iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

SMART HOSPITAL AI-BASED INTRUSION DETECTION: PERFORMANCE, EXPLAINABILITY, AND ETHICAL IMPLICATIONS

Ahmad Ikram
Virginia University Science and Technology
Seattle Washington
Ahmad.miguel@gmail.com

ABSTRACT
Smart Hospitals are based on the use of pervasive interconnections, the Internet of Medical Things (IoMT) devices, electronic health records (EHRs), and artificial intelligence (AI)-imposed systems to enhance clinical processes and patient outcomes. Nevertheless, this highly interconnected area greatly increases the scope of cyberattacks, and health infrastructure is an increasingly effective target, including ransomware, data leakage, and IoT devices. Legacy intrusion detection systems (IDSs, which are either inherently based on the notion of identifying anomalies or on largely fixed signatures, are not flexible to the dynamic threat space and heterogeneous devices, and are bound by the rigid real-time requirements of clinical facilities.
The paper views AI-generated IDSs in terms of Smart Hospitals, which encompass three interrelated aspects: performance, explainability, and ethical risks. We shall start by summarizing the current state of the art in AI/ML-based IDS models for IoMT and healthcare networks, which will include deep learning, reinforcement learning, and federated learning-based models and methods. Second, it shows the accuracy of the key performance indicators (KPIs), the detection rate, latency, scalability, and false-positive rates in the context of Smart Hospital operations and the challenges, including class imbalance, concept drift, and diverse devices. Third, we note that explainable AI (XAI) procedures (LIME, SHAP, attention mechanisms, explainable tree-based models) can be applied to enhance trust, regulatory compliance, and incident response without increasing the model's cost or performance.
Finally, the ethical and governance issues, i.e., data privacy, bias, accountability, and regulatory uncertainty are discussed in the context of HIPAA, GDPR and new AI-specific healthcare regulations. We assert that nascent AI-based IDSs in Smart Hospitals must be brought up in a fragile property amid the security performance, interpretability, and ethics and current policy pathways and research options in case of trustworthy AI security in healthcare 5.0 environments.

Keywords:
Smart Hospitals; Intrusion Detection systems; Internet of Medical things; AI Security; Explainable AI; Healthcare Cybersecurity; Ethical AI.

INTRODUCTION

1.1 Background and Motivation

Smart Hospitals: Smart Hospital can be defined as the combination of Internet of Things/Internet of Medical Things, clinical decision support wet with AI support, EHR systems, and cloud/edge computing architecture into a highly interconnected and interwoven data-driven system. Such an ecosystem provides the possibility to monitor continuously, provide customized treatment, and simplify the workflow and create thick attack surfaces and interdependency that adversaries can exploit (Khan B. et al., 2023; Kumar et al., 2025). The breach of an individual IoMT device, infusion pump, or an imaging device will cause a chain of events of impacting the operations of an establishment, heighten the risks of patient safety, or cause a massive data leak (Naghib et al., 2025; Biasin et al., 2024).
Data security and privacy are, therefore, the foundations of patient trust and clinical safety. Certain examples of such cases include ransomware and data exfiltration, cross-system changes in the network-wide traffic, and manipulation of clinical data already demonstrated the susceptibility of the healthcare infrastructures worldwide (Murdoch, 2021; Said et al., 2021).
The traditional Intrusion Detection Systems (IDSs) are mainly signature based systems where known malicious activity patterns are detected or anomaly based systems where an abnormality is detected. Such systems can be used successfully in the context of certain attacks but, unfortunately, they are not capable of detecting zero-day

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

attacks, polymorphic malware, and subtle adversarial behaviors typical of advanced threat agents of the modern world (Narayanan et al., 2020; Wahab et al., 2022).

Recent developments in AI and machine learning could help to enhance the performance of IDS, learning complex patterns in high-dimensional network and device data, changing according to emerging threats, and the capacity to identify anomalies in a context-sensitive manner (Khan A. et al., 2024; Shaikh et al., 2025; Lui et al., 2025). However, here, implementing AI in a safety-critical environment raises new questions: black box decision-making, bias in the models, inconsistency in the regulations, and ethical concerns of responsibility in the case of AI failures (Mirbabaie et al., 2021; Hussein et al., 2024; Busch et al., 2025).

1.2 Problem Statement

The challenge of acquiring Smart Hospitals is due to:

- Non-homogeneous devices and protocols (old medical equipment, new IoMT sensors, mobile communication devices, connections to the cloud).
- Clinical processes which are performed in real-time and delays or false alarms can ruin or damage care.
- Complex network topologies exist between on-premises, edge, and cloud network systems.
- The traditional IDSs are not well-adapted to the environment as they assume the uniformity of the traffic, centrality of the visibility and constrained variety of devices and are readily overwhelmed by the noise and evolving threats (Narayanan et al., 2020; Said et al., 2021). Artificial intelligence-based IDSs will only bring new potential threats: black-box nature, warnings that are hard to interpret, potential discrimination, and gaps in AI control over healthcare (Islam et al., 2024; Rjoub et al., 2023; Biasin et al., 2024).

1.3 Research Questions

This paper discusses the following three questions:

1. **Performance:** How well do AI-driven IDSs detect different intrusions within Smart Hospital settings and what are their performance based on KPIs of detection rate, false positives, latency, and scale?
2. **Explainability:** What are the main explainability problems and opportunities of AI-driven IDS in high stakes healthcare settings?
3. **Ethical Risks:** What are the ethical risks and implications of putting AI-driven IDS in Smart Hospitals, particularly privacy of patient data, bias, responsibility and human control?

1.4 Objectives of the Paper

- To investigate the performance measures as well as the realities of AI-based IDSs in the Smart Hospital environments.
- To investigate the effects of having a complex AI-based IDS model that can be explained on trust and incident response as well as to investigate how it can be made more explainable.
- To determine and mention AI-based IDSs and ethical threats and regulatory issues in cybersecurity in healthcare.

1.5 Scope of the Research

The paper focuses on:

- IoMT and Smart Hospital network intrusion detection through AI/ML as opposed to generic enterprise networks.
- These are three axes, i.e. the performance, explainability (XAI) and ethical/regulatory risks.
- New and peer-reviewed non-adolescent articles of AI IDS, XAI in the context of cybersecurity, and AI governance in health care (e.g., Khan A. et al., 2024; Hosain and Cakmak, 2025; Naghib et al., 2025; Hussein et al., 2024; Busch et al., 2025).

LITERATURE REVIEW

2.1 The paper will begin with superimposing Smart Hospital Architectures and Cybersecurity Challenges.

Smart Hospitals represent the combinations of multi-layer of the IoMT networks, EHRs, and clinical information systems, and the cloud/edge infrastructure (Naghib et al., 2025; Almalki et al., 2024). Typical layers include

Perception layer: medical field, wearable, imaging devices, sensors.

Network layer SDN/NFV, wireless/wired infrastructure, gateway

EHR systems Application layer, clinical decision support, hospital information systems.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The other vulnerabilities that are often identified include outdated legacy devices, weak authentication, poorly configured wireless, and the absence of network segregation (Said et al., 2021; Zachos et al., 2025). Such are exploited by the attackers in ransomware, DDoS, man in middle, data exfiltration and hijacking devices.

The table 1 below introduces the major cyber threats in Smart Hospital environments.

TABLE 1 : Common Cyber Threats in Smart Hospitals

Threat Type	**Description**	**Primary Impact**	**Typical Targets**
Ransomware	Encrypts hospital data and demands payment	Operational disruption; financial loss	EHR servers; file shares; backup systems
Data Exfiltration	Stealthy theft of patient/operational data	Privacy violations; regulatory penalties	EHR databases; cloud storage
IoMT Device Hijacking	Remote compromise/control of medical devices	Safety risks; altered treatment parameters	Infusion pumps; monitors; imaging devices
Lateral Movement	Spread from compromised node across hospital network	Broader compromise; persistent attacker access	Workstations; domain controllers; internal servers
DDoS Attack	Overwhelms services with malicious traffic	Service outage; degraded patient care	Patient portals; telemedicine systems; hospital gateways
Insider Misuse	Abuse of legitimate credentials	Unauthorized access; data tampering	Administrative consoles; EHR access points

2.2 The Conventional Intrusion Detection Systems.

Conventional IDSs are characterized by signature-based or anomaly-based (Narayanan et al., 2020; Said et al., 2021). Signature-based IDSs implement a compare-and-contrast approach to traffic, which is very accurate in known attacks but nearly non-existent in novel attacks. Anomaly-based IDSs are trained on a baseline of normal behavior and report deviations, but have high false positives and tend to be sensitive to changes in environment such as Smart Hospitals with variation in workload and emergency cases.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Table 2. Comparison between traditional and AI-based IDS

Dimension	Signature-Based IDS	Anomaly-Based IDS	AI-Driven IDS
Knowledge of Attacks	Requires known attack signatures	Detects unknown anomalies	Learns complex patterns; generalizes to novel attacks
False Positives	Low for known threats	High in dynamic settings	Lower with proper model training
Adaptability	Limited; manual updates required	Moderate; retraining needed	High; supports continual and federated learning
Scalability	Struggles with high traffic volumes	Moderate scalability	High scalability via distributed, edge, and federated architectures
Smart Hospital Suitability	Weak against IoMT & zero-day threats	Struggles with complex workflows	Best suited for heterogeneous, evolving Smart Hospital environments

2.3 Intrusion DetectionArtificial Intelligence/Machine Learning.

AI-based IDSs have a lot of models:

Reinforcement learning (RL): it is used in adaptive defense mechanism and policy-based dynamic configuration (Shaikh et al., 2025).

Federated learning (FL): It enables different institutions or departments to share information during the training of a model without having raw patient data (Almalki et al., 2024; Ashraf et al., 2022).

Wahab et al. (2022) propose a hybrid framework to facilitate e-health using IoT, which is a combination of feature engineering and deep models driven by AI. The case of the deep learning-based anomaly detection of the IoMT networks is demonstrated by Khan A. et al. (2024) and the Lui et al. (2025), and the case of the deep RL-based robust IDS of the IoMT healthcare networks is presented by Shaikh et al. (2025). Ashraf et al. (2022) introduce FIDChain, an IoT healthcare federated IDS in blockchains and comment on the opportunities of cooperating in a privacy-friendly manner using FL.

2.4 IDSs Performance Metrics and Performance Evaluation.

Accuracy, precision, recall, F1-score, detection rate (True Positive Rate), false positive rate (FPR) and Area Under the ROC Curve (AUC) are the most significant performance measures deployed in the evaluation of the performance of IDS (Zachos et al., 2025). Other KPIs that are vital in healthcare include:

1. **Latency**: It is the time taken after the detection of the attack and the actual initiation of the attack.
2. **Throughput and scalability:** ability to gauge the high volume of traffic and the many devices.

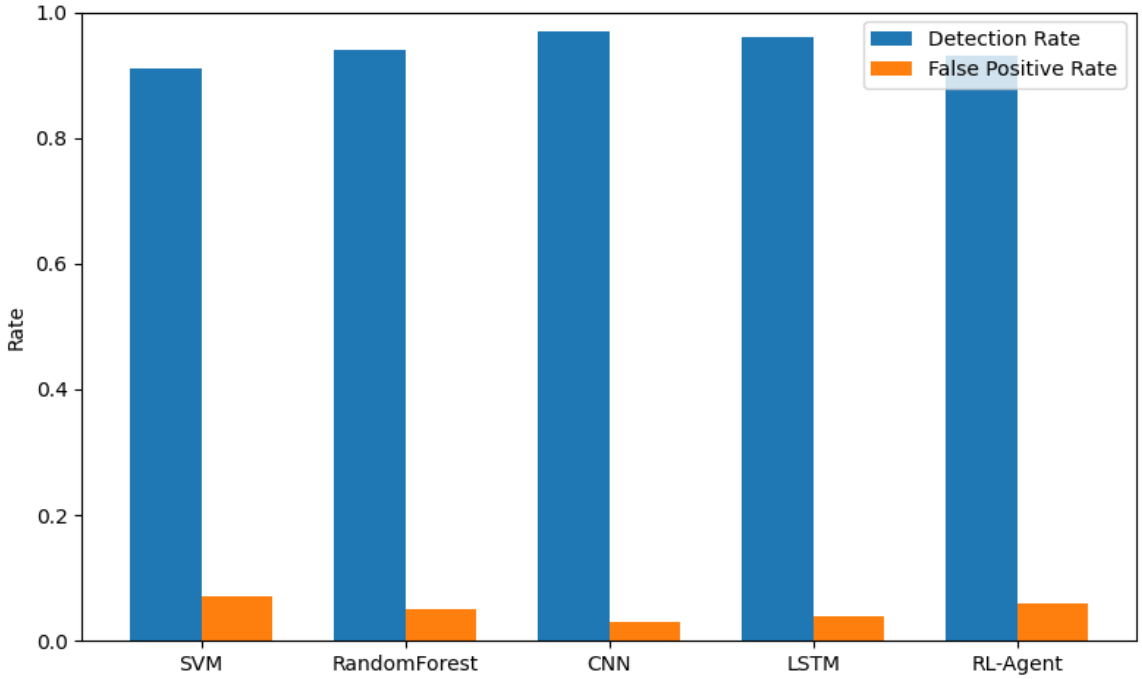
Resource overload CPU, memory and bandwidth usage in small medical devices and gateways.

The real-life evaluation is also complicated by the unbalanced data, the absence of labelled data, and the privacy issues of providing the real patient traffic (Naghib et al., 2025). Artificial data might fail to characterize delicate conduct and functioning deviations.

The bar chart (described in Figure 1) can be used to compare the rate of detection and false positives of AI models with a hypothetical data of the Smart Hospital

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Figure 1 – Detection vs. false positive rates



2.5 The AI explanation in Cybersecurity and Health.

Considering the safety-related quality of healthcare, explainability is a key to AI-based IDS adoption. Some of the techniques, including SHAP, LIME, feature importance scores, attention mechanisms, and saliency maps, become more popular in security models (Zhang et al., 2022; Rjoub et al., 2023; Islam et al., 2024).

According to Hosain and Cakmak, XAI-XGBoost is a proposed explainable intrusion detection method that uses interpretable tree boosting to IoMT settings (2025). Muhammad et al. (2025) introduce L-XAIDS, which is a LIME-based IDS decision explanation framework. Mohale and Obagbuwa (2025) offer a review of the integration of XAI in IDS systems in a systematic manner focusing on transparency and interpretability. Si-ahmed et al. (2024) dwell upon explainable machine learning-based protection of security and privacy in the IoMT and map the XAI results to security controls that could be comprehended by humans.

2.6 AI/Healthcare Ethical Frameworks and Ethical risks.

Healthcare AI should also comply with ethical values, including beneficence, non-maleficence, autonomy, and justice, and legal standards, including HIPAA, GDPR, and future AI regulation (Murdoch, 2021; Mirbabaie et al., 2021; Hussein et al., 2024; Busch et al., 2025). Rony et al. (2024) get the ideas of nurses regarding privacy and ethical issues related to the use of AI. Murdoch (2021) addresses the issue of privacy in the health data processing with AI.

Hussein et al. (2024) and Busch et al. (2025) visualize changing AI governance and regulation maturity and Naili et al. (2025) and Biasin et al. (2024) review AI medical devices legal implications and cybersecurity. Gala (2024) and Pasricha (2022) present the ethical and legal implications of health cybersecurity and smart healthcare more generally, in general, driven by AI. Jha et al. (2023) believe in the ethical and philosophical basis of efficient medical artificial intelligence.

3.0 AI-DRIVEN INTRUSION DETECTION SYSTEMS PERFORMANCE.

3.1 Artificial Intelligence/Machine Learning Models of IDS in Smart Hospitals.

Smart Hospitals that use AI-driven IDSs usually consume network flows, device logs, EHR access patterns, authentication logs, and system events (Said et al., 2021; Naghib et al., 2025). Models must handle:

- Versatile information types (ordered records, time category, categorical occurrences).
- Multi-scale time-patterns (milliseconds-based flows of packets and hours-based user sessions)
- Distributed deployments, occasionally with a restricted number of computational resources on edge gateways and medical devices.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

- Similar federated learning techniques like those by Almalki et al. (2024) and Ashraf et al. (2022) are useful in training IDS models without disclosing local data, which is in line with regulatory demands.

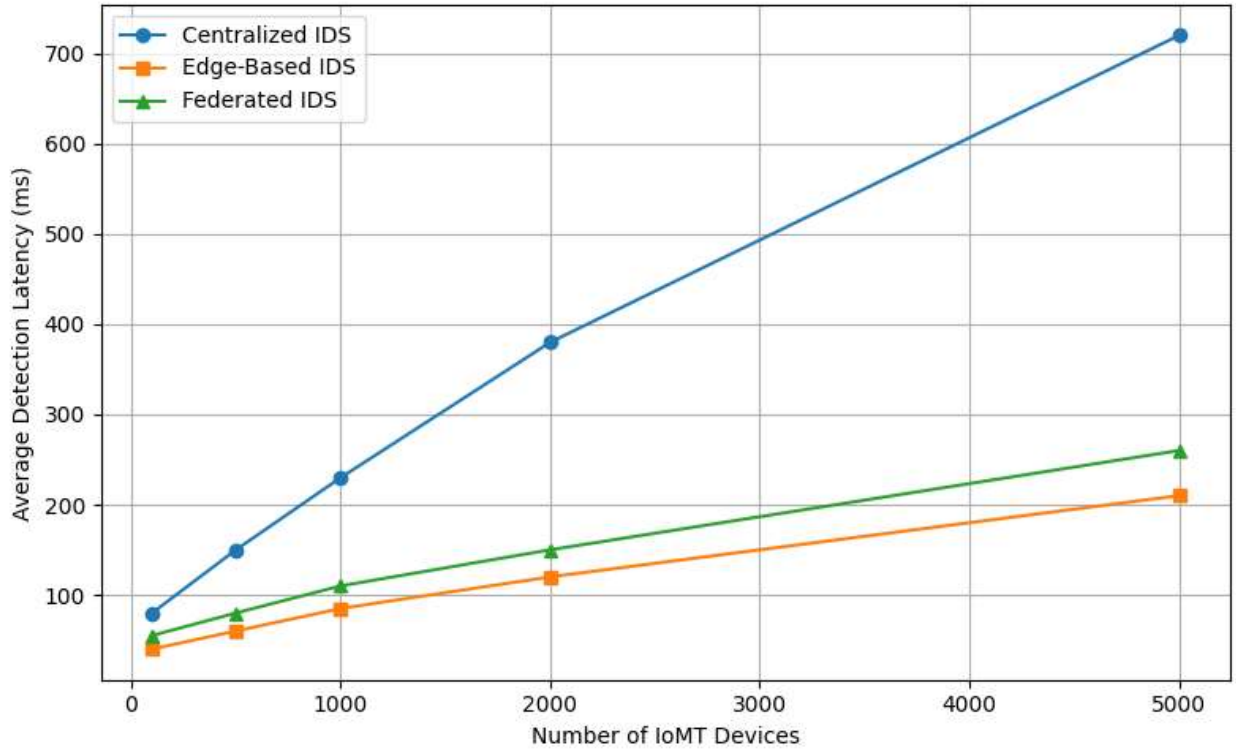
3.2 Key Performance Indicators

In case of AI-driven IDSs in Smart Hospitals, KPIs do not regard only traditional ML metrics but operational restrictions:

- The performance in terms of ransomware, data exfiltration, IoMT hijacking, and insider threats detection accuracy and per attack type (Khan A. et al., 2024; Zachos et al., 2025).
- Live detection and latency, particularly in the critical care networks where any delay in the blocking of malicious operations can be fatal (Lui et al., 2025)
- Scalability as the number of IoMT devices grows as well as the amount of telemetry (Naghib et al., 2025).
- The use of resources especially embedded devices and edge gateways (Hafid et al., 2025).

The dependence on the number of monitored devices on the average detection latency of various architectures of the IDS (centralized, edge-based, federated) can be represented in Figure 2.

Figure 2 :Latency vs. number of devices



3.3 Problems with Performance Optimization.

There is a lot of class imbalance: benign traffic much outweighs malicious events, and some classes of attacks are also infrequent. Devoid of attentive sampling, economic learning, or aberration recognition tactics, the models may be biased to the majority classes (Zachos et al., 2025; Naghib et al., 2025).

When the attacker methods or workflows of hospitals are altered (e.g. due to a pandemic) concept drift occurs. The performance of the static models decays with time, which requires constant learning and regular training (Khan A. et al., 2024; Islam et al., 2024).

False positives are especially dangerous in clinical settings, as they lead to alert fatigue and even desensitization of security personnel and IT professionals. A fundamental optimization objective is balance between detection rate and FPR is thus a core optimization objective.

3.4 Case Studies and Examples

Said et al. (2021) introduce an effective anomaly detection method of Smart Hospital IoT, stating that it has a high detection rate yet consumes only modest amounts of resources. Wahab et al. (2022) show an AI-based hybrid e-health IDS, which is highly effective compared to conventional methods on open data. Lui et al. (2025)

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

consider the aspects of cyber-biomedical in the case of the healthcare 5.0 IDS. The article by Zachos et al. (2025) describes and tests an anomaly-based IDS of the IoMT networks, generating a dataset, and comparing the ML algorithms.

TABLE 3 :Selected AI-Driven IDS Studies for Smart Hospital / IoMT Environments

Study	Context	Technique / Model	Dataset Type	Key Findings
Said et al., 2021	Smart Hospital IoT	ML-based anomaly detection	Simulated IoT traffic	High detection rate with low computational overhead
Wahab et al., 2022	IoT-enabled e-Health	Hybrid AI framework	Public IoT security data	Significantly improved accuracy over traditional IDS
Khan A. et al., 2024	IoMT-based Smart Healthcare	Deep learning anomaly detection	IoMT network traces	High detection accuracy even with imbalanced datasets
Lui et al., 2025	Healthcare 5.0	Cyber-biomedical ML feature engineering	Testbed dataset	Novel feature design dramatically boosts detection performance
Zachos et al., 2025	IoMT networks	Anomaly-based ML IDS	Custom IoMT dataset	Robust detection of IoMT-specific intrusion patterns
Almalki et al., 2024	Healthcare 5.0	Federated learning + blockchain-enabled IDS	Distributed environments	Privacy-preserving IDS architecture with improved trust and resilience

4. INTRUSION DETECTION SYSTEMS BASED ON AI ARE EXPLAINABLE.

In relation to smart hospitals, explain ability is crucial because it enables customers to trust artificial intelligence systems and their products with their vital data.

4.1 Significance of Explainability in Smart Hospitals.

There should be explanability of:

- Clinician-IT staff- administrative trust and acceptance (Mirbabaie et al., 2021; Jha et al., 2023).
- Compliance with the regulators, in particular, within the framework of the expectations of GDPR regarding meaningful information on automated decisions (Murdoch, 2021; Hussein et al., 2024).
- Such areas as forensic analysis and incident response, in which analysts have to know why an alert has been raised to research and fix as soon as possible (Zhang et al., 2022).
- Making reduction of alert fatigue through giving clear context to the operators, which would enable them prioritize important alarms.

4.1 XAI Techniques for IDSs

LIME and SHAP model-agnostic techniques will produce local explanations of individual predictions, indicating the characteristics most likely to have contributed to an alert. Muhammad et al. (2025) IDS are explained by using LIME, and Islam et al. (2024) works with the XAI techniques to address the transparency gap in network IDS. Hosain and Cakmak (2025) illustrate the use of XAI-XGBoost in which the inbuilt feature importance is enhanced with XAI techniques.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Attention mechanisms, saliency maps, and gradient-based attribution make deep learning-based IDSs able to visualize the time steps, packets or features that the classification relied on (Zhang et al., 2022; Si-ahmed et al., 2024).
An example of a confusion matrix of an IDS model is provided in Figure 3 and SHAP-based feature importance in Figure 4.

Figure 3 – Confusion matrix heatmap

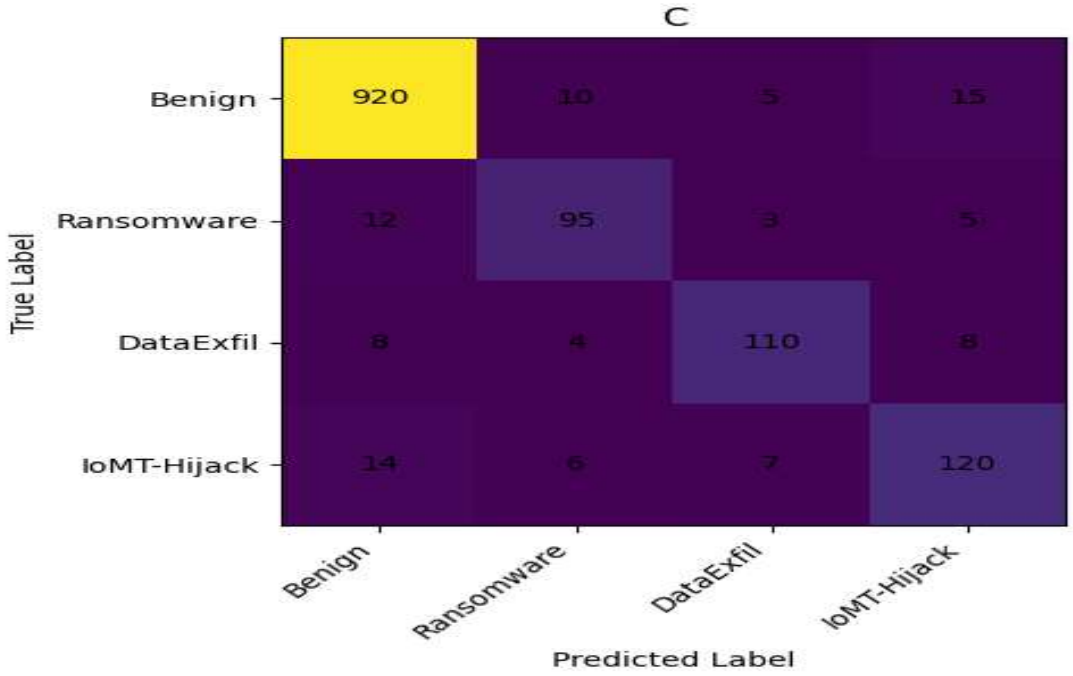
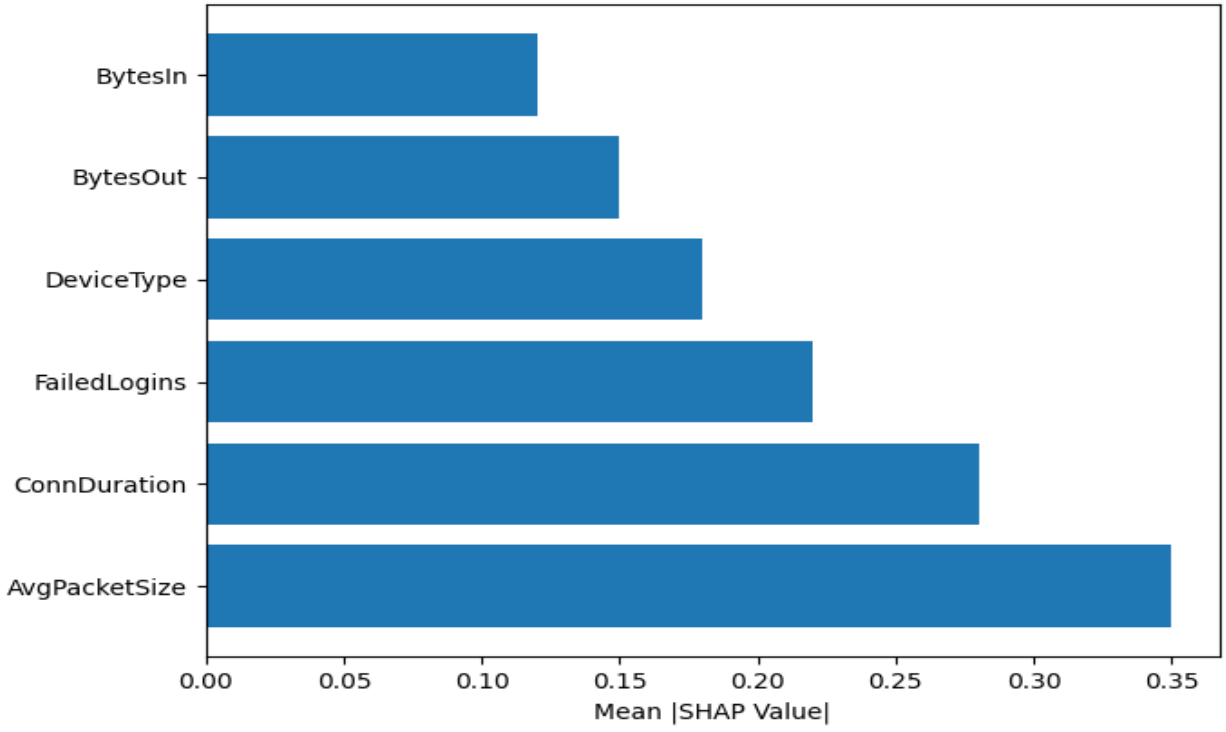


Figure 4 – SHAP-style feature importance bar chart



iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

4.2 Issues of Explainability Achievement.
Major challenges include:

- **Complexity of the model:** Deep neural networks may be viewed as black boxes, which are usually of high dimensions, and their logic cannot be determined (Zhang et al., 2022; Rjoub et al., 2023).
- **Trade-off versus performance:** Simple models are more understandable, but deep architectures may be more powerful, and also constrained devices can prefer smaller models (Hafid et al., 2025; Mohale and Obagbuwa, 2025).
- **Human factors:** Technical concepts ought to be transformed into practical wisdom to non-technical stakeholders such as the clinicians and the administrators (Mirbabaie et al., 2021; Rony et al., 2024).

4.3 Explainability Effect on Response to Incidents.
Explainable output can:

- Minimize the mean time to detect (MTTD) and mean time to respond (MTTR) by demystifying what type of behavior by what asset, user, device, etc. drew an alert.
- Promote after incident learning, whereby the explanations are inputted into the security training and policy changes.
- Audit trails and documentation support, that will assist in the compliance reporting and investigation of legal aspects (Hussein et al., 2024; Naili et al., 2025).

5.ETHICAL RISK AND ISSUES.

5.1The security of data privacy and confidentiality.
The network flows with identifiers, EHR queries, or device telemetry, which are tied to people, can be specified as specific sensitive patient information that may be processed by IDSs in Smart Hospitals. The information about health can be provided through the lack of anonymization or documentation and can be unlawful according to the law of HIPAA, GDPR, or local regulations (Murdoch, 2021; Hussein et al., 2024).
These risks can be removed by privacy-preserving analytics and federated learning (Almalki et al., 2024; Ashraf et al., 2022), which stores the raw data locally, but communicates the updates of the models, but the problem of metadata leakage also exists.

5.2 Bias and Fairness
Such AI-based IDSs may be biased towards certain device, process, or user group types in the event that the training data is biased (Jha et al., 2023; Naghib et al., 2025). One such example is that the devices in the less-resourced departments should be monitored less often or parts of the facility or suppliers should have more tags. It can lead to a skewed security position and even inequity in allocation of incident response resources.

- **Transparency and Accountability:** The company has a good financial reporting and transparency among the employees and other stakeholders.
- **The most important ethical question is the following:** who will take charge in case of a malfunction of an AI-based IDS? The hospital, the IDS vendor, and the system integrators may be liable in case of damage to the patient as a consequence of an unidentified intrusion (Gala, 2024; Biasin et al., 2024).

This is compounded by the fact that opaque models are used to produce an issue of decision process reconstruction, and this brings about questions on due process, auditability and legal defensibility (Busch et al., 2025; Naili et al., 2025).

5.3 Existing Human Oversight and Autonomy Human Oversight and Autonomy.
The danger of excessive automation is excessive trust in AI where human operators trust system reactions too much. Implementation of ethics requires:

- A human-in-the-loop must review high-impact activities (e.g. an isolating critical device).
- Clear phases of increasing the proposals of AI that get rationalized by human experts (Mirbabaie et al., 2021; Jha et al., 2023).

5.4 Regulatory and Legal Environment.
The AI regulatory landscapes in the healthcare field are evolving irregularly. Hussein et al. (2024) propose a model of maturity of AI governance in healthcare, whereas Busch et al. (2025) contaminate the world on AI governance in healthcare. Since it is pointed out as Naili et al. (2025) and Biasin et al. (2024) say, cybersecurity is pushing current legal systems to their limits, and AI-based diagnostics.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Table 4. Ethical risks and corresponding mitigation strategies

Ethical Risk	Mitigation Strategy	Relevant References / Concepts
Privacy breaches	Data anonymization; federated learning; strict log policies	Murdoch (2021); Almalki et al. (2024); Ashraf et al. (2022)
Bias & unfair detection	Representative datasets; fairness checks; bias audits	Jha et al. (2023); Naghib et al. (2025)
Lack of transparency	XAI methods; documentation; model cards	Zhang et al. (2022); Rjoub et al. (2023); Muhammad et al. (2025)
Accountability gaps	Governance models; audit trails; clear liability structures	Hussein et al. (2024); Busch et al. (2025); Gala (2024)
Over-reliance on AI	Human-in-the-loop oversight; operator training	Mirbabaie et al. (2021); Rony et al. (2024)
Regulatory non-compliance	HIPAA/GDPR alignment; AI-specific compliance processes	Naili et al. (2025); Biasin et al. (2024)

DISCUSSION

6.1 Performance, Explainability, and Ethics Interplay.

The concepts of performance, explainability, and ethics are closely interrelated:

- Models that perform highly and are opaque at the same time can have a great detection rate but be unacceptable legally and ethically.
- Very interpretable models can lose minor pattern recognition creating lower security efficacy.
- The data to train or ethics, including fairness and privacy, can limit the available data (or necessitate more intricate architectures e.g. federated learning), which impacts performance and complexity (Almalki et al., 2024; Ashraf et al., 2022).

6.2 Finding a Balance between Hospital Priorities and Security Needs.

Smart Hospitals have to choose between security, clinical efficacy, economic reasons, and usability. The security policies may cause friction to the clinicians and staff, and failure to invest adequately in security may expose patients.

AI-assisted IDSs must consequently be part of a risk-based security policy, in which models can be adjusted based on the sensitivity of the systems being secured and allowed false positives.

6.3 Implementation into Practice.

Practical suggestions would include;

- Multipurpose deployment of IDS components (device-level, network-edge, data-center) in order to minimize blind spots (Said et al., 2021; Zachos et al., 2025).
- Federated and privacy-sensitive training adoption where feasible to meet the data protection laws (Almalki et al., 2024; Ashraf et al., 2022).
- XAI dashboards incorporated into security operations centers (SOCs) to give alerts in the form of human-readable explanations (Hosain and Cakmak, 2025; Islam et al., 2024; Si-ahmed et al., 2024).

Setting up of governance frameworks to establish roles, duties, and responsibility of AI security mechanisms (Hussein et al., 2024; Busch et al., 2025)

7 CONCLUSION

7.1 Summary of Key Findings

In this paper, the authors have discussed AI-based IDSs in Smart Hospitals in three axes, namely, performance, explainability, and ethical risks. We highlighted that:

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

- The AI-based IDSs, utilizing deep learning, reinforcement learning, and federated learning, have better detection capabilities and flexibility than traditional IDSs in the face of a complex IoMT-intensive environment (Khan A. et al., 2024; Shaikh et al., 2025; Lui et al., 2025; Almalki et al., 2024).
- The KPIs that should be taken into account in the context of a performance evaluation relate to the healthcare domain, i.e., latency, scalability, and resource constraints, and address issues like class imbalance and concept drift (Zachos et al., 2025; Naghib et al., 2025).
- Easy-to-understand AI methods are necessary to establish confidence, facilitate forensic processing, and fulfill regulatory requirements; however, there are still contradictions between explainability and uncoded detection efficacy (Hosain and Cakmak, 2025; Muhammad et al., 2025; Mohale and Obagbuwa, 2025).
- The risks that are ethical in nature include privacy, bias, accountability, and over-reliance on automation, which are not trivial, and need to be proactively addressed with the help of governance, technical safeguards, and human oversight (Murdoch, 2021; Mirbabaie et al., 2021; Hussein et al., 2024; Busch et al., 2025; Biasin et al., 2024).

7.2 Importance of the Research.

The paper in question will be of great significance not only to nursing students and researchers but also to nurses and other medical professionals that want to make a step forward in their career.

The discussion highlights the fact that implementation of AI-based IDSs in Smart Hospitals is not merely technical issue, but also a socio-technical and ethical problem. Effective security cannot be pursued without explainability and governance, as otherwise the hospitals are likely to exchange one group of vulnerabilities with another.

8 FUTURE WORK

8.1 Research Directions

Future directions that are promising are:

- Constructing domain-specific AI systems that are specific to Smart Hospital processes and device systems as opposed to general network IDS systems (Naghib et al., 2025; Zachos et al., 2025).
- Moving towards real-time, explainable IDSs, where the descriptions are produced simultaneously with alerts and made as human-friendly as possible (Islam et al., 2024; Muhammad et al., 2025; Si-ahmed et al., 2024).
- Bringing ethical theories to the model construction, including fairness limitation, privacy budgets, and mechanisms of accountability embedded on processes of model training and deployment pipelines (Hussein et al., 2024; Jha et al., 2023).
- Open-ended longitudinal investigation of AI-based IDS implementations in operational hospital setting, monitoring the results of security, user confidence and organizational adjustment.

8.2 Policy and Regulatory Recommendations.

Policy-focused work should:

- Donate to guidelines and norms of AI-based IDSs in healthcare infrastructures that are critical.
- Elaborate liability structures and standards regarding transparency and explainability of the AI-based security devices.
- Promote regulatory sandboxes in which hospitals and vendors and regulators can run tests to test new security technologies in controlled settings (Busch et al., 2025; Hussein et al., 2024; Naili et al., 2025).

REFERENCES

1) Khan, B., Fatima, H., Qureshi, A., et al. (2023). Drawbacks of artificial intelligence and their potential solutions in the healthcare sector.
2) Murdoch, B. (2021). Privacy and artificial intelligence: Challenges for protecting health information in a new era.
3) Kumar, M., Kumar, R., Arisham, D. K., et al. (2025). Emerging AI impact in the healthcare sector: A review.
4) Rony, M. K. K., Numan, S. M., Akter, K., et al. (2024). Nurses' perspectives on privacy and ethical concerns regarding artificial intelligence adoption in healthcare.
5) Gala, K. M. (2024). Ethical and legal considerations in AI-driven health cybersecurity.
6) Pasricha, S. (2022). AI ethics in smart healthcare.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

7) Mirbabaie, M., Hofeditz, L., Frick, N., et al. (2021). Artificial intelligence in hospitals: Status quo of ethical considerations to guide future research.
8) Jha, D., Rauniyar, A., Srivastava, A., et a. (2023). Ensuring trustworthy medical artificial intelligence through ethical and philosophical principles.
9) Khan, A., Rizwan, M., Bagdasar, O., et al (2024). Deep learning-driven anomaly detection for IoMT-based smart healthcare systems.
10) Shaikh, J. A., Wang, C., Us Sima, M. W., et al. (2025). A deep reinforcement learning–based intrusion detection system for IoMThealthcare networks.
11) Wahab, F., Zhao, Y., Javeed, D., et al. (2022). An AI-driven hybrid framework for intrusion detection in IoT-enabled e-health.
12) Hafid, A., Rahouti, M., & Aledhari, M. (2025). Optimizing intrusion detection in IoMT networks through interpretable and cost-aware machine learning.
13) Lui, P. H., Siqueira, L. P., Kazienko, J. F., et al. (2025). Performance of cyber-biomedical features for intrusion detection in Healthcare 5.0.
14) Hosain, Y., & Çakmak, M. (2025). XAI-XGBoost: An explainable intrusion detection approach for securing Internet of Medical Things systems.
15) Islam, M. T., Syfullah, M. K., Rashed, M. G., et al. (2024). Advancing transparency and trustworthiness in network intrusion detection with explainable AI.
16) Muhammad, A. E., Yow, K. C., Bačanin-Džakula, N., et al. (2025). L-XAIDS: A LIME-based explainable AI framework for intrusion detection systems.
17) Mohale, V. Z., & Obagbuwa, I. C. (2025). Integration of explainable artificial intelligence in intrusion detection systems: A systematic review.
18) Said, A. M., Yahyaoui, A., & Abdellatif, T. (2021). Efficient anomaly detection for smart hospital IoT systems.
19) Narayanan, S. N., Joshi, A., & Bose, R. (2020). ABATe: Automatic behavioral abstraction technique to detect anomalies in smart cyber-physical systems.
20) Zhang, Z., Al Hamadi, H., Damiani, E., et al. (2022). Explainable artificial intelligence applications in cybersecurity: State-of-the-art review.
21) Rjoub, G., Bentahar, J., Abdel Wahab, O., et al. (2023). A survey on explainable artificial intelligence for cybersecurity.
22) Naghib, A., Gharehchopogh, F. S., & Zamanifar, A. (2025). Intrusion detection systems in the Internet of Medical Things: A comprehensive literature review.
23) Hussein, R., Zink, A., Ramadan, B., et al. (2024). Advancing healthcare AI governance: A comprehensive maturity model.
24) Busch, F., Geis, J. R., Wang, Y., et al. (2025). AI regulation in healthcare around the world: Status and challenges.
25) Naili, Y. T., Mangkunegara, I. S., Purwono, P., et al. (2025). Regulatory challenges in AI-based diagnostics: Legal implications in medical applications.
26) Biasin, E., Kamenjašević, E., & Ludvigsen, K. R. (2024). Cybersecurity of AI medical devices: Risks, legislation, and challenges.
27) Almalki, J., Alshahrani, S. M., & Khan, N. A. (2024). A secure Healthcare 5.0 system using federated learning, intrusion detection, and blockchain.
28) Ashraf, E., Areed, N. F. F., Salem, H., et al. (2022). FIDChain: Federated intrusion detection system for blockchain-enabled IoT healthcare applications.
29) Si-ahmed, A., Al-Garadi, M. A., & Boustia, N. (2024). Explainable machine learning–based security and privacy protection for Internet of Medical Things systems.
30) Zachos, G., Mantas, G., Porfyrakis, K., et al. (2025). Anomaly-based intrusion detection for IoMT networks: Design, dataset generation, and machine learning evaluation.