on vendors for electronic records, imaging systems, laboratory platforms, biomedical maintenance, cloud hosting, and managed security services. Third-party remote access should be time-bound, strongly authenticated, monitored, and linked to approved work orders. AI systems can flag unusual vendor sessions, unapproved tools, excessive data access, or connections outside maintenance windows. Contractual controls should require timely vulnerability disclosure, security logging, incident notification, recovery support, and participation in joint exercises. Because a supplier compromise can become a hospital outage, vendor risk should be included in the clinically weighted score rather than handled as a separate procurement issue.

IX. VALIDATION, METRICS, AND CONTINUOUS LEARNING

AI-enabled resilience requires rigorous validation. Hospitals should evaluate detection models in four environments: retrospective log analysis, controlled simulation, red-team exercises, and live supervised monitoring. Retrospective analysis tests whether a model would have detected historical suspicious activity. Simulation tests whether the model recognises synthetic ransomware staging, credential misuse, lateral movement, data exfiltration, or device anomaly scenarios. Red-team exercises examine adversarial adaptation and response coordination. Live supervised monitoring measures real false positives, alert fatigue, and clinical operational fit. Each stage should include clinical engineering and department representatives when the scenario involves patient-facing technology.

Performance metrics should be balanced. Accuracy alone is inadequate because rare high-impact attacks may be hidden within large volumes of normal traffic. Precision indicates whether analysts can trust alerts; recall indicates whether serious events are missed; time-to-detection measures speed; and time-

to-validation measures operational usefulness. For hospitals, additional indicators are required: number of alerts requiring clinical coordination, percentage of high-risk alerts with documented clinical impact, downtime avoided, recovery objective achieved during exercises, and number of unsafe automated actions prevented. Explainability should be measured through analyst review, asking whether the explanation identifies the decisive evidence and supports a safe response [21,24].

Continuous learning should be controlled rather than uncontrolled. After an incident, near miss, or exercise, investigators should update playbooks, asset criticality, model features, and detection rules. However, automatic retraining must not proceed without validation because poisoned or unrepresentative data could weaken detection. Model changes should pass change-management review, version control, rollback planning, and performance comparison. This is especially important in hospitals where new clinical systems are introduced, departments move, patient volumes change, or cloud services are reconfigured. Drift monitoring should therefore be linked to hospital change calendars and vendor maintenance plans.

The most mature hospitals will treat cyber-resilience learning as part of quality improvement. Security incidents, downtime drills, and AI-alert reviews should generate lessons that are discussed across information technology, clinical engineering, nursing, medical leadership, privacy, risk management, and executive governance. This approach recognises that cyber resilience is not located inside the security operations centre alone. It is distributed across people, devices, clinical decisions, infrastructure, vendors, and recovery plans. AI improves this ecosystem when it strengthens shared situational awareness and prioritises the issues most likely to affect safe care.

X. IMPLEMENTATION CONSIDERATIONS

Implementation should begin with asset clarity. Hospitals cannot prioritise what they cannot identify. The first phase is to establish an authoritative inventory of clinical systems, biomedical devices, cloud workloads, interfaces, privileged accounts, data stores, and third-party connections. Each asset should be mapped to a clinical service and assigned an impact level. The second phase is telemetry improvement. Critical systems should generate reliable logs, synchronised time stamps, security events, and performance signals. Where legacy devices cannot host agents, passive network monitoring and gateway telemetry should be used.

The third phase is model development and validation. Hospitals should not assume that a model trained on public datasets will perform well in a Kingdom of Saudi Arabia hospital. Local workflow, language, vendor mix, network architecture, clinical schedules, and maintenance routines shape normal behaviour. A safe approach is to begin with supervised use, where AI ranks alerts but analysts make decisions. The model should be tested against historical incidents, red-team scenarios, tabletop exercises, and benign high-volume events. Metrics should include precision, recall, false-positive burden, time-to-triage, explainability quality, and clinical disruption avoided [15,21-24].

The fourth phase is integration with security operations. AI outputs should appear in the existing case-management platform, not in a separate dashboard that analysts ignore. Each alert should include severity, rationale, affected services, recommended actions, required clinical contacts, and recovery dependencies. Security orchestration may automate low-risk enrichment, such as pulling asset details or checking vulnerability status. High-impact containment should require human authorisation because clinical consequences must be considered. This balance reduces automation bias while still improving speed.

The fifth phase is governance of AI risk. AI models should be treated as controlled hospital assets. Their training data, versions, performance, assumptions, and access rights should be documented. Model drift should be monitored after major system changes, new vendors, network redesign, or shifts in clinical operations. Adversarial resilience should be considered because attackers may attempt to evade detection, poison data, or trigger noisy false

positives. Ethical governance is also necessary, especially when staff behaviour analytics are used. Monitoring must be proportionate, transparent, and limited to legitimate security purposes [8,20].

In practice, the framework should be introduced gradually. A hospital may begin with high-value workflows such as emergency care, pharmacy, intensive care, radiology, and enterprise identity, then extend monitoring to outpatient portals, research environments, and administrative platforms. Early deployment should prioritise explainable alerts, analyst training, and clinical escalation routes before broader automation is attempted. This staged approach reduces resistance because departments see that security analytics protect continuity rather than obstruct work. It also permits local calibration of thresholds, collection of baseline traffic, and refinement of asset criticality. Over time, the same evidence base can support investment decisions, national reporting, insurance discussions, and board oversight. The framework is therefore both a technical architecture and a management discipline for sustaining trust in digital hospitals. Most importantly, it keeps patient safety as the primary measure of cyber performance, not a secondary operational concern today.

XI. CHALLENGES AND RESEARCH DIRECTIONS

Several challenges remain. First, healthcare AI security research often relies on benchmark datasets that do not capture hospital complexity. Real hospital traffic contains maintenance activity, emergency surges, device quirks, and clinical exceptions. Future research should use privacy-preserving multicentre datasets that represent Kingdom of Saudi Arabia hospital environments while protecting patient and staff confidentiality. Second, predictive models need stronger evaluation against adversarial behaviour. Detection accuracy on clean test data is insufficient if attackers can alter timing, protocol features, or login patterns to evade the model.

Third, risk prioritisation requires better measurement of patient-safety consequence. Cybersecurity teams often lack structured access to clinical dependency information, while clinical leaders may not know which technical assets support their workflows. Joint mapping between clinical operations, biomedical engineering, information technology, and security teams is required. Fourth, smaller hospitals may lack specialised AI security expertise. Managed services, shared threat intelligence, and standardised playbooks can help, but accountability must remain with the healthcare organisation.

Fifth, the interaction between national compliance and AI governance needs continued research. Kingdom of Saudi Arabia hospitals must manage cybersecurity controls, cloud responsibilities, personal-data obligations, and health-data safeguards [1-4]. AI can automate evidence collection and control monitoring, but automation must be auditable. Finally, cyber resilience should be studied as an outcome. Research should examine whether AI-enabled prioritisation actually reduces downtime, improves response accuracy, protects patient safety, and lowers recovery cost, rather than only improving technical detection metrics.

XII. CONCLUSION

AI-enabled cyber resilience offers Kingdom of Saudi Arabia digital hospitals a realistic path from reactive defence toward predictive, clinically aware protection. The evidence from 2020 to 2025 indicates that machine learning, explainable analytics, and behavioural modelling can improve intrusion detection in healthcare and IoMT environments, particularly when combined with strong governance and validated operational workflows. Yet AI is not a substitute for foundational controls. Its value depends on complete asset visibility, trustworthy identity, reliable telemetry, segmentation, secure cloud architecture, privacy safeguards, incident response, and tested recovery.

The model proposed in this review integrates these requirements into a five-layer framework: trusted assets and identity, security data fabric, AI analytics, operational resilience, and governance. Its main contribution is the clinically weighted risk-prioritisation cycle, which ranks threats not only by technical severity but by their likely consequences for patient care, data protection, lateral movement, and

recovery. For Kingdom of Saudi Arabia hospitals, such a model aligns digital transformation with national cybersecurity and data-protection expectations while supporting high-availability clinical operations. Future empirical work should validate the model in live hospital settings, quantify its effect on response time and patient-safety risk, and refine privacy-preserving learning across multiple healthcare organisations.

REFERENCES

[1] National Cybersecurity Authority. Essential Cybersecurity Controls (ECC 2-2024). Riyadh: National Cybersecurity Authority; 2024.

[2] National Cybersecurity Authority. Cloud Cybersecurity Controls (CCC-1:2020). Riyadh: National Cybersecurity Authority; 2020.

[3] Saudi Data & AI Authority. Personal Data Protection Law. Riyadh: Saudi Data & AI Authority; 2023.

[4] Saudi Data & AI Authority. Implementing Regulations of the Personal Data Protection Law. Riyadh: Saudi Data & AI Authority; 2024.

[5] Al-Kahtani N, Alruwaie S, Al-Zahrani BM, Abumadini RA, Aljaafary A, Hariri B, et al. Digital health transformation in Saudi Arabia: a cross-sectional analysis using Healthcare Information and Management Systems Society digital health indicators. Digital Health. 2022; 8:20552076221117742.

[6] World Health Organization. Global Strategy on Digital Health 2020-2025. Geneva: World Health Organization; 2021.

[7] National Institute of Standards and Technology. The NIST Cybersecurity Framework 2.0. Gaithersburg, MD: NIST; 2024.

[8] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD: NIST; 2023.

[9] European Union Agency for Cybersecurity. ENISA Threat Landscape: Health Sector. Athens: ENISA; 2023.

[10] He Y, Aliyu A, Evans M, Luo C. Health care cybersecurity challenges and solutions under the climate of COVID-19: scoping review. Journal of Medical Internet Research. 2021;23(4):e21747.

[11] Ewoh P, Vartiainen T, Nambisan P. Vulnerability to cyberattacks and sociotechnical solutions for health care systems: systematic literature review. Journal of Medical Internet Research. 2024;26: e46904.

[12] Elhoseny M, Thilakarathne NN, Alghamdi MI, Mahendran RK, Gardezi AA, Weerasinghe H, et al. Security and privacy issues in Medical Internet of Things: overview, countermeasures, challenges and future directions. Sustainability. 2021;13(21):11645.

[13] Huda S, Islam MR, Abawajy J, Kottala VNV, Ahmad S. A cyber risk assessment approach to federated identity management framework-based digital healthcare system. Sensors. 2024;24(16):5282.

[14] Radanliev P. AI security and cyber risk in IoT systems. Digital Society. 2024;3:1-26.

[15] Balhareth G, Alharbi A, Alqurashi M, Aldhyani THH. Optimized intrusion detection for IoMT networks with tree-based machine learning and feature selection. Scientific Reports. 2024; 14:20733.

[16] Khatun MA, Raza SM, Islam MS, Gawanmeh A. Machine learning for Healthcare-IoT security: a review. IEEE Access. 2023; 11:12101-12125.

[17] Hasegawa K, Daoud A, Sherif M, Aborode AT. Cybersecurity interventions in health care organizations in low- and middle-income countries: scoping review. Journal of Medical Internet Research. 2024;26:e51775.

[18] Aldosari B. Cybersecurity in healthcare: new threat to patient safety. Cureus. 2025;17(5):e84247.

[19] Alharbi A, Alotaibi Y, Alghamdi S. Cybersecurity governance in the healthcare sector during digital transformation. Frontiers in Public Health. 2025; 13:1703689.

[20] Di Palma G, Cafiero C, Frascino M, Romano A. AI-induced cybersecurity risks in healthcare: a narrative review. Healthcare. 2025;13(20):2558.

[21] Hosain Y, Islam MM, Alshamrani SS. XAI-XGBoost: an innovative explainable intrusion detection framework for Internet of Medical Things systems. Scientific Reports. 2025; 15:21095.

[22] Shaikh JA, Wang C, Sima MWU, Arshad M, Owais M, Hassan DSM, et al. A deep reinforcement learning-based robust intrusion detection system for securing IoMT healthcare networks. Frontiers in Medicine. 2025; 12:1524286.

[23] Siqueira LP, Carvalho JV, Reis R, Albuquerque VHC. A comprehensive survey on intrusion detection systems for Healthcare 5.0. Computers in Biology and Medicine. 2025; 184:109371.

[24] Alzakari SA, Alghamdi A, Alshammari T. Explainable artificial intelligence-based cyber resilience in smart healthcare systems. Scientific Reports. 2025; 15:18770.

[25] Health Sector Transformation Program. Health Sector Transformation Program Delivery Plan 2021-2025. Riyadh: Kingdom of Saudi Arabia; 2021.

[26] AlDulijand NA, Alshammari F, Alotaibi M, Alanzi T. Sustainable healthcare resilience: disaster preparedness and business continuity planning in Saudi hospitals. Sustainability. 2024;16(1):198.

[27] International Organization for Standardization. ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection: information security management systems. Geneva: ISO; 2022.

[28] International Organization for Standardization. ISO/IEC 27005:2022 Information security, cybersecurity and privacy protection: guidance on managing information security risks. Geneva: ISO; 2022.

[29] International Organization for Standardization. ISO/IEC 42001:2023 Information technology: artificial intelligence management system. Geneva: ISO; 2023.

[30] Cybersecurity and Infrastructure Security Agency. Cross-Sector Cybersecurity Performance Goals. Washington, DC: CISA; 2023.