STAP
Smart Technologies Academic Press

Journal of Cyber Security and Risk Auditing

https://www.jcsra.thestap.com/

Check for updates

Machine Learning Approaches to Mitigate Insider Threats in Electronic Health Records Systems

Abdullah Alessa[1], Yaseen Alduwayl[1], M M Hafizur Rahman[1]

[1]Department of computer networks and communications, College of Computer Sciences and Information Technology, King Faisal University, Al-Ahsa, 31982 Saudi Arabia

ARTICLE INFO

Article History

Received: 17-10-2025
Revised: 29-11-2025
Accepted: 22-12-2025
Published: First Online

Vol.2026, No.1

DOI:
https://doi.org/10.63180/jcsra.thestap.2026.1.1

***Corresponding author.**
Email:
226031653@student.kfu.edu.sa and mhrahman@kfu.edu.sa

Orcid:
https://orcid.org/0000-0001-6808-3373

ABSTRACT

The increasing of convert the healthcare system to be as digital system, becoming necessary to secure electronic health records against inside threats. Although the existing numerous research provides solutions and utilizes machine learning techniques to enhance security of the records, it remains fragmented and lacks comprehensive synthesis of approaches. In this paper aims to present a comprehensive and rigorous study for an appropriate method based on ML to detect and mitigate insider threats in EHR systems. PRISMA methodology is used to follow. It starts to analyze 537 primary studies from different major databases where it looks to identify the relevant research in the same scope. Then it reduced to 25 studies after applying the standards of inclusion and exclusion. This study focuses on comparing the algorithms and techniques that have been used, the challenges faced with the implementation process, and the gap of existing studies. The study findings reveal trends, limitations, challenges, and future directions to develop and support the intelligence system and protect privacy and secure healthcare records.

Keywords: Machine Learning, Electronic Health Records EHR, Insider Threats, Break-the-Glass, Data Security.

How to cite the article
Alessa, A., Alduwayl, Y., & Rahman, M. M. H. (2026). Machine Learning Approaches to Mitigate Insider Threats in Electronic Health Records Systems. Journal of Cyber Security and Risk Auditing, 2026(1), 1–19. https://doi.org/10.63180/jcsra.thestap.2026.1.1

1. Introduction

Technology adoption has changed the health care business at a rapid pace, and electronic health records have been one of the most important developments. Electronic Health Records (EHR) systems electronically store, maintain, and share patients' medical histories, diagnoses, treatment plans, test results, and other important data Figure 1. That leads us to enhance efficiency, accuracy, and accessibility [1] [2]. EHRs allow physicians and health institutions to make timely decisions, coordinate treatment, and apply the best practices by getting manual work documents. While these benefits are substantial, electronic health information creates new and important security and privacy problems and risk. Electronic records have to be discreet, secure, and accessible to preserve patient confidence, compliance with regulatory requirements, and sensitive medical information [1] [3].

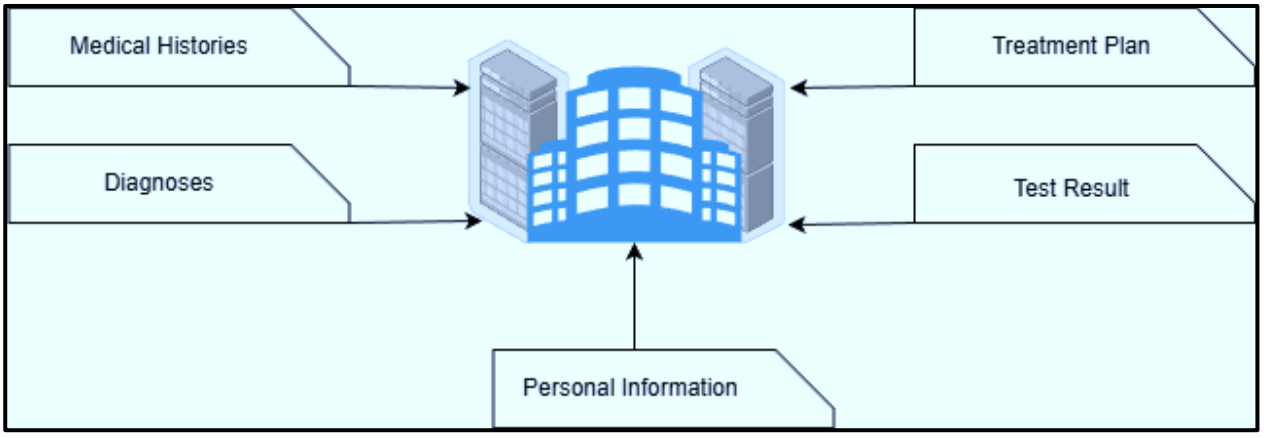

Figure 1. EHRs critical elements on Smart hospital system

EHR implementation has been driven by policy regulations, technical innovation, and desire to have interoperable health care in the world [1] [2] [4]. Alongside these the best practice and the policy of the organization must enforce the insider employees to compliance with their instructions in terms of the usage of sensitive data in a proper way and in the case, they need it only. Digital health care practices are embraced by the developing countries as they continue to modernize [5] [6]. The cyber threats evolve more rapidly than traditional security mechanisms, which make health records as an attractive goal for attackers due to the richness of such details in personal identifiers and clinical sensitive data [7] [8].

Threats include insurance fraud, blackmail or illegal sale of data may be committed with theft records [6] [9]. Furthermore, the employees who work in healthcare can bring a huge danger to EHRs, while they avoid permissible limits and access to sensitive data about patients such as infections and infertility [10]. Although significant, firewalls, encryption, and access control are often inefficient to prevent sophisticated techniques of advanced outside attacks and insider threats. In addition, healthcare organizations are also facing the problem of budgetary constraints, outdated technology, and the shortage of cybersecurity skills [11] [12].

As a result of these limitations, it becomes necessary to develop smart protection tools, flexible, and compatible with dynamic threats. Machine learning (ML) is an essential part that promises to address the dynamic cyber threats and detect misuse of privileges [2] [13] [14]. ML guides systems offer early detection, ongoing monitoring, and automatic protection measures instead of breach reactive response. It comes at a time when health care is transitioning towards predictive analytics where algorithms driven by data is used to predict the occurrence of diseases, improve treatment methods, and increase efficiency. This is natural and necessary to inject predictive capabilities into security [15] [16]. Deep learning algorithms retrieve complex information of raw data to enhance the detection rate which addressing the existing health care infrastructure requirements [17] [18] [19]. According to recent research, ML security frameworks are becoming increasingly helpful to identify abnormal patterns of access to the patient records and reduce the cases of data sensitive leakage [20]. However, the challenges like adversarial attacks, are increasingly concerning issues that where attackers change input data to manipulate ML models [21] [22].

Moreover, insider attack which may cause a great damage since they consider as an authorized user, making them hard to detect, due to the techniques such as flexibility and resiliently [23]. Many methods are already used in different techniques deploying to detect insider threats such as Local Outlier Factor, One Class Classification (OCC), Isolation Forest (IF), and Self-Organizing Map they selected to be trained in EHR logging data [24]. However, small, clinics and remote hospitals might not have difficulties to implement advanced ML solutions due to the limitations of their infrastructure, while larger institutions may suffer from weak of coordination between their units [2] [19].

Despite the development of EHRs system, still insider threats consider as a dangerous, especially in case of misuse of permitted emergency access (Break-the-Glass). Machine learning (ML) techniques show a promising capability for detecting these behaviors but often lack clear explanation. Which prevents practical implementation. The urgent need of intelligence solutions able to analysis the users' behavior and detect abnormal patterns in understanding way and reliable. From here, the importance of this study is to make a comprehensive and rigorous study for pervious primary research, which aims to identify the gap and present integrated vision towards framework that is more effective and transparent to protect and preserve patient's data.

Although of the variety of previous studies that exploring the usage of ML techniques in field of abnormal detection, predicting threats, and securing healthcare system. However, these efforts are still various and uncompleted and lack the ability to face real operational challenges. So, when a single security compromise occurred, it cost the organization huge funds, making patient's life at risk, and it brings a negative impact on organization's reputation.

The most outstanding challenges that EHR systems face in terms of cybersecurity are the difficulty detecting insider threats. Also, the lack of transparency in AI models, and the limitations of real datasets to train models. Those systems suffered from the governance of access privileges with missing precis mechanisms for tracking and monitoring the privileges misuse, especially in emergency situations (Break-the-Glass). In addition, the restrictions on technologies that deploy at small hospitals play as a main obstacle to performing complex protections models.

The fundamental problem presented by this study is the need to develop advanced techniques to detect misuse of permitted emergency access (Break-the-Glass) in EHR systems, especially by authorized users who exceed the scope of their role and tasks. This type of insider attack is usually not detected by traditional protection mechanisms that are based on authentication and privileges control. The danger of this gap is compounded due to not notice by anyone from management and technical teams for long time, which causes disclosure of the patient's privacy to be exposed, and compromise the organization to ethical and legal risk.

This standardized contemporary study aims to perform scientific literature that relevant to using machine learning to detect insider threats related to misuse of permitted access in EHR, especially in the context of ‘Break-the-Glass’ privileges. This study focuses on analyzing and evaluating how the previous research relying on analysis behavior methodologies, explainability, the accuracy of detecting abnormalities patterns, additionally covering the types of algorithms used and the challenges of applying them in healthcare environments.

For achieving this, there are certain study questions as below:

- **RQ1:** What are the most machine learning techniques used nowadays to detect insider threats in EHR systems?
- **RQ2:** To what extent the literature covering the concept of "Break-the-Glass misuse" as a type of insider threat?
- **RQ3:** How do they deploy analyzing the behavior used in detecting misuse of privileges?
- **RQ4:** To what extent is the research relaying on explainability in security solutions design?
- **RQ5:** What are the most outstanding challenges and limitations facing the implementation of these models in real healthcare environments?

This study contributes to filling a clear gap in literature through providing comprehensive and rigorous contemporary study for the research that presents the usage of ML for detecting insider threats in EHR systems, with focusing on misuse of permitted emergence access (Break-the-Glass). Also, highlighting the extent deploying analyzing behaviors and explainability with modern security solutions design.

In this study we will go through five more sections, starting from section 2 is Explaining the study methodology and showing the steps of selection of the studies and analyzing them using PRISMA framework. Then in section 3 synthesize and analyze the outcomes of included studies and discussing the new trend in the scope. We will next Discusses the outstanding challenges and clears the existing gap and illustrate the challenges and limitations on research's topic and the obstacles in the same field in section 4. Furthermore, section 5 is presenting the future research recommendations. Finally, section 6 the conclusion and the main results of the study.

2. Study selection Process – PRISMA approach

2.1 Databases and Search Keywords

To ensure the search process is a comprehensive, it is done through 4 main databases Google Scholar, Semantic Scholar, IEEE Xplore, and Saudi Digital Library (SDL). Also, been used group of keywords are relevant to the topic such as Machine Learning (ML), Electronic Health records (EHR), data security, insider security, misuse of emergence permitted (Break-the-Glass), patient's privacy and smart hospital.

2.2 Inclusion and Exclusion Criteria

The inclusion process joins studies of the last 5 years, which focus ML applications on secure healthcare records systems, especially for those related to insider usage behavior and threats coming with it. The exclusion process happens if the studies were: (1) duplicated, (2) experiential without real application, (3) not related to healthcare field, (4) not available the entire text, (5) not address research questions, (6) not appropriate methodology of preforming the research, or (7) not ineligible for inclusion in contemporary study as shown on Figure 2.

2.3 Final Selection

After studying the full text, due to weak relation to topic 61 were excluded, or usage of techniques that are not relevant, or not mention to insider behaviors. Then 32 studies were excluded because they were not focused on behavior analysis or it was relatively old. Thus, 25 papers were approved to be as main sources for securitization and summarization on this paper.

2.4 Data Extraction and Analysis Plan

After final selection of the studies, an organizing plan was followed to extract and analyze the data in way to help answer the research questions was written before and achieve the goals of this study. An analyzing table was created to contains number of elements for each study, including:

- The topic of the study.
- Publish year.
- The method used.
- The kind of insider threat was selected.
- Technology and algorithms.
- If the study uses behavior analysis.
- Dose it includes explainability of the results.
- The limitations and challenges have been mentioned.

2.5 Inclusion and Exclusion Criteria

All the studies that have been found across through several points assigned in earlier stages to match them with and select which ones appropriate with this paper and which ones have missing aspects of these points. In Table 1 shows the key points that based on.

Table 1. Inclusion and Exclusion information

Item	Inclusion	Exclusion
Type of paper	Systematic review, published in journals or scientific conferences	Unreviewed papers, unofficial reports
Scope	ML studies, EHR analysis, and insider threats detections	Irrelative studies to EHR or not focusing on insider threats
Period	Published during last 5 years	Before 2020
Language	English	Other languages
Data availability	Contains a describing method of collecting the data or using real or synthetic data	No data contained nor clear on method of collecting it

2.5 Study Selection Process (PRISMA)

To organize the process of synthesizing of contemporary research in standardized way PRISMA framework is used to accomplish this stage, which facilitates the steps of searching, filtering, and standards of selecting the studies to ensure transparency, the ability to avoid the duplication, and reduce the bias. Moreover, PRISMA considers as approval standards that internationally reliance to guarantee the quality of synthesis of contemporary research.

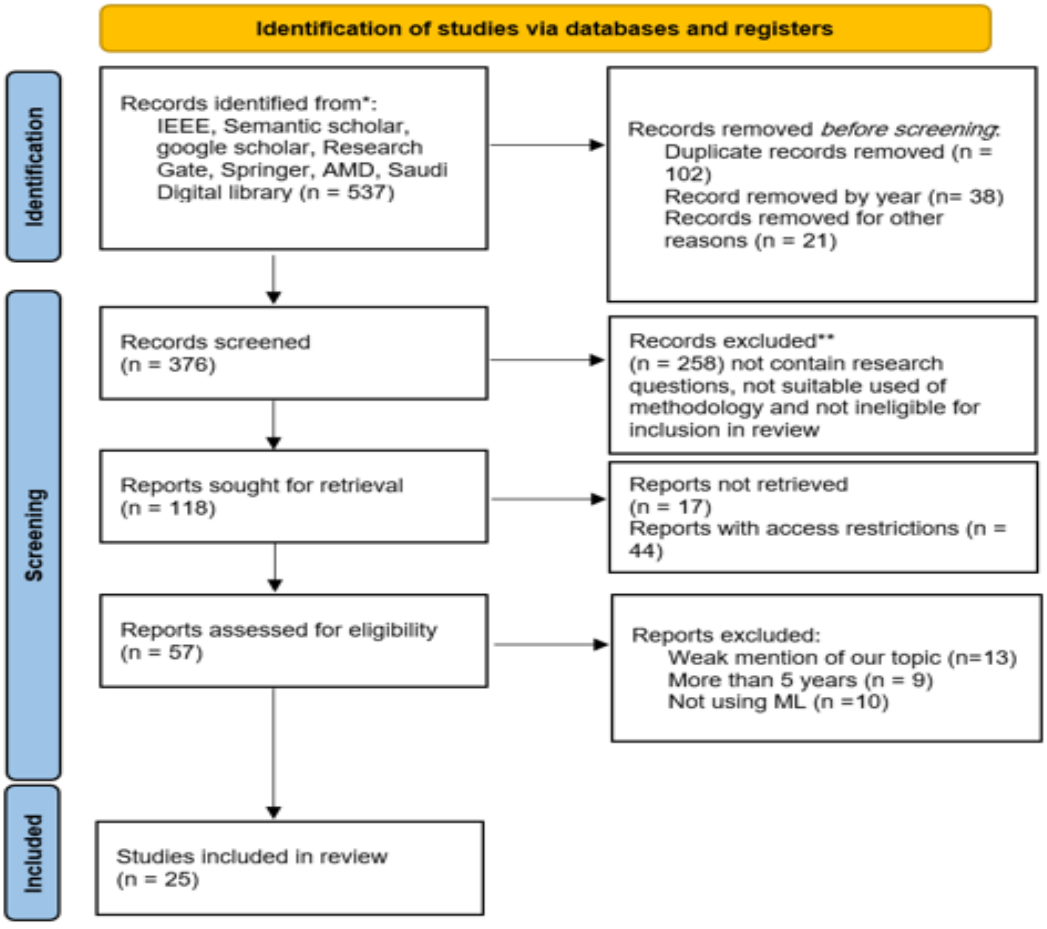

Figure 2: Methodology of selection of primary studies for comprehensive and rigorous standardized contemporary study using PRISMA

3. Challenges and Techniques

This table aims to facilitate the comparison among studies, and address the directions of research trends, and the existing gap in literature, which support systematic synthesis and evaluation process and driven to critical analysis in coming sections. This research [1] aims at the critical security and privacy issues in Internet of Medical Things (IoMT) systems using an Intrusion Detection System (IDS) framework that uses Artificial Neural Networks (ANN) for anomaly detection, Federated Learning (FL) for privacy preservation, and Explainable Artificial Intelligence (XAI) for model transparency. The research used multiple datasets such as Ton-IoT, UNSW NB15, NSL-KDD and WUSTL-EHMS to test the proposed framework against FL and centralized approaches. Performance metrics showed that FL had similar results as centralized methods, showing accuracy rates of more than 98% in most of the data sets and AUC scores of about 0.998. The methodology took advantage of ANN architectures optimized using different FL parameters, such as number of clients and fraction fit and local epochs. SHAP analysis was used for post-model interpretation to make decision making transparent. The key findings revealed that FL keeps high performance while delivering dual benefits of privacy preservation and model explained without violating regulations such as HIPAA and GDPR. Limitations included dataset specific communication rounds exhibiting inverse trends in the WUSTL-EHMS dataset due to the small sample size which is a concern for federated datasets to have larger, more representative datasets to ensure optimal model generalization.

The study [2] introduces two stages of modelling framework combining and gathering sequential encoder decoder networks with Generative Adversarial Networks (GANs) to generate high fidelity, privacy preserving synthetic Electronic Health Records. The framework was evaluated on two large-scale real-world datasets MIMIC-III and eICU, focusing critical challenges in EHR data including frequency, sparsity and timely dynamics with varying sequence lengths. The Performance metrics show that synthetic data keeping almost the same fidelity to real data with less than 3% accuracy difference for models trained on synthetic against real data while achieving high privacy metrics. The methodology handled coexisting numerical and categorical features with static and time varying components, included developed approaches for encoding and decoding, normalizing complex distributions and presenting missing data. The key findings revealed that HER Safe make a good performance in previous GAN based approaches by large margins on both fidelity and privacy evaluation metrics. Limitations included exclusion of unstructured data such as images and free form text from the scope, which did not include important components in real world EHR systems.

This paper [3] presents many models of security approaches integrating advanced technologies to strengthen EHR data privacy and resilience. The model combines BERT-CNN for data standard, game theory based hyperparameter concept to be used for performance optimization, and fully homomorphic encryption (TFHE) to secure data confidentiality. Moreover, Ethereum blockchain and RESTful API mechanisms are used for authentication and decentralized data management. They used the SyntheaTM dataset, the suggestion system achieved a 99% accuracy which represent a 23% improvement over conventional BERT and Random Forest baselines. The article verifies the scalability, interoperability, and ethical adherence of the health care system based on the cloud. Its underlying innovation is the combination of AI-powered analytics and privacy controls that are enabled by blockchain to enable end-to-end protection against unauthorized access points and insider abuse. The results outline the need of hybrid and resilient structures that can guarantee privacy protection and efficiency of operations in healthcare settings.

This research [4] proposes a framework connecting Differential Privacy (DP) with Federated Learning (FL) to improve data protection with Internet of Medical Things (IoMT) applications particularly for Tuberculosis detection use chest X-ray datasets. The study developed and compared seven different Convolutional Neural Network (CNN) models among three available and public chest X-ray datasets applying ensemble learning with noise addition to model result before differentiation. The Performance evaluation utilized comprehensive metrics presenting the framework privilege over baseline models in keeping data confidentiality. The methodology cleverly integrated FL's decentralized feature with the mechanism of DP to prevent model output rebuild and hence tackle data rebuild privacy concerns in sharing of healthcare data. The results showed that the DP aided FL framework is effective in preserving privacy while preserving model accuracy offering a strength solution for secure AI applications in healthcare. Lack of generalization: The focus on one specific medical imaging application (Tuberculosis detection) implies that the privacy preservation approach should be evaluated on various healthcare tasks and modalities to identify the generalizability of the approach.

This research [8] presents a multi layers encryption security framework for patient information monitoring systems integrating Advanced Encryption Standard (AES) and Secure Hash Algorithm 256 (SHA-256) with machine learning based anomaly detection mechanisms for real time threat identification. The methodology used multiple machine learning algorithms for capturing unusual behavior and suspicious activities using Electronic Health Records while ensuring strong data protection during storage and transmission. The framework achieved high accuracy in unusual behavior capturing while keeping effective data encryption presenting practical activity in healthcare security applications. The method use Hash based Message Authentication Code (HMAC) to make sure about integrity and authenticity of transmitted data through dual stage hashing with secret keys. The key findings highlighted the effectiveness of gathering encryption with machine learning in generating comprehensive security solutions. Limitations included insufficient detail regarding specific performance metrics and dataset characteristics and the paper focused as primary on framework design without overall experiential validation among diverse healthcare environments.

Nowrozy et al. [9] introduce the universal privacy framework of EHR systems based on ontology modeling and the use of machine learning to enhance privacy management across healthcare platforms. The study focuses on the long-standing conflict between accessibility, usability, and compliance to the international requirements in the form of GDPR and HIPAA. The authors were able to draw a distinction between legitimate and illegitimate privacy policies and automatically identify personally identifiable information (PII) in textual records by using BERT-based models and especially DistilBERT. This integration supports a context-sensitive enforcement of privacy, which enforces adaptive privacy control in distributed health data systems such as the UK NHS and the Australian My Health record. The ontology part of the paper formalizes the management and categorization of data, which improves the interoperability of healthcare stakeholders. Such an ontology-ML hybrid approach is essential in building dynamic, ethical, and legally compliant EHR privacy systems, helping to resolve the data sharing/patient trust issue.

Khalid and co-authors [11] provided a comprehensive survey of privacy-preserving AI techniques in healthcare sector, showing and highlighting challenges in balancing creation with patient confidentiality. The paper division the privacy attacks including such inference, data leakage, and model reversing and surveys countermeasures such as federated learning, differential privacy, homomorphic encryption, and secure many parties computation. It also discusses privacy threats in medical IoT systems and confirming the importance of GDPR and HIPAA compliance for AI deployment. The study highlights that while AI offers transforming service for a potential for clinical to take the decision, not standard datasets and ethical risks preventing the real-world adoption. The authors propose a classification of hybrid and cryptographic privacy-preserving models, making sure of securing data pipelines for machine learning and deep learning systems. This work as foundation for designing transparent and auditable AI systems that provide protection for patient data integrity while enabling secure innovation in medical research.

This paper [13] proposes an adversarial learning approach for automatic deidentification of medical records to facilitate privacy preserving data sharing for research while adhering to GDPR and HIPAA regulations. The methodology addresses the "chicken and egg" paradox in deidentification strong training sets are required to enhance automatic deidentification classifiers yet sharing large medical text corpora in a privacy preserving manner without the classifier is challenging. The method transforms and changes medical text that can't be revert back into vector representations through adversarial learning avoiding the need for manual pseudonymization and enabling secure training data transferences. The technique uses Named Entity Recognition (NER) metadata for sequence tagging of Protected Health Information (PHI) like names, locations, dates and account numbers. The key finding in this paper presents that adversarial methods avoid strong pseudonymization while keeping privacy protections fit for multi-institutional data gathering. The paper also shows limitations included the continuation of what necessity for human annotation of PHI during training and they can't revert back into transformation of data may limit downstream analysis requiring the original text information.

Bin Sarhan and Altwaijry [14] come up with a detailed insider threat detection model that uses machine learning to detect data leaks in organizational networks. They used the CERT insider threat dataset, which was applied to Deep Feature Synthesis to obtain more than 69,000 behavioral features per user, which was reduced using PCA. Both the anomaly detecting and the classification models were also tested and the SVM classifier score was 100 and the anomaly detector model was 91. The SMOTE technique was also used in the research to address the imbalance in the data set so that it increased recall but reduced precision. The authors concentrate on the creation of a scaled, automated system that can detect abnormalities in the user activity logs on a real-time basis. Their paradigm shows that further feature engineering and guided learning can be used to improve the detection of malicious insiders, which can help in the creation of smarter and more versatile cybersecurity infrastructures.

This research [17] develops dynamic feature dataset for ransomware detection comprising 2000 registers with 50 carefully chosen dynamic characteristics extracted from sandbox analysis of encryptor and locker ransomware combined with goodware samples among five platforms. The study evaluated by machine learning algorithms including Gradient Boosted Regression Trees, Random Forest, and Neural Networks using 10-fold among validation. The performance metrics show average accuracy excellent to 0.99 among all three algorithms showing high effectiveness in ransomware detection. The systematic methodology analyzed 326 features from Cuckoo reports choosing the 50 most pertinent features while focusing redundancy through analysis engagement. Key findings show that the automatic feature approach could be effectively detect all of current and emerging ransomware changes by identifying shared behavioral attributes. The limitations included the limited scope to dynamic or active analysis without connecting static analysis methods and the focus on specific ransomware families may limit generalization to novel ransomware variants with principally different operational behaviors.

The research [19] is important for this specific study which addressing insider threats within EHR systems using supervised machine learning to detect unauthorized data misuse by legitimate users. The study holding and dealing with real-world data from a UK hospital and applies classification algorithms such as Decision Tree, Random Forest, and Support Vector Machine (SVM), between the previous, SVM algorithm which achieved the highest accuracy of 0.99, effectively identifying anomalous access behavior. The authors classify insider attacks into malicious, careless, and compromised categories, focusing on the first and third types. Their findings give confirmation that the traditional security tools often overlook the necessary misuse patterns happen by the employee that ML-based anomaly detection can reveal. The research shows the feasibility of incorporating predictive models into healthcare IT infrastructures to keep the insider behavior monitored around the clock. This approach adds value to data confidentiality and compliance; it is one of the strong cases in favor of adopting ML-driven security frameworks in today's hospital environment.

This study [20] shows how the technology of ML can make Electronic Health Record (EHR) strong and secure by evaluating vulnerabilities and improving resilience within healthcare environments. The authors performing advanced classifiers such as XGBoost and LightGBM to enhance system protection, achieving more than good ROC-AUC scores of 1.00, linking excellent accuracy in distinguishing legitimate and malicious cases. The paper confirms that it is compatible with data privacy, and the ethical connect of large-scale EHR, aligning with HIPAA requirements. It highlights that combining AI with encryption and secure communication protocols is important in order to enhance the confidentiality, integrity, and availability of patient data. Furthermore, it brings that future EHR frameworks should merge ML-based anomaly detection with cryptographic methods to allocate emerging cyber threats and ensure sustainable privacy protection. Overall, this work contributes to the growing literature approval of intelligent, adaptive, and data can be protection mechanisms for sensitive healthcare systems.

This paper [21] proposes an access control scheme for Electronic Medical Record (EMR) privacy protection combining Ciphertext Policy Attribute-Based Encryption (CP ABE) with SM4 cryptographic algorithm, utilizing Ordered Binary Decision Diagrams (OBDD) as the access structure. The research uses security analysis and experimental verification against the Decisional Bilinear Diffie Hellman (DBDH) possibilities to validate the scheme's security. The given hybrid encryption method shows constant key length and high decryption efficiency compared to old CP ABE schemes utilizing access trees or Linear Secret Sharing Scheme (LSSS) structures. The methodology addressed basics challenges in EMR applications by supporting positive and negative attribute values with flexible access policy terms. The key findings present that the OBDD based methods achieved huge computational and storage performance advantages, with minimizing overhead in both key generation and decryption operations. The distinguished limitations combining the scheme's performance being close to specific OBDD structures, where too many effective paths could increase encryption overhead relative to other CP-ABE schemes, importing careful design of access policies for better efficiency.

This research [23] introduces an anomaly-based insider threat detection model that utilizes the Isolation Forest (IF) algorithm to address data not balancing in insider threat datasets, especially in the CERT dataset. Unlike prior models relying on oversampling or lack of samples techniques, this method mitigates not directed class distribution in the algorithms. The model effectively focuses on insider data leakage and malicious access by authorized users, by achieving 98% accuracy across varied violation and discloser ratios. The paper underscores the challenge of detecting minority attack cases in highly unbalanced environments and illustrates that anomaly-based detection provides superiors able to work with compared to conventional supervised classifiers. In addition, it compares the IF strategy against the fundamental models, highlighting the enhance in precision and recall, while identifying malicious insiders. This contribution provides a robust, unsupervised ML framework for health sector organizations seeking to detect precise abnormal behavior that the normal access control security systems often not notice.

This research [25] shows unusual detection methodology for Electronic Health Records (EHRs) within big and complex healthcare environments using unsupervised machine learning algorithms gathering Isolation Forest and Local Outlier Factor (LOF) clustering. Also, the study used EHR data gathered from hospitals in North England, using and applying systematic preprocessing, labeling, and evaluation methods with performance validation through Silhouette Score and Dunn Score metrics. Results present that Isolation Forest SVM reaches the highest accuracy of 99.21%, sensitivity of 99.75%, specificity of 99.32%, and F1 Score of 98.72%, while Isolation Forest Decision Tree performed well with 98.92% accuracy and 99.35% F1 Score. The method gathers both unsupervised algorithms with supervised classifiers SVM, Decision Tree, Random Forest to effectively make sure they predict unlabeled anomalies with reduced false positives. The key findings in this research show revealed the approach could identify new contextual anomalies not captured in baseline methodologies, presenting effectiveness in capturing insider threats within EHR systems. The limitations included the hospital specific dataset suggesting the need for validation among multiple healthcare institutions to ensure generalizability.

Table 3. ML Algorithms Used is used in the contemporary primary studies

Algorithms	Number of Study	Data Type
SVM	6	Real
Isolation Forest	4	Simulated
Random Forest	3	Synthetic

4. Discussion

Through systematic synthesizing and analyzing for studies that discuss the security of EHR using AI techniques and ML, it seems there is significant progress in mechanisms of detection threats design, but the clear gaps stay especially when dealing with complex insider threats. Figure 3 shows differentiation between the types of insider threats, which each type has its own mechanism of accessing files whether it was for manipulating patient's medical data or only for eavesdropping on sensitive data.

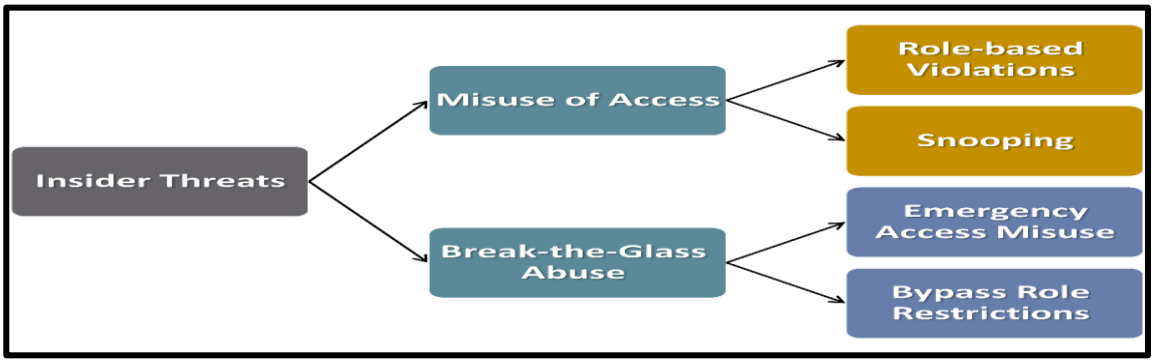

Figure 3: EHRs critical elements on Smart hospital system

Table 2. Summary of Contemporary Primary Studies on Privacy-Preserving and Secure ML in Healthcare

Ref	Objective	Problem Statement	Methodology	Results	Key Findings	Limitations
[1]	Develop explainable ML security framework for IoMT using federated learning and XAI	IoMT faces security challenges, Centralized IDS raises privacy concerns and Black-box models lack transparency	ANN + Federated Learning + XAI (SHAP), Multiple datasets, LIME and SHAP explanations	FL comparable to centralized, High accuracy, SHAP interpretable decisions	FL comparable while preserving privacy, Single point of failure eliminated	Google Colab only, Device capability assumptions, Network connectivity assumptions and Heterogeneity challenges
[2]	Develop generative framework for high-fidelity, privacy-preserving synthetic EHR data	Privacy blocks data sharing, Anonymization tedious / costly / distorts features, need synthetic data maintaining fidelity	Two-stage: Encoder-decoder networks + GANs, Heterogeneous, sparse, time-varying features	MIMIC-III: Best model only 0.026 AUC worse, eICU: Only 0.009 AUC worse, Membership inference near random	<3% fidelity gap, Realistic statistics maintained, GANs handle complex EHR	Limited to numerical/categorical features, Excludes images/text, Two datasets only
[3]	Propose integrated methodology with Game Theory, BERT-CNN	Insufficient EHR security, Centralized vulnerabilities, Data exchange difficulties and Privacy compromise	Game Theory hyperparameter optimization + BERT-CNN, Synthea dataset	99% accuracy, 99.4% precision, 98.7% recall, 99.5% F1-score, 23% improvement over BERT+RF	Game Theory significantly improves performance, Blockchain ensures transparency	Theoretical without clinical deployment, Scalability concerns, Computational efficiency trade-offs needed
[4]	Develop privacy-preserving framework combining FL and DP for secure IoMT healthcare	Centralized AI poses privacy risks, FL leaks info via attacks, Privacy-productivity tradeoff	FL + Differential Privacy; NoisyMax + Laplacian noise, Ensemble CNN, TB detection on chest X-rays	FL maintains centralized performance; High accuracy across datasets, Privacy-utility balanced	FL+DP maintains accuracy with memorization; HIPAA / GDPR adherent; Eliminates single point	Google Colab, Device capability assumptions, Connectivity assumptions; Scalability unclear
[8]	Enhance patient monitoring security via dual-layer encryption and ML anomaly detection	Ransomware / insider / breach vulnerabilities, Single-layer insufficient, Real-time detection crucial	Dual-layer: AES + SHA-256, ML anomaly detection, 2FA, PKI, Digital Signatures, SSL/TLS; HIPAA compliance	Effectively secures data; High anomaly accuracy; Real-time low latency; Multiple auth prevention	Multi-layer effective; AES-SHA256 robust; Real-time suspicious pattern detection; Multiple threat vectors	Limited ML algorithm discussion; No clinical validation; Scalability missing; Overhead not quantified
[9]	Develop universal privacy model for EHR systems using ontology and ML	Privacy breaches despite HIPAA / GDPR; Generic standards fail to protect; Inadequate policies expose data	Techniques: BERT, DistilBERT, Albert, RoBERTa models	DistilBERT: 94% accuracy, 94% precision, 94% F1-score	DistilBERT most accurate for policy classification	Limited 169 policies dataset; Google Colab only; No diverse patient group evaluation; Needs larger datasets

[11]	Provide overview of privacy-preserving AI techniques	Non-standardized records, limited datasets, stringent privacy requirements, Privacy concerns with ML/DL	Survey of cryptographic (homomorphic, SMPC), non-cryptographic (differential privacy)	FL COVID-19: AUC 88.15%, Sensitivity 73.31%, Accuracy 91.93%	FL maintains privacy with high accuracy, Hybrid techniques combine methods and encrypted computation	Trade-offs between accuracy and efficiency, Privacy attacks still threaten FL, Scalability limitations, Need standardization
[13]	Develop automatic de-identification using adversarial learning to remove PHI preserving privacy	Manual de-identification labor-intensive / error-prone, Pseudony mization costly, Need universal tools	Domain-Adversarial Neural Network (DANN), LSTM-CRF for NER, Gradient reversal and Noise injection	FastText: 97.4% F1, 96.89% GloVe	FastText superior for medical text; Noise / reversal effective; Sharing without re-id risks	Limited datasets; Lack of real-world deployment
[14]	Develop intelligent ML model for insider threat detection using Deep Feature Synthesis	Difficulty extracting behavioral features, Lack of advanced engineering, Imbalanced datasets bias results	DFS generating 69,738 features, PCA dimensionality reduction, SVM, NN, AdaBoost, RF, SMOTE balancing, CERT dataset	OCSVM: 86% accuracy; Isolation Forest: 91% accuracy; With SMOTE SVM: 100% accuracy / precision / recall /F1	SVM perfect performance, Feature engineering critical, SMOTE improves recall	CERT dataset only; SVM may overfit; Scalability unclear; Generalization needs evaluation
[17]	Detect encryptorlocker before encryption	Signatures fail due to evolution, Lack of public dynamic datasets, Parameters not explained	Five Windows platforms, 326 to 50 features; 10-fold CV	>0.99 accuracy all algorithms; Gradient Boosting best; RF and NN high; 2000 records created	Relevant features identified, Early detection before damage, Reproducible research	Windows only; Feature selection may bias; Generalization not validated
[19]	Detect insider data misuse in EHR systems using supervised ML	EHRs equally at risk from insider attacks, Security apps focus on outer boundary, Missing internal misuse	Supervised classification (SVM, Random Forest, Decision Tree)	SVM: Accuracy 98.96%, Sensitivity 97.96%, Specificity 98.97%	SVM ideal for insider threat detection, Anomalous patterns effectively detected	Dataset size differences, Limited to UK hospital data, Focused on malicious / compromised attacks only
[20]	Enhance EHR security and assess vulnerabilities using ML classifiers and multi-layer defense	Data confidentiality, privacy, integrity attacks major concerns, Traditional methods insufficient for cloud EHR	XGBoost + LightGBM ensemble; Triple-layered defense: ECC, AES, Hyperledger blockchain	AUC = 1.00 (near-perfect), Perfect ROC curves, Flawless classification performance	Blockchain guarantees data integrity, Outstanding ML performance	Theoretical framework without extensive real-world validation, Scalability testing needed, Advanced protocols needed
[21]	Propose access control scheme using attribute-based encryption with OBDD	Need efficient scheme with flexible policies	SM4 encryption, Hybrid approach	Constant key length, Superior performance	OBDD enables flexible policies, Hybrid reduces overhead	Theoretical framework, No recent CP-ABE comparison and Scalability analysis missing, No clinical deployment

[23]	Detect insider threats using anomaly-based Isolation Forest addressing class imbalance	Insider threats exploit authorized access, ML fails with skewed class distribution, Poor minority detection	Isolation Forest addressing class imbalance at algorithm level	IF: 98% accuracy, Baseline methods poor minority detection, IF outperformed in balanced performance	IF addresses imbalance at algorithm level, Generalizable to other anomalies	Limited to CERT dataset, Scalability concerns, Needs real-world org data, Limited DL comparison
[25]	Secure EHR in smart healthcare using unsupervised ML for anomaly	Anomalies in EHR from insider access; Unlabeled anomalies, Challenge accurate results with low false positives	Hybrid IF-LOF + SVM/DT/RF, Six model variations, Cross-correlation features	IF-SVM: 99.21% accuracy, 99.75% sensitivity, 99.32% specificity, 98.72% F1	IF-SVM optimal, Hybrid superior, Cross-correlation reveals hidden anomalies	Single hospital validation, Scalability concerns, Institution-specific anomalies

The following details show the most vital notes that come up from literature and analyzing it in light of the goal of this paper study:

4.1 The lack of focus on insider threats compared to external attacks

The majority of studies indicate constraining on external attacks such as ransomware or intrusion that come from internet, while insider threats like misuse of the permitted access usually ignore or reduction within general framework. As illustrate it in [19] the EHR systems are depending heavily on controls metrics of privileges, which makes it not efficient in detecting some cases like Break-the-Glass, where the user has the privilege of complete access in emergency conditions, but it may produce illegitimate usage without easily detect or notice. As it shown the most of models is lacking to depth of behavioral analyzing for the users or the ability to explain their decisions in transparency way in Figure 4 presents the percentage of each attack that covered on previous studies. Which limit of its effectiveness in sensitive environment for instance healthcare sector. The numbers that are shown in the Figure 4, consider as evidence that indicates an existence of clear gaps in dealing with insider threats, which confirm the necessity of switching the research focusing on developing advance mechanisms to monitor the threats that inside the system.

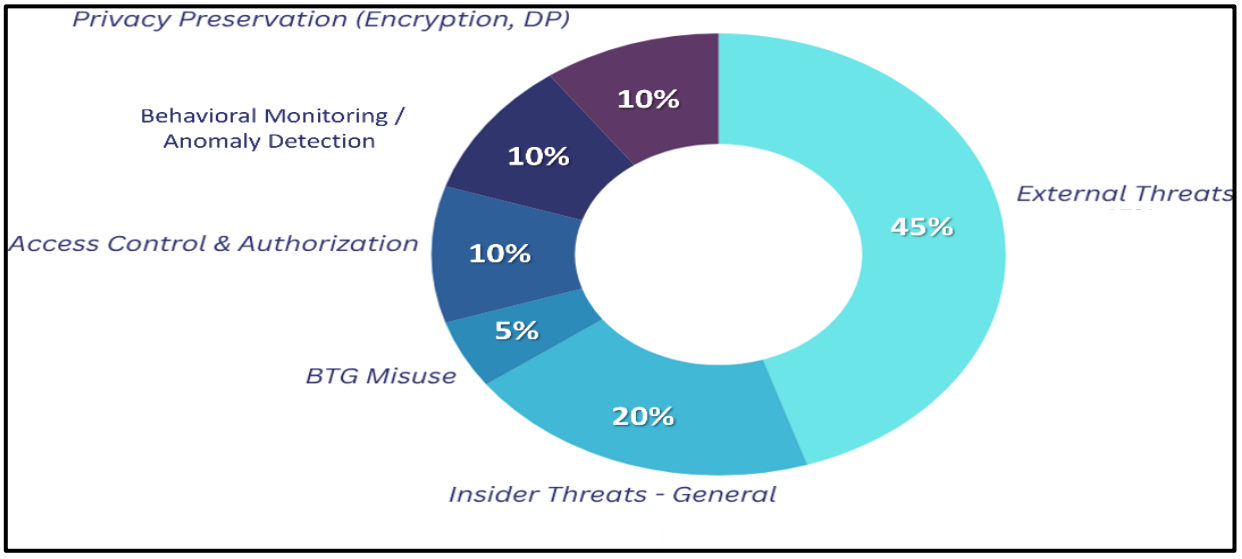

Figure 4: The distribution Focusing of studies on EHR security

4.2 The missing of user behavioral profiling

The analyzing of studies such as [14] [25] showing that majority of algorithms relaying on surface features for instance type of query and time, without building dynamic behavior model monitor the actual pattern of user. Which makes it a main weakness point, because the insider threats do not necessarily appear as compromise attempts style, rather in change of normal behavior that is make it unusual. So, from this point the need of Context-aware Behavioral Modeling (CBM) techniques that considering the usual personal pattern for each user and compare it with existing activities to identify the abnormality. This analysis indicates to the reliance on traditional access data without merging dynamic behavior models that represent a clear gap on modern secure systems, which increase the need for developing models that concentrate on intelligence user behavior.

4.3 The missing of the systematic classification of Break-the-Glass case

The difficulty of distinguishing between the legitimate and illegitimate usage of data by the permitted staff, especially for BTG or critical cases, possesses a huge methodological challenge, it brings the hardness of drawing a clear boundary between normal and suspicious behavior.

By synthesizing and studying papers such as [1] [24] it been noticed that most of models deal misuse such as Break-the-Glass as a part of general insider threats without deducting algorithms or special features for this sensitive scenario. Where this kind of usage differs in its nature due to it usually occurs in emergency cases and allows for wider privileges. The challenge is located in the distinguishes between the legitimate use and otherwise, and this is what is required to develop a model that linking among the patent's clinical case record, context of using BTG, and the time of accessing. Inserting BTG as independent threat case helps to enhance the model accuracy and increase its effectiveness in distinguishing between the legitimate use and suspicious use.

4.4 The weakness of explainability in AI models

One of the most technical constraints is that many of DL's Models work as black boxes which make it difficult to interpret its decision or justify its behavior, this weakens the trust in these systems by the health-care's employees. Inability to interpret the results has negative effects on adopting those models of environment that require transparency and organizational compliance such as healthcare sector.

Number of studies, one of them [13] [11] present models depend on DNN that characterize of strong predictive, but not able to easily explain, and this consider as an obstacle in medical environment, while it need to be the transparency a required ethically and legitimately. Explainable AI (XAI) such as SHAP and LIME offering the capability of results explain in understanding way. However, we not found from the analysis studies an actual implementation for it in detection of misuse of emergency permitted scenarios. This study observes that the explainable models represent a strategic solution among detection accuracy and user confidence, and it is one of the promising trends for future research.

4.5 The limitations of real-world dataset

Although the huge development of AI and ML techniques applications in the field of healthcare, still the deployment of these applications to detect insider threats in EHR systems remain faces a set of fundamentals challenges effect on the suggestions of the effectiveness of solutions and its suitability of apply [26-30]. One of the most challenges is the difficulty of gaining realistic data that contains an example of scenarios about the actual misuse, that are due to the ethical and organizational constraints relevant to patient's privacy [31-35]. This challenge limits the researchers' abilities of developing and enhancing trained models in sufficient way to face the realistic insider threats conditions.

As it notice in studies such as [2] [4] a lot of researches depending on synthetic data or close-source, which effect on the ability of publishing its results. Some of the new research suggests high-fidelity synthetic data but remains the accessing to actual data after hiding the identity an essential to evaluate the model’s effectiveness in realistic environment. This challenge confirms that the availability of real data after hiding the identity is a main condition to train effective models and evaluate their performance on real-world environments.

4.6 The weakness of adaptability with limited resources environment

The variety of data is showing up and their inhomogeneity as one of essential obstacles, where the electronic records differ in their format between the healthcare organizations as well as the big size and continuous flow of data enforce a pressure on ML's algorithms' performance, especially in low resources environment such as rural clinics and limit-technology of infrastructure's organization.
Papers like [8] points to the most of ML models that use are needed to advance computation environments, which makes it difficult to apply on small hospitals and not prepared very well locations. The need for lightweight models still stands that able to present a precise results using simple resources, which facilitate to publish it wider range. The studies illustrate the importance of developing lightweight models compatible with limited resources, which ensure cyber security protection on several of health sectors environments.

4.7 The variety of evaluation metrics

Some of the existing research suffering from methodological divergence and weakness of comparative results, where differentiation of the standards that use for assessment the models' performance limits the ability to unify the research effort, focusing on theoretical aspects, or developing a clear reference framework. As a large aspect of the studies focusing on theoretical aspects or perfect cases without models' effective test in realistic operational environment. The studies showing a huge variety in evaluations metrics, while some of the papers using accuracy and recall, where the other studies are tending to use only the ratio of detection without presenting a comprehensive statistics analyzing, which indicates in [20] the importance of designing a unifying framework to assess the models that include AUC, F1-score, and precision, which increase the ability of comparative and rise the reliability of results. The results highlight the need to establish a unified framework evaluation for comparison among models in scientific way and can build on for later research.

This discussion illustrates that the progress in securing healthcare records using ML is still suffering from lacks in some aspects in dealing with subtle insider threats, especially those related to BTG. To achieve a qualitative step in securing healthcare data, it is necessary to develop explainability behavioral algorithms, supported with realistic data, and the ability of implementing on limited resource environments. This research direct not only focusing on increasing the patient's protection privacy, rather it strength the trust of raped digital transformation in healthcare sector.

5. Strategic Pathways for Future Research

After comprehensive synthesis and analysis of contemporary relevant primary studies to deploy ML techniques in EHR protection, there are number of gaps and challenges that need to be addressed and precis research focusing on future. This section aims to present a detailed research map built upon criticism analyzing previous studies and realistic requirements in healthcare sector. It is depicted in Figure 5, the future direction will have such a strategy of monitoring and detecting the insider threats in the healthcare sector.

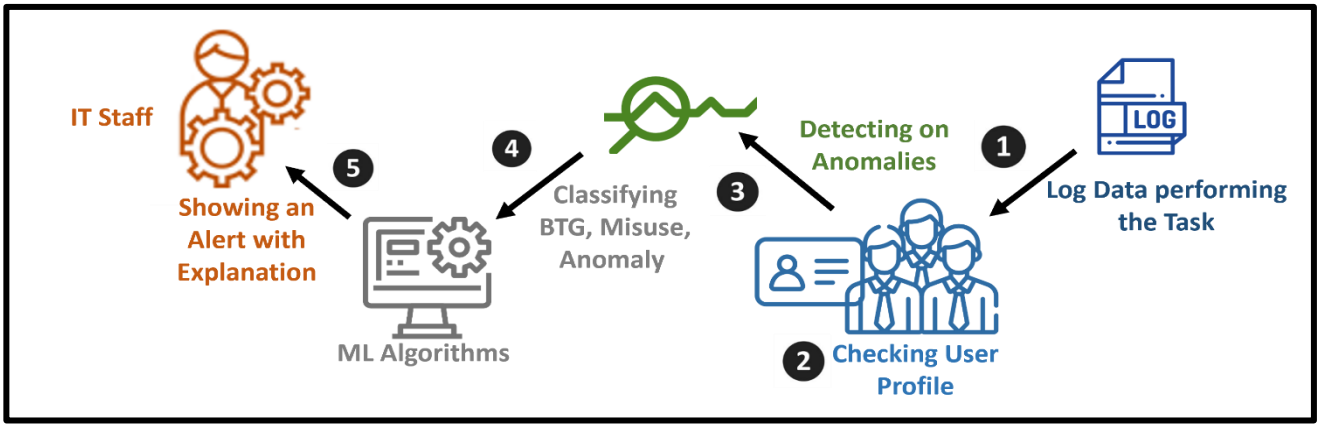

Figure 5: The Procedure of Proposed Study

5.1 increasing the transparency on AI models – Explainable AI

Despite the success that was achieved by DL models in detecting abnormal patterns and recording a high accuracy at classification techniques. However, it is often classified as a black box categorize, where the difficulty of tracking the logic behind the decisions it takes. This issue is getting sharp in context of sensitive medical, which consider understanding the decision and justifying it an ethical and organizational importance. In this context, it showing the importance of toward to XAI models, that allow the medical and management teams of tracking the reasons behind classifying a certain case as suspicious or safe, for instance marginalized between LIME and SHAP with classification models facilitating of building the confidence, support the investigation in case misuse occurred, and presenting an evidence able to approved legally. The transparency on models leads to increase the confidence and achievement of compatibility with sensitive organizational and ethical requirements in healthcare sectors.

5.2 Building smart models to analysis the dynamic behavioral for users

One of the most complicated threats on EHR systems is permitting insider users, that have legal privileges, but he uses it in unconventional ways. Most of the existing studies relaying on analyzing the records of logging-in and logging-out or type of files have been accessed, but it doesn't consider the depth of user's behavior patterns such as unexpected time repetition, alternating in type of data that is access, or moving from different or multiple portions that not align with fundamental role of the user. Therefore, developing User Behavior Modeling (UBM) has been very important issue which ensuring including those models an ability of self-learning and adaptability with changes across the time. The dynamic behavioral analysis supports the ability of system to early detection for abnormal use even with having legitimate privileges.

5.3 Distinguishing between misusing emergency privileges and other threats scenarios

Most of the studies presented insider threats in EHR systems not detailed sufficiently between the normal use of privileges and especial scenario of BTG. Privileges of BTG are designed to allow critical access in exceptional circumstances but often exposed by insider users to unjustified access to sensitive data. Therefore, it is vital that developing detecting models specialize on BTG cases, focusing on context of access, time circumstances, deployment history for the users, and the connection or the relation between the patient and the user. These models must have the ability to distinguish between legitimate use and fraud use although it technically is permitted and has the proper privileges. It shows the importance of creating specialize models to understand and explain the usage of emergency privileges within its time and behavior context.

5.4 Increasing the field experiment and depending on realistic data

Contemporary studies showing the majority ratio of it depending on synthetic data or simulation environment unrealistic, where these data considerations are useful in an initiate phase of modeling but complete relaying on it, making the model ability weak against real scenarios. Therefore, it is essential to build a partnership between researchers and healthcare organizations to offer a realistic experimental environment, and data has been preprocessing with preserving privacy (anonymized real-world data), must these environments include specifications such as number of users, variety of privileges -different in their access permissions-, and different types of emergency cases, to grantee the effectiveness of models under the real operational pressure. The movement to real experiment environments allows the researchers to have the ability to assess their models in accurate way and enhance their reliability and practical effectiveness.

5.5 Developing lightweight algorithms for healthcare organizations with limited resources

The literatures indicate to some of the advanced AI models require a high resource computation that does not exist on small clinical and rural hospitals. Therefore, consider the development of lightweight ML models inevitable trend, in condition of not effecting that on detecting accuracy. That can be done through optimizing computation performance for simple models, reducing the elements of deep network, or using effective algorithms such as Random Forest (RF) or Isolation Forest (IF) that design with an efficient way for low capabilities environments. The lightweight models represent a strategic solution to expand the electronic protection to include the limited abilities of resources on small healthcare organizations.

5.6 Merging supportive technologies – IoMT with Blockchain

While the studies largely focus on algorithms, the comprehensive security system requires margining with supportive technologies such as IoMT and blockchain. Where IoMT is offering additional data about the devices used and location and time context. Which increases the precise understanding for user behavior. As for blockchain, it provides a mechanism of not centralized checking and verification, unverifiable of manipulation, record every access operation and changing on records. The future researches are inviting to evaluate and test the effectiveness of merging those technologies within comprehensive protection framework dealing with insider threats from multiple perspectives. The merging of supportive technologies offer additional dimension in understanding the user behavior and the comprehensive protection achievement by increasing the transparency and continuous monitoring [34-38].

5.7 The need to unify the evaluation standards and test

The existing studies suffer from inconsistency on evaluation metrics, where some models use accuracy, while the others depend on recall or False Negative rate. This differentiation makes it difficult to perform subjective comparation among models or adapting the best one from them. Therefore, we are advising to develop a unify frame to evaluate the security models in context of EHR, including quantitative metrics such as AUC, precision, and F1-score, and qualitative metrics such as explainability, impacting on ongoing business, and the speed of detecting, providing results able to compare, publish, and practically application. Unified the metrics will contribute in facilitating comparing the models and improving the credibility of outcomes on research and applied community

The pervious recommendations illustrate the future of EHR security can't be only on optimizing the statistics models or algorithms, but rather in offering comprehensive and context solutions taking on account the organizing challenges, ethical, and technical in modern healthcare environments. Then through filling those gaps, an intelligent and reliable system can be built that has the ability of resisting and blocking inside misuse scenarios, and achieving digital clinical environment more secure, trust, and sustainability.

The process of building it will be through several architecture layers as illustrated in Figure 6. The architecture provides a feature for each layer that specializes in certain tasks. Two upper layer and the last layer related to operation and management tasks to execute the proposed procedure. Where the layers are in the middle for behavior of the user, utilizing appropriate algorithms, the functional and enhancement layer for alerting and explaining, and the layer responsible to interpret the text content to the person who is concerned and responsible.

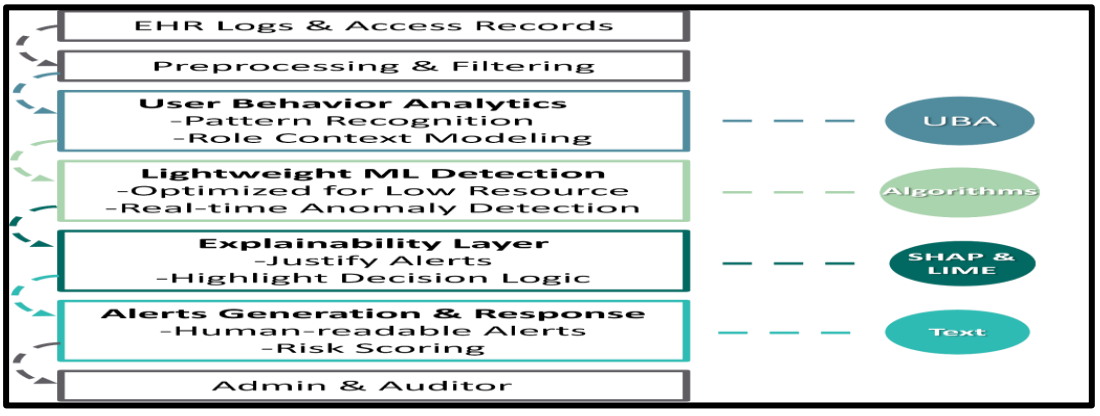

Figure 6. The process of Architecture layers.

6. Conclusion

In this study, we accomplished a comprehensive and rigorous systematic study aimed to analyzing trend directions and standing challenges in the field of AI and ML techniques to increase the security of EHR, with especially focusing on insider threats and misuse emergency privileges BTG. Moreover, we classified and analyzed the studies directly related to the specific issue of this study's topic and that based on precise systematic standards, with the aim of understanding the level of readiness of the proposed solutions in order to deal with the actual challenges in healthcare environment. The study investigated that most of research largely concentrated on external attacks and overlooked the complex of insider threats that are difficult to recognize and detect due to the legitimate privileges that the attackers have.

This paper contributes to highlighting systematic gaps and the technology that limit applying the models on realistic scenarios, and it proposed specific future directions able to be as a baseline to develop more reliable and adaptable solutions, especially through analyzing behavior. In conclusion, this study is not sufficient in contributing on collecting and analyzing the literatures, it seeks to drive the researchers and decision-makers towards developing environment more secure and aware of dealing with sensitive clinical data, through depth understanding for insider threats and creation solutions that viable to apply and deploy support the trust, transparency, and sustainability of digital healthcare system.

Despite the remarkable progress and advance development of technologies using AI to protect EHR systems. There are still clear research gaps, especially those related to the lack of real data. Also, miss of effective explainability models, not distinguishing between misuse scenarios and incompatibility of models for limited resources. Providing solutions for these gaps will be pivotal for developing more secure systems and more effective in future.

Corresponding author

Abdullah Alessa

226031653@stduent.kfu.edu.sa

Acknowledgments

The authors wish to express their gratitude to the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia. We would like to acknowledge the anonymous reviewers who made great contributions with their brilliant scholarly intuitive comments and sagacious recommendations to improve the quality and clarity of this paper.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia under the [GRANT No. KFU254709].

Author Contribution: All authors have equally contributed. All authors have read and agreed to the published version of the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used AI tools in order to assist in language refinement, proofreading, and structural suggestions during the preparation of this manuscript. After using this tool, the author reviewed and edited the content as needed and took full responsibility for the content of the published article.

Ethics declarations:

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

[1] Si-Ahmed, A., Al-Garadi, M. A., & Boustia, N. (2024). Explainable machine learning-based security and privacy protection framework for Internet of Medical Things systems. *arXiv Preprint*. https://arxiv.org/abs/2403.09752

[2] Yoon, J., Mizrahi, M., Ghalaty, N. F., et al. (2023). EHR-Safe: Generating high-fidelity and privacy-preserving synthetic electronic health records. *npj Digital Medicine*, *6*, Article 141. https://doi.org/10.1038/s41746-023-00888-7

[3] G. A., & Prasanna, S. (2024). Secure and resilient: An integrated methodology for enhancing electronic health record (EHR) data security and privacy in healthcare. In *Proceedings of the 9th International Conference on Science, Technology, Engineering and Mathematics (ICONSTEM)* (pp. 1–10). IEEE. https://doi.org/10.1109/ICONSTEM60960.2024.10568848

[4] Barnawi, A., Chhikara, P., Tekchandani, R., Kumar, N., & Alzahrani, B. (2024). A differentially privacy-assisted federated learning scheme to preserve data privacy for IoMT applications. *IEEE Transactions on Network and Service Management, 21*(4), 4686–4700.
[5] Saraswat, B. K., Saxena, A., & Vashist, P. C. (2023). Machine learning techniques for analysing security practices in electronic health records. In *Proceedings of the 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 998–1005). IEEE.
[6] Lee, J., Kim, J., Jeong, H., & Lee, K. (2025). A machine learning-based ransomware detection method for attackers' neutralization techniques using format-preserving encryption. *Sensors, 25*(8), 2406. https://doi.org/10.3390/s25082406
[7] Sharma, D., & Prabha, C. (2023). Security and privacy aspects of electronic health records: A review. In *Proceedings of the International Conference on Advancement in Computation & Computer Technologies (InCACCT)* (pp. 815–820). IEEE.
[8] Sundar, S., & Priyadharshini, A. (2024). Advanced security framework for patient monitoring systems: Integrating machine learning and encryption for enhanced data protection. In *Proceedings of the International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1–8). IEEE.
[9] Nowrozy, R., Ahmed, K., Wang, H., & McIntosh, T. (2023). Towards a universal privacy model for electronic health record systems: An ontology and machine learning approach. *Informatics, 10*(3), 60. https://doi.org/10.3390/informatics10030060
[10] Alarfaj, K. A., & Rahman, M. H. (2024). The risk assessment of the security of electronic health records using risk matrix. *Applied Sciences, 14*(13), 5785. https://doi.org/10.3390/app14135785
[11] Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine, 158*, 106848. https://doi.org/10.1016/j.compbiomed.2023.106848
[12] Naidu, U. G., Lakkshmanan, A., Krishna, J. G., Elamathi, E., & Reddy, T. S. (2025). Federated AI framework for privacy-preserving differential diagnosis across distributed medical networks. In *Proceedings of the 6th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 932–940). IEEE.
[13] Anupama, H. S., Pradeep, K. R., Niranjanamurthy, M., Kanthraju, V., Darshan, C., & Murthy, S. (2024). Adversarial learning for de-identification of medical records. In *Proceedings of the International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (Vol. 1, pp. 1–5). IEEE.
[14] Bin Sarhan, B., & Altwaijry, N. (2022). Insider threat detection using machine learning approach. *Applied Sciences, 13*(1), 259. https://doi.org/10.3390/app13010259
[15] Agrawal, A., Baniya, P., Alazzawi, E. M., Rakesh, N., Bhushan, B., & Jamil, A. (2024). Detection of DoS and DDoS attacks using machine learning and blockchain in IoMT networks. In *Proceedings of the 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)* (pp. 1–5). IEEE.
[16] Jahan, I., Mahabub, S., & Hossain, M. R. (2024). Optimizing data analysis and security of electronic health records (EHR): Role of machine learning for usability interface revolution. *Nanotechnology Perceptions*, 4011–4022.
[17] Herrera-Silva, J. A., & Hernández-Álvarez, M. (2023). Dynamic feature dataset for ransomware detection using machine learning algorithms. *Sensors, 23*(3), 1053. https://doi.org/10.3390/s23031053
[18] ElGawish, R., Abo-Rizka, M., ElGohary, R., & Hashim, M. (2022). Detecting ransomware within real healthcare medical records adopting Internet of Medical Things using machine and deep learning techniques. *International Journal of Advanced Computer Science and Applications, 13*(2).
[19] Hurst, W., Tekinerdogan, B., Alskaif, T., Boddy, A., & Shone, N. (2022). Securing electronic health records against insider threats: A supervised machine learning approach. *Smart Health, 26*, 100354. https://doi.org/10.1016/j.smhl.2022.100354
[20] Jaafar, H. S., Abed, A. A., & Al-Shareeda, M. A. (2026). A Secure Industrial Internet of Things (IIoT) Framework for Real-Time PI Control and Cloud-Integrated Industrial Monitoring. STAP Journal of Security Risk Management, 2026(1), 77–86. https://doi.org/10.63180/jsrm.thestap.2026.1.5
[21] Saraswat, B. K., Varshney, N., & Vashist, P. C. (2024). Machine learning-driven assessment and security enhancement for electronic health record systems. *International Journal of Experimental Research and Review, 43*, 160–175.
[22] Ramesh, H., Ismail, N., Abd Rahman, N. A., & Ali, A. (2026). PhishGuard: AI-Driven Graph-Based Analysis for Smarter Email Security. STAP Journal of Security Risk Management, 2026(1), 31–45. https://doi.org/10.63180/jsrm.thestap.2026.1.2
[23] Zhang, S., Guo, F., Jing, C., & Wu, C. (2024). Electronic medical record privacy protection scheme based on attribute encryption technology. In *Proceedings of the IEEE 7th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)* (Vol. 7, pp. 402–412). IEEE.
[24] Alotaibi, A., Aldawghan, H., & Rahman, M. M. H. (2026). IoT Security Concerns with Non-Fungible Tokens: A Review. STAP Journal of Security Risk Management, 2026(1), 1–30. https://doi.org/10.63180/jsrm.thestap.2026.1.1
[25] Yeng, P. K., Fauzi, M. A., Yang, B., & Yayilgan, S. Y. (2022). Analysing digital evidence towards enhancing healthcare security practice: The KID model. In *Proceedings of the 1st International Conference on AI in Cybersecurity (ICAIC)* (pp. 1–9). IEEE.

[26] Ang, S., Ho, M., Huy, S., & Janarthanan, M. (2026). A Multi-Layered Adaptive Cybersecurity Framework for the Banking Sector Integrating Next-Gen Firewalls with AI-Driven IDPS. STAP Journal of Security Risk Management, 2026(1), 67–76. https://doi.org/10.63180/jsrm.thestap.2026.1.4
[27] Addula, S. R., Norozpour, S., & Amin, M. (2025). Risk Assessment for Identifying Threats, vulnerabilities and countermeasures in Cloud Computing. Jordanian Journal of Informatics and Computing, 2025(1), 38–48. https://doi.org/10.63180/jjic.thestap.2025.1.5
[28] Al-Na'amneh, Q., Aljawarneh, M., Alhazaimeh, A. S., Hazaymih, R., & Shah, S. M. (2025). Securing Trust: Rule-Based Defense Against On/Off and Collusion Attacks in Cloud Environments. STAP Journal of Security Risk Management, 2025(1), 85–114. https://doi.org/10.63180/jsrm.thestap.2025.1.5

[29] Alghareeb, M. S., Almaiah, M., & Badr, Y. (2024). Cyber Security Threats in Wireless LAN: A Literature Review. International Journal of Cybersecurity Engineering and Innovation, 2024(1).
[30] Abu Laila, D. (2025). Responsive Machine Learning Framework and Lightweight Utensil of Prevention of Evasion Attacks in the IoT-Based IDS. STAP Journal of Security Risk Management, 2025(1), 59–70. https://doi.org/10.63180/jsrm.thestap.2025.1.3
[31] Alrajeh, M., Almaiah, M., & Mamodiya, U. (2026). Cyber Risk Analysis and Security Practices in Industrial Manufacturing: Empirical Evidence and Literature Insights. International Journal of Cybersecurity Engineering and Innovation, 2026(1).
[32] Al-Shehari, T., Al-Razgan, M., Alfakih, T., Alsowail, R. A., & Pandiaraj, S. (2023). Insider threat detection model using anomaly-based isolation forest algorithm. *IEEE Access, 11*, 118170–118185. https://doi.org/10.1109/ACCESS.2023.3325032
[33] Addula, S. R., Norozpour, S., & Amin, M. (2025). Risk Assessment for Identifying Threats, vulnerabilities and countermeasures in Cloud Computing. Jordanian Journal of Informatics and Computing, 2025(1), 38–48. https://doi.org/10.63180/jjic.thestap.2025.1.5
[34] Al-shareeda, M., & Alrudainy, H. (2026). Sustainable and Secure Energy Optimization Strategies in the Internet of Healthcare Things (IoHT). International Journal of Cybersecurity Engineering and Innovation, 2026(1).
[35] Ali, A. (2024). Adaptive and Context-Aware Authentication Framework Using Edge AI and Blockchain in Future Vehicular Networks. STAP Journal of Security Risk Management, 2024(1), 45–56. https://doi.org/10.63180/jsrm.thestap.2024.1.3
[36] Albinhamad, H., Alotibi, A., Alagnam, A., Almaiah, M., & Salloum, S. (2025). Vehicular Ad-hoc Networks (VANETs): A Key Enabler for Smart Transportation Systems and Challenges. Jordanian Journal of Informatics and Computing, 2025(1), 4–15. https://doi.org/10.63180/jjic.thestap.2025.1.2
[37] Chandak, A., & Chandak, P. (2026). Blockchain technology in health care an extensive scoping review of the existing applications, challenges, and future directions. International Journal of Cybersecurity Engineering and Innovation, 2026(1).
[38] Tabassum, M., Mahmood, S., Bukhari, A., Alshemaimri, B., Daud, A., & Khalique, F. (2024). Anomaly-based threat detection in smart health using machine learning. *BMC Medical Informatics and Decision Making, 24*(1), 347. https://doi.org/10.1186/s12911-024-02430-9

Biographies

Abdullah Alessa. Graduated with B.Sc. Degree from Valparaiso University, USA, in 2016, and He is currently pursuing a Master in Cybersecurity at King Faisal University, respectively. He is currently working as GRC Cybersecurity in Aramco, Saudi Arabia. His current research interests in cybersecurity.
226031653@student.kfu.edu.sa

Yaseen Alduwayl. Received the B.Sc. Degree from Technical Trainers College, Saudi Arabia, in 2018, and He is currently pursuing a Master in Cybersecurity at King Faisal University, respectively. He is currently an ICT instructor in Technical and Vocational Training Corporation, Saudi Arabia. His current research interest is cybersecurity.
226030848@student.kfu.edu.sa

M. M. HAFIZUR RAHMAN. Received the B.Sc. Degree in EEE from KUET, Khulna, Bangladesh, in 1996, and the M.Sc. and Ph.D. degrees in information science from JAIST, Japan, in 2003 and 2006, respectively. He is currently an Associate Professor with the Department of CN, CCSIT, KFU, and Saudi Arabia. Prior to joining KFU, he was an Assistant Professor with Xiamen University Malaysia, and IIUM, Malaysia; and an Associate Professor with the Department of CSE, KUET. He was also a Visiting Researcher with the School of Information Science, JAIST, in 2008; and a JSPS Postdoctoral Research Fellow with the Graduate School of Information Science (GSIS), Tohoku University, Japan, in 2009, and the Center for Information Science, JAIST, from 2010 to 2011. His current research interests include hierarchical interconnection networks, optical switching networks, and cybersecurity.
mhrahman@kfu.edu.sa