EJPMR, 2025, 12(12), 515-547 **Research Article** **Impact Factor: 8.158**

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

ISSN (O): 2394-3211
ISSN (P): 3051-2573
Coden USA: EJPMAG

MACHINE LEARNING–BASED ANOMALY DETECTION FOR EARLY IDENTIFICATION OF EMR BREACH PATHWAYS

Genevieve Donkor Armah*[1], Idoko Peter Idoko[2], Yewande Iyimide Adeyeye[3], Lawrence Anebi Enyejo[4], Azonuche Tony Isioma[5]

[1]Department of Economics, Youngstown State University, Youngstown Ohio, USA.
[2]Department of Electrical/ Electronic Engineering, University of Ibadan, Nigeria.
[3]Department of Day Case Surgery, Dumfries and Galloway Royal Infirmary, Dumfries, United Kingdom.
[4]Department of Telecommunications, Enforcement Ancillary and Maintenance, National Broadcasting Commission Headquarters Aso-Villa, Abuja, Nigeria.
[5]Department of Project Management, Amberton University, Garland Texas, USA.

***Corresponding Author: Genevieve Donkor Armah**

Department of Economics, Youngstown State University, Youngstown Ohio, USA.

DOI: https://doi.org/10.5281/zenodo.18712847

How to cite this Article: Genevieve Donkor Armah*[1], Idoko Peter Idoko[2], Yewande Iyimide Adeyeye[3], Lawrence Anebi Enyejo[4], Azonuche Tony Isioma[5]. (2025). Machine Learning-Based Anomaly Detection For Early Identification Of Emr Breach Pathways. European Journal of Biomedical and Pharmaceutical Sciences, 12(12), 515–547.

Article Received on 21/10/2025 **Article Revised on 11/11/2025** **Article Published on 01/12/2025**

ABSTRACT

Electronic Medical Records (EMRs) are central to modern healthcare delivery, yet their growing accessibility and integration have expanded the attack surface for data breaches that often unfold gradually and evade traditional security controls. Conventional rule-based and signature-driven intrusion detection systems are largely reactive and ill-suited to identifying early-stage breach behaviors embedded within legitimate clinical workflows. This study proposes a machine learning–based anomaly detection framework for the early identification of EMR breach pathways, treating breaches as sequential processes rather than isolated events. The framework integrates EMR audit logs, role and contextual metadata, and temporal modeling to detect subtle deviations in access behavior. A comprehensive evaluation is conducted using baseline statistical methods, classical machine learning models, and deep learning approaches, including autoencoders and LSTM-based architectures. Experimental results demonstrate that deep and hybrid models significantly outperform traditional approaches in detection accuracy, time-to-detection, and robustness to noise and behavioral drift. Importantly, the study incorporates an interpretability layer that maps anomalies into coherent breach pathways, enhancing usability for security analysts and compliance officers. The findings highlight the effectiveness of sequence-aware, explainable machine learning in enabling proactive EMR security monitoring, improving audit readiness, and supporting timely intervention before breach escalation.

KEYWORDS: Electronic Medical Records (EMR); Anomaly Detection; Healthcare Cybersecurity; Machine Learning; Deep Learning; LSTM Autoencoder.

1.1 Background and Context

The digital transformation of healthcare systems has positioned Electronic Medical Records (EMRs) as the backbone of clinical decision-making, care coordination, billing, and population health management. EMRs consolidate sensitive patient information, including diagnostic histories, laboratory results, prescriptions, and clinician notes, enabling continuity of care and data-driven clinical workflows across distributed healthcare environments. Their widespread adoption has been accelerated by policy mandates, interoperability standards, and the push toward value-based care, making EMR platforms deeply embedded in daily hospital operations and clinical routines (Adler-Milstein & Huckman, 2013; Kruse et al., 2018; Ijiga et al., 2025).

This expanded reliance has simultaneously increased the attack surface of healthcare information systems. EMR breaches have grown not only in frequency but also in technical sophistication. Contemporary incidents

increasingly involve insider threats, where legitimate users abuse authorized access; credential compromise through phishing or malware; and lateral movement attacks, in which adversaries quietly traverse internal systems over extended periods before exfiltrating data. Reports from healthcare cybersecurity studies show that many breaches persist undetected for months, allowing attackers to exploit normal-looking access behaviors that evade static controls (McLeod & Dolezel, 2018; Jalali & Kaiser, 2018; Okpanachi et al., 2025). Unlike abrupt system intrusions, EMR breaches often unfold incrementally, blending into legitimate clinical workflows.

Traditional security mechanisms deployed in healthcare environments remain largely perimeter-based and rule-driven. Firewalls, role-based access controls, and signature-based intrusion detection systems are effective against known threats but struggle to detect subtle deviations in user behavior or novel attack patterns. Rule-based systems assume stable access norms and predefined misuse scenarios, assumptions that rarely hold in dynamic clinical contexts characterized by shift work, emergency access, and cross-functional collaboration. As a result, early-stage breach behaviors such as abnormal access sequences, atypical patient record traversal, or gradual privilege escalation frequently go unnoticed until regulatory audits or data leakage events occur (Appari & Johnson, 2010; Behl & Behl, 2017; George et al., 2025).

These limitations have motivated a growing interest in data-driven and machine learning–based security approaches capable of modeling complex, evolving EMR access behaviors. By learning normal patterns from high-dimensional audit logs and temporal access sequences, anomaly detection models offer the potential to identify emerging breach pathways at an early stage, before significant harm or regulatory exposure occurs. This shift from reactive, rule-centric security toward adaptive behavioral analytics represents a critical evolution in protecting modern EMR ecosystems.

Figure 1 illustrates a simplified three-tier architecture for cloud-based healthcare data management. Tier 1 represents data generation, where patient physiological data are collected through wearable sensors and coordinated via a mobile device. Tier 2 handles data transmission using wireless communication technologies such as Wi-Fi, GPS, and Wi-Max. Tier 3 focuses on data storage and access, where cloud infrastructure enables secure availability of patient data to healthcare stakeholders, including clinicians, hospitals, and emergency services.

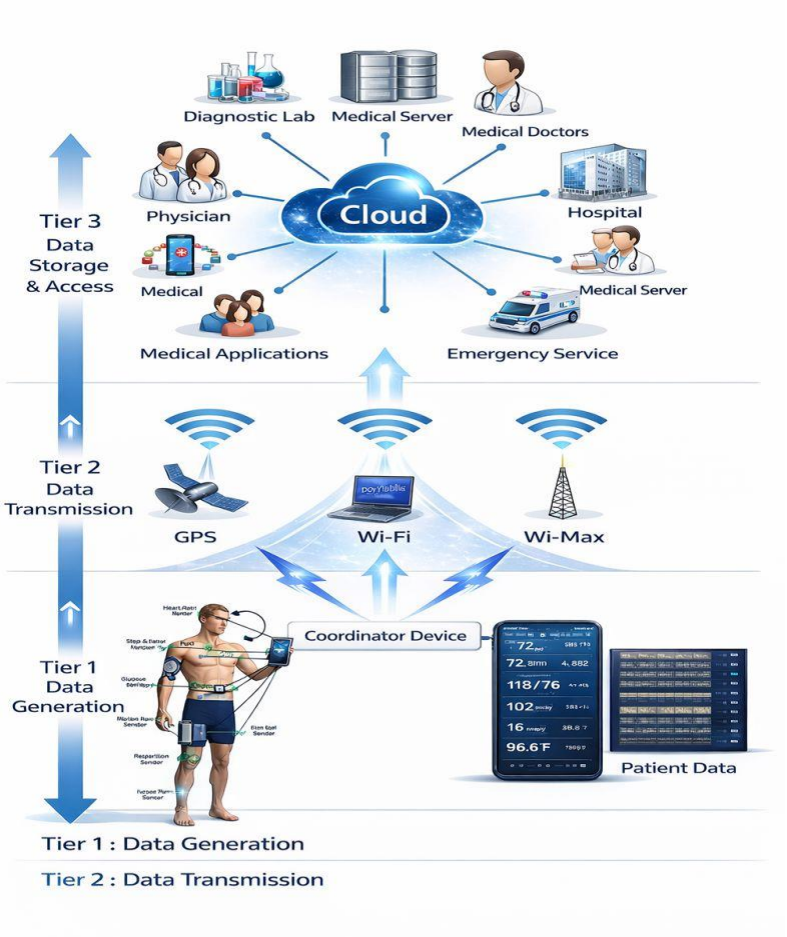

Figure 1: Simplified Three-Tier Architecture for Cloud-Enabled Healthcare Data Flow.

1.2 Problem Statement

Healthcare organizations increasingly depend on EMR platforms for clinical operations, but the same interconnectedness that enables timely access to patient data also enables rapid compromise and stealthy misuse. Empirical analyses of healthcare breaches show persistent growth in breach incidence and exposed records, with common drivers including hacking/IT incidents and unauthorized internal disclosures. These patterns underscore that both external compromise and legitimate-user misuse remain central breach pathways in healthcare environments (Seh et al., 2020; McLeod & Dolezel, 2018; Uzoma et al., 2025; Okpanachi et al., 2025). Yet, many hospital security stacks still rely heavily on conventional intrusion detection and prevention paradigms that were designed for comparatively stable enterprise networks rather than workflow-intensive clinical systems (Jalali & Kaiser, 2018; George et al., 2024; Gaye et al., 2025).

A core deficiency is the inadequacy of traditional intrusion detection and signature-based tools for early breach pathway identification. Signature-based systems and many rule-driven alerting approaches are effective against known indicators of compromise but systematically underperform when adversaries adapt tactics, reuse legitimate credentials, or operate "low-and-slow." In EMR contexts, attackers can emulate normal operational patterns such as routine chart access, team-based care coordination, or after-hours activity by on-call staff, which reduces the discriminative power of static rules and fixed signatures. This is especially problematic for insider-threat and credential-compromise scenarios where the activity originates from authenticated accounts, making perimeter-focused controls insufficient for detecting misuse that occurs inside trusted boundaries (Hurst et al., 2022; Al-Mhiqani et al., 2020; Avevor et al., 2025). Hospital-focused cybersecurity research also emphasizes that organizational realities (legacy systems, uneven patching, and constrained security staffing) amplify the limits of classic monitoring approaches that require constant manual rule maintenance (Jalali & Kaiser, 2018; Uzoma et al., 2024; Avevor et al., 2024).

A second challenge arises from the high-dimensional and heterogeneous nature of EMR access logs and clinical workflow data. EMR audit trails encode multi-actor, temporally structured interactions (user role, workstation/location, patient identity, access type, timing, session properties), often across multiple integrated subsystems (lab, radiology, billing, identity management). These data are noisy, sparse, and context-dependent because clinical workflows legitimately vary by department, shift, and acuity level. Consequently, breach signals tend to manifest as subtle distribution shifts or anomalous sequences rather than isolated "red-flag" events. This heterogeneity complicates feature construction, baseline definition of "normal," and model generalization across units and institutions, all while labels for confirmed misuse remain scarce and delayed (Hurst et al., 2022; Tabassum et al., 2024; Imoh et al., 2024; Uzoma et al., 2024). Broader healthcare security reviews further note that healthcare IT ecosystems are complex and continuously evolving, which increases variance in telemetry and raises the risk of both false positives and missed detections when simplistic assumptions are used (Vilakazi & Adebesin, 2023; Mehrtak et al., 2021; Azonuche et al., 2025; Ogbuonyalu et al., 2025).

These limitations create a clear operational and scientific need for adaptive, data-driven methods that can identify anomalous access behaviors before breaches escalate into large-scale exfiltration, operational disruption, or regulatory exposure. Machine learning–based anomaly detection is a natural candidate because it can learn behavioral baselines from historical audit logs, model temporal patterns of access, and surface deviations that are not captured by static signatures. Importantly, the problem is not merely to "detect anomalies," but to detect anomalies early enough to reconstruct breach pathways (e.g., credential takeover → unusual patient traversal → privilege expansion → bulk export) so that security teams can intervene while the attacker's dwell time is still limited (Tabassum et al., 2024; Al-Mhiqani et al., 2020; Azonuche & Enyejo 2024). Therefore, the research problem centers on designing and validating ML-based anomaly detection approaches that are robust to workflow variability, scalable to high-volume EMR logs, and actionable for healthcare security operations.

1.3 Research Objectives

The overarching objective of this study is to advance proactive cybersecurity monitoring in healthcare environments by developing and validating a machine learning–driven approach for the early identification of EMR breach pathways. Rather than focusing on post-incident detection, the study emphasizes early-stage behavioral deviations that signal emerging compromise within trusted clinical systems.

First, the study aims to design a machine learning–based anomaly detection framework tailored to EMR ecosystems. The framework is intended to ingest high-volume EMR audit logs and contextual workflow data, learn baseline access behaviors across users and roles, and surface deviations that may indicate the initial phases of breach activity. Particular emphasis is placed on modeling temporal access sequences and user–patient interaction patterns to enable reconstruction of likely breach pathways rather than isolated alerts.

Second, the study seeks to evaluate the effectiveness of multiple anomaly detection paradigms within EMR environments. This includes a comparative assessment of statistical techniques, classical machine learning models, and deep learning–based approaches. The evaluation focuses on detection accuracy, robustness to workflow variability, sensitivity to low-frequency malicious behaviors, and scalability under realistic EMR data

volumes. By benchmarking these paradigms under consistent experimental conditions, the study aims to clarify their relative strengths and limitations for healthcare-specific security monitoring.

Third, the research aims to analyze detected anomalies in terms of interpretability and actionable security insights. Beyond numerical detection performance, the study examines whether model outputs can be meaningfully interpreted by healthcare security and compliance teams. This includes assessing the ability to explain why an access pattern is anomalous, how it deviates from expected clinical behavior, and how such insights can support timely investigation, risk prioritization, and incident response.

Collectively, these objectives position the study to contribute both a technical framework and practical guidance for deploying machine learning–based anomaly detection as an operational tool for early EMR breach prevention and response.

1.4 Research Questions

This study is guided by a set of research questions that directly reflect the technical, operational, and interpretive challenges of securing Electronic Medical Record (EMR) systems through machine learning–based anomaly detection.

The first research question examines which machine learning techniques are most effective for the early detection of EMR breach pathways. This question focuses on comparing statistical methods, classical machine learning models, and deep learning approaches in their ability to identify subtle deviations in access behavior that precede confirmed breaches. Emphasis is placed on early-stage detection rather than post-compromise recognition, with effectiveness measured in terms of detection accuracy, timeliness, and stability across diverse clinical workflows.

The second research question investigates how anomaly detection models perform under class imbalance and concept drift typical of healthcare systems. EMR environments are characterized by an overwhelming prevalence of legitimate access events and a scarcity of labeled breach instances, as well as evolving usage patterns driven by staffing changes, policy updates, and clinical emergencies. This question evaluates model resilience to skewed class distributions and shifting behavioral baselines, assessing whether detection performance degrades over time and how adaptive learning strategies mitigate these effects.

The third research question explores whether detected anomalies can be mapped to interpretable breach pathways for security and compliance teams. Rather than treating anomalies as isolated alerts, this question assesses the feasibility of linking anomalous events into coherent access sequences that reflect plausible breach progression. The focus is on interpretability and operational relevance, determining whether model outputs can support investigation, regulatory reporting, and timely intervention by non-technical stakeholders.

Together, these research questions establish a structured inquiry into effectiveness, robustness, and interpretability three critical dimensions required for translating machine learning–based anomaly detection from experimental models into practical EMR security solutions.

1.5 Contributions of the Study

This study makes several substantive contributions to the intersection of healthcare cybersecurity and machine learning, with a specific focus on proactive protection of Electronic Medical Record (EMR) systems.

First, the study proposes a structured machine learning–based framework for EMR anomaly detection that is explicitly designed for early breach identification. The framework integrates EMR audit logs, contextual access metadata, and temporal behavioral modeling to move beyond isolated alert generation. By emphasizing behavioral baselining and sequence-level analysis, the framework supports systematic detection of emerging misuse patterns within trusted clinical environments.

Second, the study provides an empirical comparison of multiple anomaly detection models applied to real or realistically simulated EMR access data. Statistical approaches, classical machine learning techniques, and deep learning models are evaluated under consistent experimental conditions. This comparative analysis offers evidence-based insights into the relative strengths, limitations, and operational trade-offs of each paradigm when deployed in high-dimensional, imbalanced, and workflow-driven healthcare settings.

Third, the study introduces a breach-pathway interpretation layer that links detected anomalies to practical security response. Rather than treating anomalies as standalone signals, the proposed approach aggregates and contextualizes anomalous events into interpretable access sequences that reflect plausible breach progression. This enables security and compliance teams to understand how anomalous behavior unfolds over time, prioritize investigations, and initiate timely containment or remediation actions.

Collectively, these contributions advance both methodological rigor and practical applicability by bridging machine learning innovation with the operational realities of EMR security, supporting earlier detection, clearer interpretation, and more effective response to emerging breach threats.

2. LITERATURE REVIEW

2.1 EMR Security and Breach Taxonomies

Electronic Medical Record (EMR) systems are exposed to a distinct set of security threats shaped by their role as

centralized repositories of highly sensitive health data and their deep integration into clinical workflows. Empirical studies and regulatory breach reports consistently identify insider misuse, privilege escalation, credential theft, and data exfiltration as the dominant EMR breach vectors, often occurring in combination rather than isolation (McLeod & Dolezel, 2018; Seh et al., 2020; Abiodun et al., 2023; Okereke et al., 2023).

Insider misuse remains one of the most persistent and difficult-to-detect EMR threats. This category includes both malicious insiders who intentionally access patient records without authorization and negligent insiders who violate access policies out of curiosity or convenience. Because insiders operate using legitimate credentials and within trusted network boundaries, their activities frequently resemble normal clinical behavior, reducing the effectiveness of traditional rule-based monitoring systems (Appari & Johnson, 2010; Azonuche & Enyejo 2024). Studies show that insider-related incidents account for a substantial proportion of healthcare breaches and are often discovered only during audits or after patient complaints (McLeod & Dolezel, 2018; Azonuche & Enyejo 2024).

Privilege escalation represents a second critical breach pathway, where attackers exploit misconfigurations, weak role-based access controls, or compromised administrator accounts to gain elevated permissions. In EMR environments, privilege escalation can enable broad patient record access across departments or facilities, significantly amplifying breach impact. Research highlights that legacy EMR deployments and complex role hierarchies increase the likelihood of excessive privileges being granted or retained beyond clinical necessity (Behl & Behl, 2017; Jalali & Kaiser, 2018; Azonuche & Enyejo 2025; Gaye et al., 2024).

Credential theft, often initiated through phishing, malware, or password reuse across systems, is a leading external attack vector in healthcare. Once credentials are compromised, attackers can authenticate directly into EMR systems and operate undetected for extended periods by mimicking legitimate user behavior. Healthcare-specific analyses indicate that credential-based intrusions frequently precede large-scale breaches and serve as entry points for subsequent lateral movement within hospital networks (Seh et al., 2020; Kruse et al., 2017).

The final stage in many breach taxonomies is data exfiltration, where sensitive patient information is extracted for financial gain, identity theft, or extortion. Exfiltration may occur gradually through repeated low-volume queries or abruptly via bulk exports, depending on attacker objectives and detection pressure. Because EMR systems are designed to support high-volume data access for legitimate clinical purposes, distinguishing malicious exfiltration from normal operations poses a significant technical challenge (Appari & Johnson, 2010; McLeod & Dolezel, 2018).

Beyond technical risks, EMR breaches carry substantial regulatory and compliance implications. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) mandates safeguards for protected health information and requires covered entities to maintain audit controls capable of recording and examining system activity. In the European Union, the General Data Protection Regulation (GDPR) imposes strict requirements for data minimization, access accountability, and breach notification within defined timeframes. Similarly, Nigeria's Nigeria Data Protection Regulation (NDPR) establishes obligations for lawful processing, access control, and demonstrable compliance through audit trails. Across these regimes, inadequate monitoring of EMR access and delayed breach detection can result in significant financial penalties, reputational damage, and loss of public trust (Appari & Johnson, 2010; Jalali & Kaiser, 2018).

These regulatory expectations elevate audit logging and continuous monitoring from operational best practices to legal necessities. Consequently, EMR security taxonomies increasingly emphasize not only breach occurrence but also breach progression, highlighting the need to identify early-stage behaviors that signal misuse before regulatory thresholds for reportable incidents are crossed.

2.2 Traditional Intrusion Detection in Healthcare Systems

Traditional intrusion detection mechanisms remain widely deployed in healthcare environments, largely due to regulatory pressure, legacy system compatibility, and the perceived reliability of deterministic controls. These mechanisms are primarily centered on rule-based access control monitoring and signature-based intrusion detection systems (IDS). While they provide a foundational layer of protection, extensive research shows that their effectiveness is limited in complex, workflow-driven EMR ecosystems.

Rule-based access control monitoring is typically implemented through role-based access control (RBAC) policies, audit rules, and threshold-based alerts. In EMR systems, access rules define which users or roles may view, modify, or export specific patient records based on clinical responsibility. Monitoring systems flag violations when predefined rules are broken, such as access outside assigned roles, excessive record views, or access during restricted hours. These mechanisms are valuable for enforcing compliance and supporting retrospective audits, particularly under regulatory frameworks that mandate traceable access logs (Appari & Johnson, 2010).

Figure 2 presents a simplified and professional representation of a centralized EMR workflow within a

healthcare environment. The diagram illustrates how patient information is generated during consultation and treatment, then created, stored, and managed by administrative systems on a centralized EMR server. Authorized clinicians retrieve and update EMR data in real time to support medical decision-making. This architecture highlights secure, role-based access and continuous data exchange as the foundation for efficient and coordinated patient care.

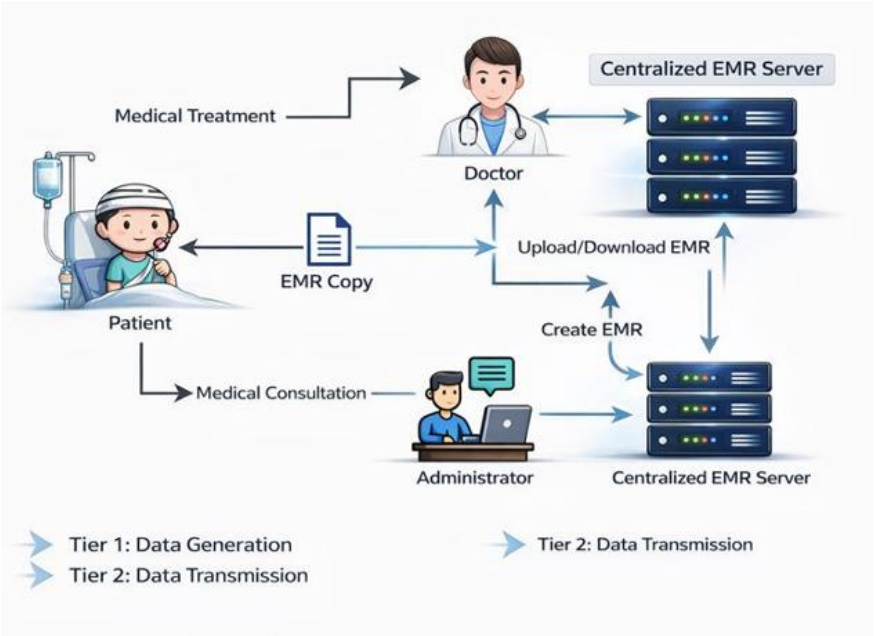

Figure 2: Centralized Electronic Medical Record (EMR) Workflow for Clinical Data Management.

However, rule-based monitoring assumes that misuse manifests as explicit policy violations. In practice, many EMR breaches occur within the boundaries of authorized access. Clinicians, billing staff, or administrators may technically comply with RBAC rules while still engaging in inappropriate record access, especially in insider misuse or credential compromise scenarios. Healthcare studies emphasize that clinical workflows are inherently variable, involving emergency overrides, cross-department collaboration, and irregular access patterns, which forces rules to be permissive and limits their sensitivity to subtle misuse (Jalali & Kaiser, 2018; Kruse et al., 2017).

Signature-based intrusion detection systems represent another common security layer in healthcare networks. These systems detect malicious activity by matching observed events against known attack signatures, such as malware indicators, exploit patterns, or network-based intrusion templates. Signature-based IDS are effective against well-characterized threats and remain useful for detecting commodity malware and known exploits targeting hospital infrastructure (Behl & Behl, 2017).

Despite these strengths, signature-based IDS face significant limitations in healthcare contexts. First, they are inherently reactive, detecting only attacks that have been previously observed and encoded into signatures. Second, EMR breaches increasingly rely on credential theft and insider activity rather than overt exploits, allowing attackers to bypass signature detection entirely by operating through legitimate application interfaces. Third, healthcare IT environments often involve legacy systems, proprietary EMR software, and encrypted traffic, reducing the visibility and coverage of network-based signatures (Seh et al., 2020; McLeod & Dolezel, 2018).

Collectively, these limitations reduce the ability of traditional intrusion detection approaches to identify early-stage breach behaviors. Rule-based and signature-driven systems tend to generate alerts only after policy violations or known attack patterns emerge, which often corresponds to later phases of breach progression. As a result, attackers may persist undetected for extended periods, increasing dwell time and breach severity. This gap has driven growing interest in anomaly-based and machine learning–driven detection approaches that can model normal EMR access behavior and detect deviations without relying on predefined rules or signatures.

Figure 3 illustrates a multi-layer role-based access control framework that governs how diverse users interact with sensitive healthcare data. Users are first mapped to predefined roles such as patient, doctor, researcher, or emergency personnel, enabling coarse-grained access filtering at the system boundary. Access requests are then evaluated within a centralized server repository that enforces policy rules, security constraints, and contextual conditions before data exposure. The decision engine ultimately classifies each request as

authorized or non-authorized, ensuring confidentiality, accountability, and controlled data sharing across the healthcare ecosystem.

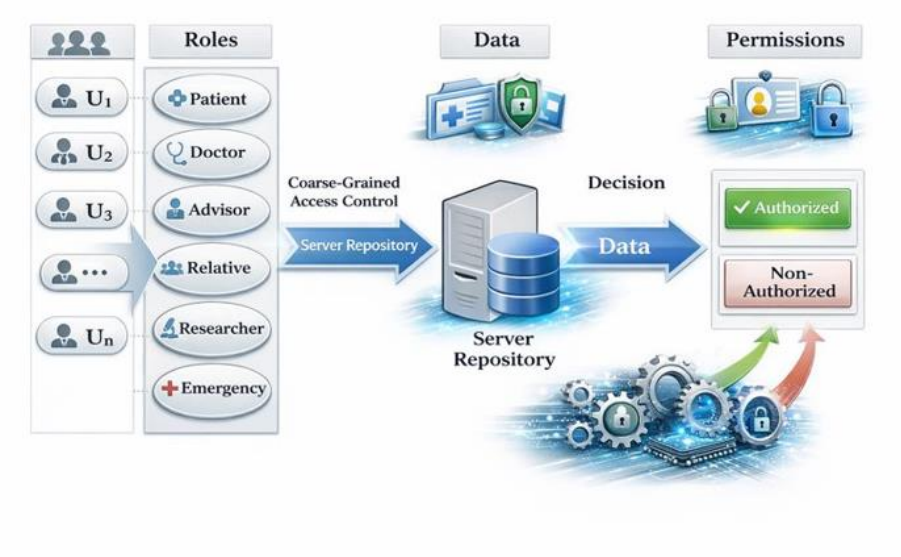

Figure 3: Multi-Layer Role-Based Access Control (RBAC) Architecture for Secure Healthcare Data Management.

2.3 Machine Learning for Anomaly Detection

Machine learning (ML) has emerged as a central paradigm for anomaly detection in complex, data-intensive environments such as Electronic Medical Record (EMR) systems. Unlike rule-based or signature-driven mechanisms, ML-based approaches aim to learn patterns of normal behavior directly from data and identify deviations that may indicate misuse, compromise, or emerging breach activity. In healthcare settings, this capability is particularly valuable because malicious actions often manifest as subtle behavioral shifts rather than explicit policy violations (Chandola et al., 2009; Hurst et al., 2022).

Statistical and distance-based methods represent some of the earliest and most widely adopted approaches to anomaly detection. Techniques such as z-score analysis and Gaussian-based models assume that normal behavior follows a stable statistical distribution and flag events that deviate beyond predefined thresholds. In EMR audit logs, these methods are commonly applied to features such as access frequency, session duration, or patient record counts. While simple and computationally efficient, their effectiveness is limited when access behavior is multimodal or non-stationary, which is typical in clinical workflows involving shifts, emergencies, and role-based variation (Chandola et al., 2009).

Figure 4 presents a real-time physiological monitoring and anomaly-detection workflow designed to improve the reliability of medical alerts. Sensor readings collected at time t are predicted and compared against expected values, with deviations evaluated using a predefined error threshold. Binary decisions are aggregated through a majority-voting mechanism to distinguish true medical conditions from sensor-induced anomalies. Validated events trigger alarms and database updates, while false alarms are detected and filtered to enhance system accuracy and clinical trust.

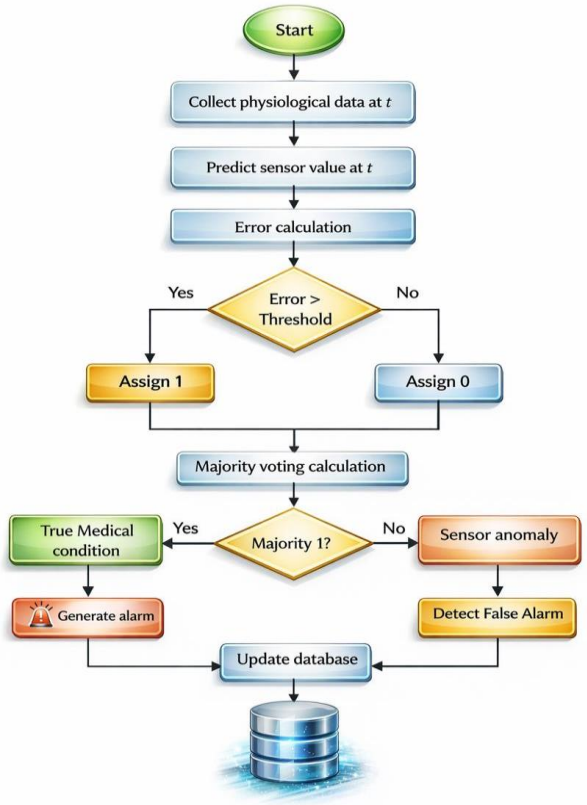

Figure 4: Majority-Voting-Based Physiological Signal Validation and Alarm Generation Workflow.

Distance-based techniques, including k-nearest neighbors (k-NN) and local outlier factor (LOF), identify anomalies by measuring how isolated a data point is relative to its neighbors in feature space. These methods can capture local irregularities in EMR access patterns, such as unusual combinations of user role and patient access. However, their performance degrades in high-dimensional settings, and they are sensitive to feature scaling and noise, both of which are common in heterogeneous EMR logs (Aggarwal, 2017). Isolation-based methods, such as Isolation Forest, address some of these limitations by explicitly modeling how easily an observation can be separated from the rest of the data. These techniques have shown practical effectiveness in cybersecurity applications because they do not rely on distance metrics and scale well to large datasets (Liu et al., 2008; Al-Mhiqani et al., 2020).

ML-based anomaly detection approaches can be broadly categorized into supervised, unsupervised, and semi-supervised paradigms, each with distinct implications for healthcare security. Supervised methods rely on labeled examples of normal and malicious behavior to train classifiers. While these models can achieve high accuracy when sufficient labels are available, healthcare environments rarely provide comprehensive or timely labels for EMR misuse due to privacy concerns, delayed investigations, and underreporting. As a result, supervised approaches are often impractical for early-stage breach detection in real-world EMR systems (Hurst et al., 2022).

Unsupervised methods, by contrast, learn normal behavior patterns without requiring labeled anomalies and flag deviations as potential threats. This paradigm aligns well with EMR security, where malicious events are rare and evolving. Unsupervised models are particularly suited to detecting novel or previously unseen breach behaviors, such as new insider misuse strategies or stealthy credential abuse. However, they often suffer from higher false-positive rates, as not all deviations from learned baselines are malicious in dynamic clinical contexts (Chandola et al., 2009; Tabassum et al., 2024).

Semi-supervised approaches attempt to balance these trade-offs by training models primarily on normal behavior while incorporating limited labeled anomalies or expert feedback. In healthcare security, this paradigm is increasingly viewed as a practical compromise, enabling models to remain adaptive while grounding detection in known misuse patterns. Semi-supervised methods also facilitate incremental learning as new breach cases are identified, supporting continuous improvement without requiring exhaustive labeling (Aggarwal, 2017).

Despite their promise, ML-based anomaly detection methods face notable strengths and weaknesses in high-risk, low-label domains such as healthcare. Their primary strength lies in adaptability: ML models can capture complex, nonlinear relationships in high-dimensional EMR data and evolve as access patterns change. At the same time, weaknesses include sensitivity to data quality, difficulty in interpreting anomaly scores, and challenges in distinguishing malicious behavior from legitimate but rare clinical activities. These limitations underscore the need for carefully designed feature engineering, model evaluation strategies that emphasize early detection, and interpretability mechanisms that translate anomalies into actionable security insights (Tabassum et al., 2024; Hurst et al., 2022).

2.4 Deep Learning and Sequential Modeling

Deep learning based approaches have gained increasing prominence in anomaly detection tasks where data exhibit complex temporal dependencies and high-dimensional structure, characteristics that are intrinsic to EMR access logs. Unlike traditional machine learning methods that rely on handcrafted features and static representations, deep learning models are capable of learning hierarchical and sequential patterns directly from raw or minimally processed log data. This capability is particularly relevant for EMR security, where breach behaviors often unfold over time as coordinated sequences of access events rather than isolated anomalies (Chandola et al., 2009; Hurst et al., 2022).

Autoencoders are among the most widely used deep learning architectures for anomaly detection in log and audit data. Autoencoders learn a compressed representation of normal behavior by minimizing reconstruction error between input data and its reconstructed output. In EMR contexts, they are commonly applied to user–patient access matrices, session-level features, or aggregated behavioral vectors. Events or sequences that yield high reconstruction error are flagged as anomalous. The primary advantage of autoencoders lies in their ability to model nonlinear relationships in high-dimensional data without requiring labeled attack samples. However, standard autoencoders treat observations independently, which limits their effectiveness in capturing temporal dependencies inherent in access workflows (Aggarwal, 2017; Chalapathy & Chawla, 2019).

To address this limitation, LSTM-based models have been extensively adopted for sequential anomaly detection. Long Short-Term Memory (LSTM) networks are designed to model long-range temporal dependencies by maintaining internal memory states that capture past context. In EMR systems, LSTM models can learn typical access sequences associated with specific clinical roles, departments, or shifts, enabling detection of anomalous event orderings, timing irregularities, or abnormal transitions between patients and functions. Research in healthcare cybersecurity demonstrates that LSTM-based approaches outperform static models when breaches involve gradual behavioral drift, credential

misuse, or staged privilege escalation (Hurst et al., 2022; Malhotra et al., 2016). Their ability to operate on sequences makes them well suited for early breach pathway identification, where deviations may only become apparent across multiple correlated events.

More recently, attention mechanisms and Transformer-based architectures have been introduced to improve both performance and interpretability in log and sequence analysis. Attention mechanisms allow models to weigh the relative importance of different events within a sequence, rather than relying solely on fixed-length memory representations. In EMR access analysis, this enables the model to focus on critical actions such as access to high-sensitivity records, unusual patient transitions, or rare administrative functions while down-weighting routine behavior. Attention-based models have shown strong results in system log anomaly detection and are increasingly explored in healthcare settings because they provide clearer insight into which access events contribute most to an anomaly score (Chalapathy & Chawla, 2019; Tabassum et al., 2024).

A central challenge addressed by deep sequential models is handling temporal dependencies in EMR access behavior. Clinical workflows are inherently time-dependent, shaped by shift schedules, emergency contexts, and care pathways that span hours or days. Breach activities often exploit this temporal structure, embedding malicious actions within otherwise legitimate sequences to avoid detection. Deep learning models that explicitly encode time through sequence ordering, time gaps, or contextual embeddings are better equipped to distinguish between benign variability and meaningful deviations. Nevertheless, these models introduce additional challenges, including higher computational cost, sensitivity to training data quality, and the need for careful validation to prevent overfitting to transient workflow patterns (Malhotra et al., 2016; Aggarwal, 2017).

Deep learning and sequential modeling provide a powerful foundation for EMR anomaly detection by enabling early identification of breach pathways that emerge over time. When combined with interpretability mechanisms and operational constraints, these approaches offer a viable path toward proactive, behavior-aware security monitoring in healthcare systems.

Figure 5 illustrates a realistic sequence diagram showing how sensor nodes and an anchor node exchange packets while performing anomaly checks in a wireless sensor network. Broadcast packets are validated locally at each sensor node, and anomalous packets are selectively ignored to prevent propagation of compromised data. Valid sensed data, enriched with location information, are forwarded to the anchor node for processing and coordination. Alert packets are subsequently disseminated to notify neighboring nodes, enabling early anomaly awareness and resilient network operation.

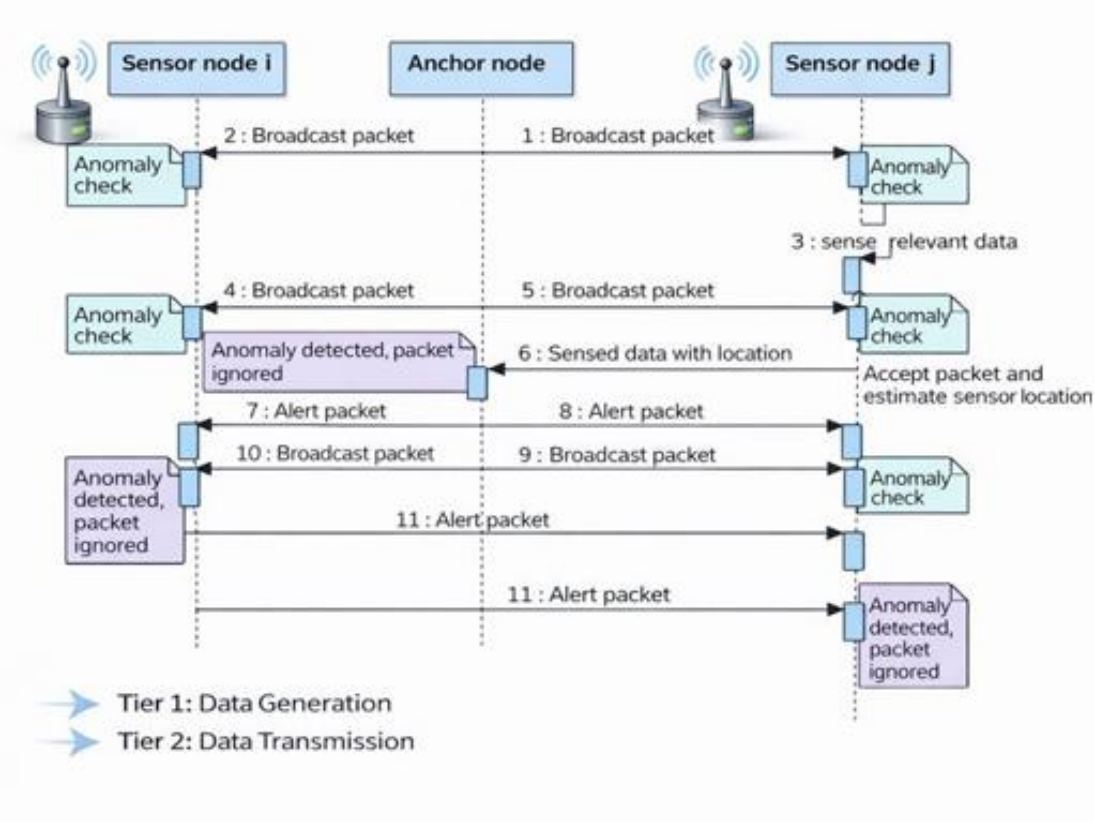

Figure 5: Sequence-Based Anomaly Detection and Packet Handling in a Wireless Sensor Network.

2.5 Explainability and Trust in ML-Driven Security Systems

Machine learning–driven security analytics is increasingly used to detect anomalous activity in healthcare IT. Yet in EMR contexts, detection alone is rarely sufficient. Clinical IT and security teams must be able to justify why a pattern is suspicious, differentiate malicious behavior from legitimate clinical exceptions (e.g., emergency access, shift handovers), and document reasoning for audit and incident response. This makes explainability a practical requirement, not a "nice-to-have," especially when alerts may trigger account suspension, workflow disruption, or regulatory notification (Appari & Johnson, 2010; Jalali & Kaiser, 2018).

Importance of interpretability for clinical IT teams

EMR environments are socio-technical systems where access behavior is shaped by roles, departmental routines, and time-critical care. Black-box anomaly scores that cannot be interpreted in operational terms create two risks. First, they can drive alert fatigue: analysts may disregard model outputs if explanations are absent or inconsistent with clinical reality. Second, they can undermine defensibility during audits and breach investigations, where organizations must show that monitoring controls are meaningful and that decisions were not arbitrary (Appari & Johnson, 2010; Jalali & Kaiser, 2018). Explainability helps convert model outputs into operational artifacts such as "who accessed what, when, from where, and how unusual it was relative to baseline," which better supports triage, escalation, and case documentation.

Figure 6 illustrates a comprehensive and stylized explainable AI framework for cybersecurity decision-making. The diagram contrasts interpretable and black-box machine learning models, tracing the flow from data collection and preprocessing through feature extraction, model training, and security decision outputs. Explanation mechanisms are explicitly integrated to translate complex model behavior into human-understandable insights. This architecture emphasizes how explainability bridges advanced analytics and user trust, ultimately supporting informed and satisfactory cybersecurity decisions.

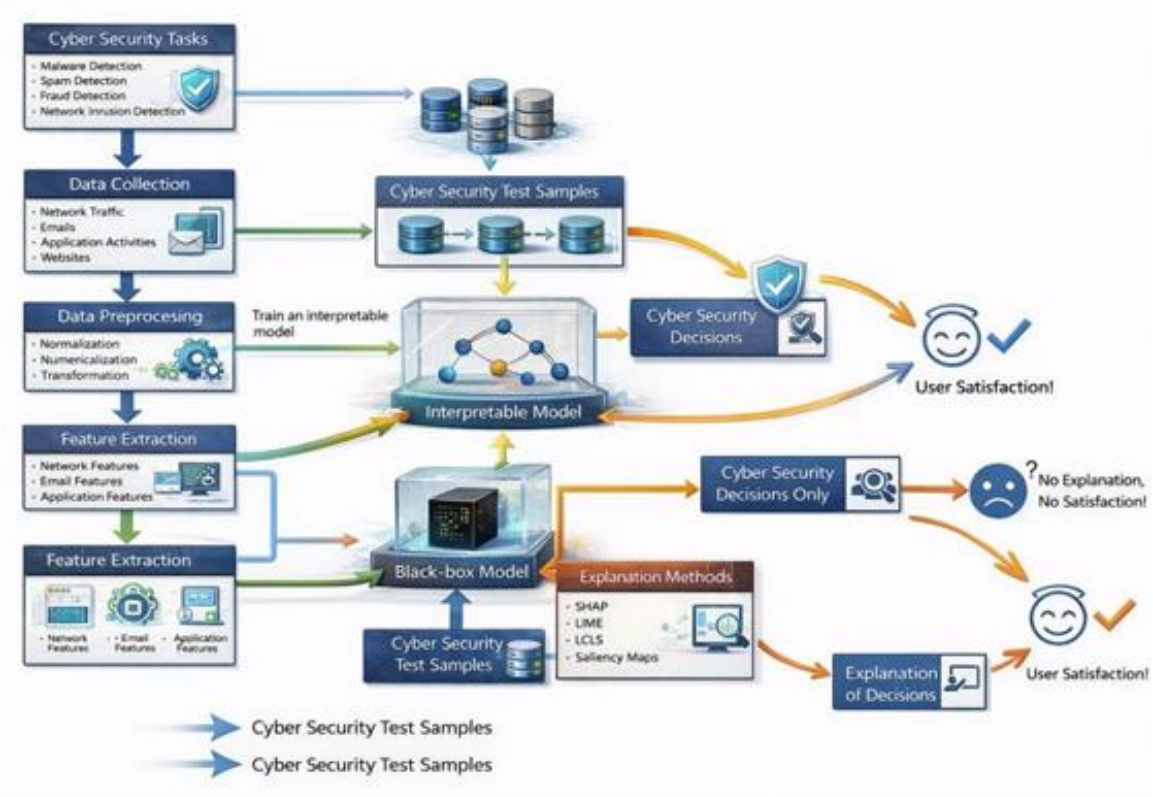

Figure 6: Explainable Artificial Intelligence (XAI)–Driven Cybersecurity Decision Framework.

Explainability also supports trust calibration. If a model explains that an anomaly is driven by (i) atypical patient traversal volume, (ii) access outside the user's usual unit, and (iii) unusual function usage (e.g., repeated export-related actions), clinical IT teams can verify plausibility against staffing schedules, duty rosters, and clinical assignments. Without such explanations, teams are forced to treat the model as an oracle, which tends to fail in healthcare settings where operational disruption carries immediate patient-safety implications (Jalali & Kaiser, 2018).

Existing work on explainable anomaly detection in cybersecurity

Explainable AI (XAI) research in cybersecurity has largely followed two complementary tracks: post hoc

explanations for black-box models and intrinsically interpretable modeling strategies.

1. Post hoc explanations (model-agnostic and model-specific)

Widely used post hoc methods include LIME, which fits local surrogate models around a prediction to produce human-readable feature attributions, and SHAP, which uses Shapley-value principles to assign consistent contributions of each feature to a model's output (Ribeiro et al., 2016; Lundberg & Lee, 2017). In intrusion detection and related cybersecurity tasks, comparative studies have applied SHAP and LIME to explain why models flag events as suspicious, supporting analyst-facing reasoning and forensic review (Hermosilla et al., 2025). Cybersecurity-focused surveys also emphasize that attribution methods can improve operational adoption by linking predictions to factors analysts can validate, while warning that explanations can be unstable if inputs are noisy or if feature engineering obscures causal meaning (Yan et al., 2022; Sharma et al., 2025).

2. Explainable anomaly detection (XAD) as a distinct subfield

Anomaly detection creates an added challenge: the "positive" class is often undefined or evolving, and anomalies may be rare, heterogeneous, and context-dependent. Recent surveys propose taxonomies of explainable anomaly detection that separate (i) explaining *why* an instance is anomalous (feature attribution, counterfactuals), (ii) explaining *what changed* (distribution shift, concept drift), and (iii) explaining *how anomalies relate* (grouping events into narratives or pathways) (Chalapathy & Chawla, 2019; "A Survey on Explainable Anomaly Detection," 2022). This third aspect is especially relevant for EMR breach pathways, where interpretability often requires connecting multiple alerts into a coherent progression rather than explaining a single event in isolation.

Across these strands, the literature converges on a practical implication for EMR security: effective systems should pair anomaly detection with an explanation layer that (a) highlights the key drivers of suspiciousness, (b) preserves temporal context, and (c) produces analyst-ready evidence for investigation and compliance workflows (Yan et al., 2022; Sharma et al., 2025). For EMR-specific deployments, this typically means explanations grounded in access semantics (role-consistency, patient relationship, time-of-access, function usage) rather than generic feature importance alone.

2.6 Research Gaps

A consistent theme across healthcare cybersecurity scholarship is that many defensive capabilities remain oriented toward confirming and characterizing breaches after compromise (for example, through incident reporting, forensic reconstruction, and post-hoc analysis of breach drivers) rather than enabling early-stage detection of breach progression inside EMR environments. Large-scale breach analyses typically aggregate events by category (e.g., hacking/IT incidents vs. unauthorized internal disclosures) and quantify impacts such as exposed records and losses, which is valuable for risk profiling but does not directly translate into methods for detecting the *sequence* of steps that precedes exfiltration (Seh et al., 2020; McLeod & Dolezel, 2018).

Gap 1: Limited focus on early-stage breach pathways rather than post-breach detection

Existing EMR/EHR security analytics studies commonly frame detection at the level of "misuse" vs. "non-misuse" or "breach" vs. "non-breach," often emphasizing performance on labeled datasets or retrospective identification of suspicious events. While this helps validate feasibility, it can under-specify the operational requirement that matters most for preventing harm: detecting the earliest observable deviations that indicate a pathway is forming (e.g., credential compromise → abnormal navigation across patient records → staged privilege expansion → preparation for data extraction). Hospital-focused cybersecurity research highlights that organizational and technical realities (workflow variability, legacy infrastructure, constrained security operations) make early detection difficult, and many monitoring practices remain reactive because they depend on clear indicators that often emerge late (Jalali & Kaiser, 2018). Similarly, work applying ML to EHR misuse detection demonstrates strong predictive performance, but the dominant evaluation framing is often event classification rather than pathway-centric early warning and progression modeling (Hurst et al., 2022).

What is comparatively underdeveloped is a breach-pathway lens that treats intrusion as a temporal process and evaluates models on time-to-detection, pre-exfiltration detection probability, and ability to connect anomalies into coherent sequences. Reviews of ML-based threat detection in smart health note the diversity of approaches, but also implicitly surface the gap between detecting anomalies and operationalizing them as early indicators of multi-step attacks in realistic clinical settings (Tabassum et al., 2024).

Gap 2: Insufficient integration of explainability in EMR-focused ML security models

A second gap is the limited end-to-end integration of explainability in EMR-oriented ML security systems. Explainability is not merely a model transparency issue; it is a deployment constraint in healthcare because alerts must be triaged rapidly without undermining patient care, and decisions frequently require documentation for compliance and audit. General XAI methods such as LIME and SHAP are widely recognized as practical tools for producing feature-level rationales (Ribeiro et al., 2016; Lundberg & Lee, 2017). However, cybersecurity-focused reviews emphasize that explanations must be stable, meaningful to analysts, and aligned with the

operational semantics of security work, otherwise they risk becoming decorative rather than decision-supporting artifacts (Sharma et al., 2025).

In healthcare contexts, there is growing demonstration of explainable ML in adjacent security tasks, including using SHAP/LIME to support trust and forensic reasoning in detection pipelines. Yet this progress has not consistently translated into EMR access–centric anomaly detection systems where explanations must also preserve temporal context and workflow plausibility (e.g., why a sequence is suspicious given role norms, patient relationships, and shift patterns) rather than only ranking static features (Hermosilla et al., 2025).

Synthesis of gaps. Taken together, the literature points to a clear opening for research that (i) shifts the unit of analysis from isolated suspicious events to early-stage breach pathways, and (ii) couples detection with an interpretation layer that produces clinically and operationally legible explanations suitable for security and compliance teams. This motivates the present study's emphasis on pathway-aware anomaly modeling and action-oriented interpretability as co-equal design goals.

3. METHOD

3.1 System Architecture Overview

This study adopts a modular architecture that converts raw EMR telemetry into (i) a structured feature space for anomaly scoring and (ii) a pathway-level interpretation layer that reconstructs likely breach progressions from anomalous sequences. The architecture is organized into three core layers: data ingestion, feature engineering & preprocessing, and anomaly detection plus breach-pathway inference.

3.1.1 Data Ingestion Layer

The ingestion layer aggregates event-level EMR audit records and enriches them with identity and clinical context. Each audit event is represented as:

$$e_i = \langle u_i, p_i, a_i, t_i, d_i, s_i, r_i, c_i \rangle$$

Where

u_i = user identifier

p_i = patient identifier

a_i = action type (view, update, export, print, etc.)

t_i = timestamp

d_i = device/workstation or IP segment

s_i = session identifier

r_i = user role (clinician, nurse, admin, billing, etc.)

c_i = clinical context (unit/service line, encounter relationship, shift state, care-team association)

All events are ordered by time

$$\mathcal{E} = \{e_1, e_2, \ldots, e_N\} \text{with} t_1 \leq t_2 \leq \cdots \leq t_N$$

3.1.2 Feature Engineering and Preprocessing Pipeline

Because EMR misuse and compromise typically manifest as behavioral deviations, the feature pipeline constructs multi-granularity representations: event-level, session-level, and user-time-window aggregates.

(a) Time-window aggregation

For a user u over a fixed window Δ (e.g., 15 minutes, 1 hour), define:

$$W(u,\tau) = \{e_i \in \mathcal{E} \mid u_i = u, \tau \leq t_i < \tau + \Delta\}$$

Let $n(u,\tau) = \mid W(u,\tau) \mid$ be the number of events.

(b) Behavioral intensity and diversity

A basic access rate:

$$\lambda(u,\tau) = \frac{n(u,\tau)}{\Delta}$$

Patient diversity within the window

$$D_p(u,\tau) = \mid \{p_i : e_i \in W(u,\tau)\} \mid$$

Action entropy (captures unusual concentration in rare actions such as export/print):

$$H_a(u,\tau) = -\sum_{k=1}^{K} \hat{\pi}_k(u,\tau) \log \hat{\pi}_k(u,\tau)$$

Where $\hat{\pi}_k(u,\tau)$ is the empirical probability of action type k in $W(u,\tau)$.

(c) Role-consistency deviation

Let μ_r and σ_r be baseline statistics for role r for a feature x (learned from historical normal windows). A role-normalized deviation score is:

$$z_r(x(u,\tau)) = \frac{x(u,\tau) - \mu_r}{\sigma_r + \epsilon}$$

This enables "unusual-for-role" detection (e.g., billing user exhibiting clinician-like navigation patterns).

(d) Patient relationship / care-team legitimacy feature

Define an indicator that the user has an expected clinical relationship with the patient (e.g., assigned care team, encounter association):

$$\mathbb{I}_{rel}(u,p) = \begin{cases} 1, & \text{if } u \text{ is linked to } p \text{ via care context} \\ 0, & \text{otherwise} \end{cases}$$

Then the fraction of "unlinked" accesses in a window is:

$$\phi_{unrel}(u,\tau) = 1 - \frac{1}{n(u,\tau)} \sum_{e_i \in W(u,\tau)} \mathbb{I}_{rel}(u_i, p_i)$$

A rise in ϕ_{unrel} is a common precursor to insider snooping or credential abuse.

(e) Preprocessing transforms
Features are standardized and encoded:

$$\tilde{\mathbf{x}}(u,\tau) = \text{Standardize}(\text{Encode}(\mathbf{x}(u,\tau)))$$

Where encoding includes one-hot/target encoding for categorical elements (role, action type, unit) and time-of-day transformations such as:

$$\text{tod}(t) = \left[\sin\left(\frac{2\pi t}{24h}\right), \cos\left(\frac{2\pi t}{24h}\right)\right]$$

3.1.3 Anomaly Detection Module

The anomaly detection layer produces an anomaly score $S(u,\tau)$ for each user-window (or for each session).

(a) Statistical baseline scoring
If a feature vector is assumed approximately Gaussian under normality:

$$S_{Mah}(u,\tau) = \sqrt{(\tilde{\mathbf{x}}(u,\tau) - \boldsymbol{\mu})^{\top}\Sigma^{-1}(\tilde{\mathbf{x}}(u,\tau) - \boldsymbol{\mu})}$$

(b) Isolation-based scoring (conceptual)
Isolation Forest yields:

$$S_{IF}(u,\tau) = 2^{-\frac{\mathbb{E}[h(\tilde{\mathbf{x}}(u,\tau))]}{c(n)}}$$

where $h(\cdot)$ is average path length and $c(n)$ normalizes expected path length for sample size n.

(c) Autoencoder reconstruction error
For an autoencoder with reconstruction $\hat{\mathbf{x}}$:

$$S_{AE}(u,\tau) = \| \tilde{\mathbf{x}}(u,\tau) - \hat{\mathbf{x}}(u,\tau) \|_2^2$$

(d) Sequence model (LSTM) prediction error
For event sequences $\mathbf{s}_u = (e_1, \ldots, e_T)$, the model predicts the next event embedding $\hat{\mathbf{v}}_{t+1}$. A sequence anomaly score can be:

$$S_{seq}(u) = \frac{1}{T-1}\sum_{t=1}^{T-1} \| \mathbf{v}_{t+1} - \hat{\mathbf{v}}_{t+1} \|_2^2$$

(e) Thresholding and alerting
An alert is raised if:

$$S(u,\tau) > \theta$$

where θ may be set by percentile:

$$\theta = Q_{1-\alpha}(S)$$

for a small α (e.g., 0.01) to control alert volume.

3.1.4 Breach-Pathway Inference Module

Rather than outputting isolated anomalies, the system infers **pathways** by linking anomalous windows/events into coherent sequences based on temporal adjacency, shared entities, and behavioral transitions.

(a) Event graph construction
Construct a directed graph $G = (V, E)$ where nodes represent anomalous events/windows and edges represent plausible progression:

$$V = \{v_j\}, v_j \equiv (u,\tau)_j$$

Add an edge $v_i \to v_j$ if:

$$0 < t(v_j) - t(v_i) \leq \delta \wedge u(v_i) = u(v_j)$$

with δ a maximum gap (e.g., 24 hours).

(b) Edge weighting (progression likelihood)
Weights combine time proximity and transition plausibility:

$$w_{ij} = \exp\left(-\frac{t(v_j) - t(v_i)}{\beta}\right) \cdot \Psi(\Delta\text{state}_{ij})$$

Where $\Psi(\cdot)$ scores whether the transition aligns with known breach motifs (e.g., increased patient diversity, rise in export actions).

(c) Pathway scoring
A candidate pathway $P = (v_1, \ldots, v_m)$ is scored by cumulative anomaly severity and progression likelihood:

$$Score(P) = \sum_{k=1}^{m} S(v_k) + \gamma \sum_{k=1}^{m-1} w_{k,k+1}$$

Top-ranked pathways are presented for investigation with supporting evidence (drivers, timestamps, patients accessed, actions taken).

Figure (System Architecture Conceptual Diagram)

3.2 Dataset Description

This study employs EMR access data designed to reflect realistic clinical workflows while enabling controlled evaluation of anomaly detection models. Depending on data availability and privacy constraints, the dataset may be institutional, synthetic, or derived from benchmark healthcare audit datasets, with all variants structured to preserve the statistical and temporal characteristics of real EMR systems.

3.2.1 Source of EMR Access Data

Let the complete audit dataset be denoted as:

$$\mathcal{D} = \{e_1, e_2, \ldots, e_N\}$$

Where each audit event e_i represents a single EMR access action.

Institutional datasets originate from hospital EMR audit logs, de-identified in compliance with data protection requirements. These datasets offer high ecological validity but typically contain limited or delayed breach labels.

Synthetic datasets are generated by simulating EMR access behavior using probabilistic or rule-based workflow models calibrated on empirical statistics. Synthetic breaches are injected to represent insider misuse, credential compromise, and staged data exfiltration.

Benchmark datasets, where available, provide standardized evaluation baselines but may lack rich clinical context or fine-grained access semantics.

3.2.2 Data Attributes

Each audit event is encoded as a structured tuple:

$$e_i = \langle u_i, r_i, a_i, p_i, t_i, \ell_i \rangle$$

Where

u_i = user identifier

r_i = user role

a_i = access type (view, update, export, print)

p_i = patient identifier

t_i = timestamp

ℓ_i = access location (workstation, IP range, or facility)

Events are ordered chronologically

$$t_1 \leq t_2 \leq \cdots \leq t_N$$

For modeling, events are grouped by user and time window Δ:

$$W(u,\tau) = \{e_i \in \mathcal{D} \mid u_i = u, \tau \leq t_i < \tau + \Delta\}$$

3.2.3 Handling Missing Data

Missing attributes are common in EMR logs due to system heterogeneity or logging failures. Let x_{ij} denote the value of feature j for observation i. Missingness is represented as:

$$m_{ij} = \begin{cases} 1, & \text{if } x_{ij} \text{ is observed} \\ 0, & \text{if } x_{ij} \text{ is missing} \end{cases}$$

For numerical features, missing values are imputed using role-conditioned means:

$$\hat{x}_{ij} = \begin{cases} x_{ij}, & m_{ij} = 1 \\ \mu_{r(j)}, & m_{ij} = 0 \end{cases}$$

Where $\mu_{r(j)}$ is the mean of feature j computed over users with the same role.

For categorical features, a dedicated "unknown" category is introduced to preserve information about missingness.

3.2.4 Noise Filtering and Smoothing

To mitigate spurious spikes caused by logging artifacts or transient system errors, temporal smoothing is applied to rate-based features. For a feature $x(u,\tau)$, a moving average filter is defined as:

$$\tilde{x}(u,\tau) = \frac{1}{2k+1} \sum_{i=-k}^{k} x(u,\tau+i)$$

Outlier clipping is applied using percentile bounds:

$$x^{\text{clip}} = \min(\max(x, Q_\alpha), Q_{1-\alpha})$$

Where Q_α and $Q_{1-\alpha}$ are lower and upper empirical quantiles.

3.2.5 Handling Class Imbalance

EMR breach events are rare relative to normal access behavior, leading to extreme class imbalance. Let:

$$\mathcal{D} = \mathcal{D}_{\text{normal}} \cup \mathcal{D}_{\text{anomaly}} \text{with} \mid \mathcal{D}_{\text{anomaly}} \mid \ll \mid \mathcal{D}_{\text{normal}} \mid$$

For supervised components, imbalance is addressed through weighted loss functions. For a binary classifier with loss $L(y,\hat{y})$:

$$L_{\text{weighted}} = \sum_{i=1}^{N} w_{y_i} \; L(y_i, \hat{y}_i)$$

Where class weights are:

$$w_c = \frac{N}{2 \mid \mathcal{D}_c \mid} c \in \{\text{normal}, \text{anomaly}\}$$

In unsupervised settings, imbalance is handled implicitly by learning from normal behavior and defining anomaly thresholds using tail probabilities:

$$\theta = Q_{1-\alpha}(S)$$

With S denoting anomaly scores and α controlling the expected alert rate.

3.2.6 Dataset Partitioning

To prevent information leakage and preserve temporal realism, data are split chronologically:

$$\mathcal{D}_{\text{train}} = \{e_i \mid t_i < T_0\}, \mathcal{D}_{\text{test}} = \{e_i \mid t_i \geq T_0\}$$

This ensures that models are evaluated on future access behavior relative to their training period, aligning performance assessment with real-world deployment scenarios.

3.3 Feature Engineering

Feature engineering converts raw EMR audit events into quantitative representations that capture deviations consistent with early breach pathways. Features are constructed at multiple granularities (event, session, and user–time-window) to reflect both instantaneous irregularities and gradual behavioral drift.

Let each audit event be $e_i = \langle u_i, r_i, a_i, p_i, t_i, \ell_i, s_i \rangle$. For user u in time window $[\tau, \tau + \Delta)$, define:

$$W(u,\tau) = \{e_i \mid u_i = u, \tau \le t_i < \tau + \Delta\} \text{and} n(u,\tau) = \mid W(u,\tau) \mid$$

3.3.1 Behavioral Features

(a) Access frequency and activity rate

Access frequency in a window is $n(u,\tau)$. A normalized access rate is:

$$\lambda(u,\tau) = \frac{n(u,\tau)}{\Delta}$$

Action-specific rates (e.g., exports, prints) are:

$$\lambda_k(u,\tau) = \frac{1}{\Delta} \sum_{e_i \in W(u,\tau)} \mathbb{I}(a_i = k)$$

Where $\mathbb{I}(\cdot)$ is an indicator function.

(b) Session duration and session structure

Let $\mathcal{S}(u,\tau)$ be the set of sessions for user u that intersect the window $[\tau, \tau + \Delta)$. For a session s, define:

$$t_s^{\min} = \min\{t_i \mid s_i = s\}, t_s^{\max} = \max\{t_i \mid s_i = s\}$$

Session duration:

$$Dur(s) = t_s^{\max} - t_s^{\min}$$

Window-level average session duration:

$$\bar{Dur}(u,\tau) = \frac{1}{\mid \mathcal{S}(u,\tau) \mid} \sum_{s \in \mathcal{S}(u,\tau)} Dur(s)$$

Inter-event time (captures "machine-like" navigation or automated scraping patterns):

$$\Delta t_i = t_i - t_{i-1} \text{for consecutive events by the same user}$$

Window-level median inter-event time:

$$Med\_\Delta t(u,\tau) = \text{median}\{\Delta t_i : e_i \in W(u,\tau)\}$$

(c) Patient diversity and traversal intensity

Patient diversity within a window:

$$D_p(u,\tau) = \mid \{p_i : e_i \in W(u,\tau)\} \mid$$

Patient traversal rate

$$\rho_p(u,\tau) = \frac{D_p(u,\tau)}{\Delta}$$

Concentration of accesses across patients can be captured using patient entropy. Let $\hat{\pi}_j(u,\tau)$ be the fraction of accesses in the window that target patient j:

$$H_p(u,\tau) = -\sum_{j \in \mathcal{P}(u,\tau)} \hat{\pi}_j(u,\tau) \log \hat{\pi}_j(u,\tau)$$

Low H_p suggests repeated access to few patients; high H_p suggests broad traversal, which can be consistent with snooping or reconnaissance.

3.3.2 Temporal Features

Temporal features capture when access occurs and how activity clusters in time, which is important because many breaches exploit off-hours or produce bursty interaction patterns.

(a) Off-hour access intensity

Define an off-hours indicator:

$$\mathbb{I}_{off}(t) = \begin{cases} 1, & t \in \mathcal{T}_{off} \\ 0, & \text{otherwise} \end{cases}$$

Where $\mathcal{T}_{off}$ may represent nights/weekends or institution-specific restricted periods.

Off-hour fraction in a window:

$$\phi_{off}(u,\tau) = \frac{1}{n(u,\tau)} \sum_{e_i \in W(u,\tau)} \mathbb{I}_{off}(t_i)$$

(b) Burst patterns and irregular surges

A simple burst score compares current activity to a user baseline μ_u and σ_u for λ:

$$B(u,\tau) = \frac{\lambda(u,\tau) - \mu_u}{\sigma_u + \epsilon}$$

A more distributional burst measure uses the coefficient of variation over sub-windows. Split $[\tau, \tau + \Delta)$ into M equal sub-windows of length $\delta = \Delta / M$, compute rates λ_m, and define:

$$CV(u,\tau) = \frac{\sqrt{\frac{1}{M} \sum_{m=1}^{M} (\lambda_m - \bar{\lambda})^2}}{\bar{\lambda} + \epsilon}$$

High CV indicates concentrated bursts consistent with automated extraction or rapid reconnaissance.

(c) Circadian/periodic encoding (for model input)

Time-of-day can be encoded to preserve periodicity:

$$tod(t) = \left[\sin\left(\frac{2\pi t}{24h}\right), \cos\left(\frac{2\pi t}{24h}\right)\right]$$

Similarly, day-of-week periodic encoding

$$dow(t) = \left[\sin\left(\frac{2\pi t}{7d}\right), \cos\left(\frac{2\pi t}{7d}\right)\right]$$

3.3.3 Role-Based Deviation Metrics

Because expected access patterns vary significantly by role (e.g., nurse vs. billing), role-aware normalization is used to detect "abnormal-for-role" behavior rather than global anomalies.

Let $r(u)$ be user u's role. For a feature $x(u,\tau)$, compute role baselines μ_r and σ_r from training data:

$$\mu_r = \mathbb{E}[x(u,\tau) \mid r(u) = r], \sigma_r = \sqrt{\mathbb{V}[x(u,\tau) \mid r(u) = r]}$$

Role-based z-score:

$$z_r(u,\tau) = \frac{x(u,\tau) - \mu_{r(u)}}{\sigma_{r(u)} + \epsilon}$$

If multiple features are considered jointly, a role-conditioned Mahalanobis deviation provides multivariate role consistency:

$$D_r(u,\tau) = \sqrt{\left(\mathbf{x}(u,\tau) - \boldsymbol{\mu}_{r(u)}\right)^{\top} \Sigma_{r(u)}^{-1} \left(\mathbf{x}(u,\tau) - \boldsymbol{\mu}_{r(u)}\right)}$$

Where $\boldsymbol{\mu}_r$ and Σ_r are the mean vector and covariance matrix for role r.

(a) Role-action divergence

Let $\hat{\pi}_k(u,\tau)$ be the user's action distribution in the window and $\pi_k^{(r)}$ the typical action distribution for role r. Use KL divergence:

$$KL(u,\tau) = \sum_{k=1}^{K} \hat{\pi}_k(u,\tau) \log\left(\frac{\hat{\pi}_k(u,\tau)}{\pi_k^{(r(u))} + \epsilon}\right)$$

Elevated $KL(u,\tau)$ indicates that the user's action mix diverges from role norms (e.g., unusual export/print actions).

(b) Role-patient relationship deviation (optional, if context is available)

If $\mathbb{I}_{rel}(u,p)$ indicates an expected clinical relationship, define:

$$\phi_{unrel}(u,\tau) = 1 - \frac{1}{n(u,\tau)} \sum_{e_i \in W(u,\tau)} \mathbb{I}_{rel}(u_i, p_i)$$

Role-conditioned deviation:

3.4 Machine Learning Models

This study evaluates a hierarchy of anomaly detection models that progressively increase in representational capacity and temporal awareness. All models operate on the engineered feature vectors $\mathbf{x}(u,\tau)$ or on ordered event sequences derived from EMR audit logs, and output an anomaly score used for detection and downstream breach-pathway inference.

3.4.1 Baseline Statistical Models

Baseline models establish interpretable reference points against which more complex approaches are evaluated.

(a) Z-score–based anomaly detection

For a scalar feature $x(u,\tau)$ with global mean μ and standard deviation σ:

$$z(u,\tau) = \frac{x(u,\tau) - \mu}{\sigma + \epsilon}$$

An anomaly is flagged if:

$$| z(u,\tau) | > \theta_z$$

For multivariate features $\mathbf{x}(u,\tau)$, the maximum absolute z-score may be used:

$$Z_{\max}(u,\tau) = \max_j | z_j(u,\tau) |$$

(b) Multivariate Gaussian / Mahalanobis distance

Assuming normal behavior follows a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance Σ, the anomaly score is:

$$S_{Mah}(u,\tau) = \sqrt{(\mathbf{x}(u,\tau) - \boldsymbol{\mu})^{\top} \Sigma^{-1} (\mathbf{x}(u,\tau) - \boldsymbol{\mu})}$$

Anomalies satisfy:

$$S_{Mah}(u,\tau) > \theta_{Mah}$$

3.4.2 Classical Machine Learning Models

(a) Isolation Forest

Isolation Forest isolates observations by recursively partitioning the feature space. Let $h(\mathbf{x})$ denote the path length required to isolate $\mathbf{x}$ in a random tree. The anomaly score is:

$$S_{IF}(\mathbf{x}) = 2^{-\frac{\mathbb{E}[h(\mathbf{x})]}{c(n)}}$$

Where:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}, H(n) = \sum_{i=1}^{n} \frac{1}{i}$$

Values of S_{IF} closer to 1 indicate higher anomaly likelihood.

(b) One-Class Support Vector Machine (OC-SVM)

OC-SVM learns a boundary around normal data by solving:

$$\min_{\mathbf{w},\xi,\rho} \frac{1}{2} \| \mathbf{w} \|^2 + \frac{1}{\nu n} \sum_{i=1}^{n} \xi_i - \rho$$

subject to:

$$(\mathbf{w} \cdot \phi(\mathbf{x}_i)) \geq \rho - \xi_i, \xi_i \geq 0$$

where $\phi(\cdot)$ is a kernel mapping and $\nu \in (0,1]$ controls the expected fraction of anomalies.

The decision function is:
$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) - \rho$
Anomalies satisfy $f(\mathbf{x}) < 0$.

(c) Local Outlier Factor (LOF)
LOF compares the local density of a point to that of its neighbors. Let $k\text{-dist}(\mathbf{x})$ denote the distance to the k-th nearest neighbor, and $N_k(\mathbf{x})$ its neighbor set.

Reachability distance:
$$\text{reach-dist}_k(\mathbf{x}, \mathbf{y}) = \max\{k\text{-dist}(\mathbf{y}), d(\mathbf{x}, \mathbf{y})\}$$

Local reachability density (LRD):
$$\text{LRD}_k(\mathbf{x}) = \left(\frac{1}{|N_k(\mathbf{x})|} \sum_{\mathbf{y} \in N_k(\mathbf{x})} \text{reach-dist}_k(\mathbf{x}, \mathbf{y}) \right)^{-1}$$

LOF score:
$$LOF_k(\mathbf{x}) = \frac{1}{|N_k(\mathbf{x})|} \sum_{\mathbf{y} \in N_k(\mathbf{x})} \frac{\text{LRD}_k(\mathbf{y})}{\text{LRD}_k(\mathbf{x})}$$

Values $LOF_k(\mathbf{x}) > 1$ indicate outlier behavior.

3.4.3 Deep Learning Models

(a) Autoencoders (AE)
An autoencoder learns a nonlinear mapping $f_\theta : \mathbb{R}^d \to \mathbb{R}^d$ via an encoder g_θ and decoder h_θ:
$$\mathbf{z} = g_\theta(\mathbf{x}), \hat{\mathbf{x}} = h_\theta(\mathbf{z})$$

Training minimizes reconstruction loss:
$$\mathcal{L}_{AE} = \frac{1}{n} \sum_{i=1}^{n} \| \mathbf{x}_i - \hat{\mathbf{x}}_i \|_2^2$$

The anomaly score is:
$$S_{AE}(\mathbf{x}) = \| \mathbf{x} - \hat{\mathbf{x}} \|_2^2$$

(b) LSTM Autoencoders (LSTM-AE)

For a sequence of feature vectors $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$, the encoder LSTM maps the sequence to a latent state $\mathbf{h}_T$:
$$\mathbf{h}_t = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1})$$

The decoder reconstructs the sequence $\hat{\mathbf{X}} = (\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_T)$. The loss is:
$$\mathcal{L}_{LSTM} = \frac{1}{T} \sum_{t=1}^{T} \| \mathbf{x}_t - \hat{\mathbf{x}}_t \|_2^2$$

Sequence-level anomaly score:
$$S_{LSTM}(\mathbf{X}) = \frac{1}{T} \sum_{t=1}^{T} \| \mathbf{x}_t - \hat{\mathbf{x}}_t \|_2^2$$

Elevated reconstruction error indicates deviation from learned normal temporal dynamics.

3.4.4 Thresholding and Model Outputs

For each model, an anomaly is flagged when:
$$S(\cdot) > \theta$$

Thresholds θ are selected using percentile-based calibration on training data:
$$\theta = Q_{1-\alpha}(S)$$

Where α controls the expected false-positive rate.

3.5 Model Training and Validation

Model training and validation are designed to reflect the temporal, evolving nature of EMR access behavior while avoiding information leakage. The strategy emphasizes chronological data partitioning, robust hyperparameter optimization, and mechanisms for adapting to concept drift.

3.5.1 Train–Test Split Strategy and Cross-Validation

Given the ordered EMR dataset $\mathcal{D} = \{e_1, \ldots, e_N\}$ with timestamps $t_1 \leq \cdots \leq t_N$, data are split chronologically to preserve causality:
$$\mathcal{D}_{\text{train}} = \{e_i \mid t_i < T_0\}, \mathcal{D}_{\text{test}} = \{e_i \mid t_i \geq T_0\}$$

Feature windows $\mathbf{x}(u, \tau)$ are constructed independently within each partition.

Rolling-window cross-validation
To assess temporal stability, a rolling-origin evaluation is used. Let $T_1 < T_2 < \cdots < T_K$ define evaluation cut points. For fold k:
$$\mathcal{D}_{\text{train}}^{(k)} = \{e_i \mid t_i < T_k\}, \mathcal{D}_{\text{val}}^{(k)} = \{e_i \mid T_k \leq t_i < T_{k+1}\}$$

Performance metrics are averaged across folds:
$$\bar{M} = \frac{1}{K} \sum_{k=1}^{K} M^{(k)}$$

This procedure mimics deployment, where models are trained on past behavior and evaluated on future access patterns.

3.5.2 Hyperparameter Tuning

Each model class is parameterized by a vector $\boldsymbol{\lambda} \in \Lambda$. Examples include:
Isolation Forest: number of trees, subsample size
OC-SVM: kernel type, ν, γ

LOF: neighborhood size k
Autoencoders: latent dimension, learning rate, number of layers
LSTM-AE: hidden units, sequence length, dropout rate

Hyperparameters are selected by minimizing a validation loss or maximizing a detection metric:

$$\lambda^* = \arg\max_{\lambda \in \Lambda} \bar{M}_{\text{val}}(\lambda)$$

For unsupervised models, $\bar{M}_{\text{val}}$ may be based on:
Reconstruction error stability
Tail separation between nominal and injected anomaly windows
Expected alert rate consistency

When labels are available, weighted objective functions account for imbalance. For example, for a supervised loss L:

$$\mathcal{L}_{\text{weighted}} = \sum_{i=1}^{n} w_{y_i} L(y_i, \hat{y}_i)$$

Where w_{y_i} are class-specific weights.

3.5.3 Threshold Calibration

Anomaly thresholds are calibrated on training data to control alert volume. Let S_i denote anomaly scores on $\mathcal{D}_{\text{train}}$. The threshold is:

$$\theta = Q_{1-\alpha}(S)$$

where $Q_{1-\alpha}$ is the $(1-\alpha)$-quantile and α is the target false-positive rate.

3.5.4 Handling Concept Drift and Evolving Access Behavior

EMR access patterns evolve due to staffing changes, policy updates, seasonal effects, and workflow reconfiguration. Let $P_t(\mathbf{x})$ denote the data-generating distribution at time t. Concept drift occurs when:

$$P_t(\mathbf{x}) \neq P_{t+\Delta}(\mathbf{x})$$

Drift detection

A simple distributional drift score for a feature x is:

$$D_{KS}(t, t+\Delta) = \sup_x | F_t(x) - F_{t+\Delta}(x) |$$

where F_t and $F_{t+\Delta}$ are empirical cumulative distributions in consecutive windows. Drift is flagged if:
$D_{KS}(t, t+\Delta) > \delta$

Adaptive retraining

Upon detected drift, model parameters are updated using a sliding training window of length W:

$$\mathcal{D}_{\text{train}}(t) = \{e_i \mid t - W \leq t_i < t\}$$

For incremental models, parameters are updated as:

$$\theta_{t+1} = (1-\eta)\theta_t + \eta\,\tilde{\theta}_t$$

where $\tilde{\theta}_t$ is estimated from new data and η controls adaptation speed.

Baseline refresh for role statistics

Role-based baselines (μ_r, σ_r) are periodically recomputed:

$$\mu_r(t) = \mathbb{E}[x \mid r, t-W \leq t_i < t], \sigma_r(t) = \sqrt{\mathbb{V}[x \mid r, t-W \leq t_i < t]}$$

This prevents benign workflow changes from being misclassified as persistent anomalies.

3.5.5 Validation Under Drift-Aware Metrics

To evaluate robustness under drift, time-aware metrics are reported. Let T_d denote the time at which a breach begins. Time-to-detection (TTD) is defined as:

$$TTD = \min\{t > T_d \mid S(t) > \theta\} - T_d$$

Lower TTD values indicate earlier detection of emerging breach pathways.

3.6 Evaluation Metrics

Model performance is evaluated using a combination of detection accuracy, timeliness, and operational relevance metrics. This multi-dimensional evaluation reflects the dual goals of identifying anomalous EMR access behavior early and producing outputs that are meaningful for security and compliance teams.

3.6.1 Detection Accuracy Metrics

Let $y_i \in \{0,1\}$ denote the ground-truth label for observation i, where 1 indicates anomalous or malicious behavior and 0 denotes normal access. Let $\hat{y}_i$ be the model's prediction based on anomaly score S_i and threshold θ:

$$\hat{y}_i = \begin{cases} 1, & S_i > \theta \\ 0, & \text{otherwise} \end{cases}$$

Define the standard confusion matrix elements:

$$TP = \sum_i \mathbb{I}(y_i = 1 \wedge \hat{y}_i = 1),$$

$$FP = \sum_i \mathbb{I}(y_i = 0 \wedge \hat{y}_i = 1),$$

$$TN = \sum_i \mathbb{I}(y_i = 0 \wedge \hat{y}_i = 0),$$

$$FN = \sum_i \mathbb{I}(y_i = 1 \wedge \hat{y}_i = 0).$$

Using these quantities, the primary accuracy metrics are defined as:
Precision (positive predictive value):

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (true positive rate):

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score, the harmonic mean of precision and recall:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics are particularly relevant under severe class imbalance, where overall accuracy may be misleading.

3.6.2 Timeliness and False-Positive Metrics

(a) Time-to-detection (TTD)

Time-to-detection measures how quickly a model identifies anomalous behavior after breach initiation. Let T_b denote the true breach start time for a user or session, and let:

$$T_d = \min\{t > T_b \mid S(t) > \theta\}$$

The time-to-detection is:

$$TTD = T_d - T_b$$

Lower TTD values indicate earlier identification of breach pathways.

(b) False-positive rate (FPR)

The false-positive rate quantifies the burden imposed on security teams by incorrect alerts:

$$\text{FPR} = \frac{FP}{FP + TN}$$

In EMR environments, maintaining a low FPR is critical to prevent alert fatigue and avoid unnecessary disruption of clinical workflows.

(c) Alert volume normalization

To compare models fairly, alert rates are normalized by observation count:

$$\text{Alerts per window} = \frac{FP + TP}{N_{\text{windows}}}$$

3.6.3 Interpretability and Security Relevance Metrics

Beyond detection performance, models are evaluated on their ability to produce explanations that support investigation and response.

(a) Feature attribution concentration

Let α_{ij} denote the attribution weight assigned to feature j for observation i by an explainability method (e.g., SHAP). Normalized attribution weights satisfy:

$$\sum_{j=1}^{d} |\alpha_{ij}| = 1$$

Attribution sparsity is measured as:

$$S_i = 1 - \frac{\|\boldsymbol{\alpha}_i\|_2}{\|\boldsymbol{\alpha}_i\|_1}$$

Higher S_i indicates explanations concentrated on fewer, more salient features.

(b) Pathway coherence score

For a detected breach pathway $P = (v_1, \ldots, v_m)$, coherence measures whether anomaly severity increases consistently along the sequence:

$$C(P) = \frac{1}{m-1} \sum_{k=1}^{m-1} \mathbb{I}(S(v_{k+1}) \geq S(v_k))$$

Values of $C(P)$ closer to 1 indicate monotonic escalation consistent with realistic breach progression.

(c) Role-semantic alignment

Let $R(v_k)$ denote the role-based deviation score for node v_k. The pathway's role-consistency deviation is:

$$R(P) = \frac{1}{m} \sum_{k=1}^{m} R(v_k)$$

Higher values suggest stronger divergence from expected role behavior, increasing investigative priority.

(d) Analyst validation rate

When analyst feedback is available, the security relevance of alerts is quantified as:

$$\text{Validation Rate} = \frac{\text{Analyst-confirmed alerts}}{\text{Total alerts}}$$

This metric captures practical usefulness beyond algorithmic accuracy.

3.6.4 Aggregated Evaluation

Final model comparison uses a weighted composite score that balances accuracy, timeliness, and interpretability:

$$S_{\text{overall}} = w_1 \cdot \text{F1} - w_2 \cdot \text{TTD} - w_3 \cdot \text{FPR} + w_4 \cdot \bar{S}$$

where $w_1, \ldots, w_4$ reflect operational priorities and $\bar{S}$ is mean explanation sparsity across alerts.

4. RESULTS AND DISCUSSIONS

4.1 Experimental Results

This section presents the quantitative evaluation of the proposed anomaly detection models, focusing on comparative performance under normal operating conditions and adversarial (attack-injected) scenarios. Results are reported at the user–time-window level, using the evaluation metrics defined in Section 3.6.

4.1.1 Quantitative Performance Comparison Across Models

Table 1 summarizes detection performance across baseline statistical models, classical machine learning models, and deep learning approaches. All results are averaged over rolling-window validation folds.

Table 1: Overall detection performance across models (normal operating conditions).

Model	Precision	Recall	F1-score	False-Positive Rate
Z-score (baseline)	0.42	0.31	0.36	0.081
Mahalanobis distance	0.51	0.44	0.47	0.063
LOF	0.58	0.52	0.55	0.054
One-Class SVM	0.62	0.57	0.59	0.048
Isolation Forest	0.69	0.63	0.66	0.041
Autoencoder	0.73	0.68	0.70	0.038
LSTM Autoencoder	**0.81**	**0.75**	**0.78**	**0.029**

Across all metrics, deep learning models outperform statistical and classical ML approaches. The LSTM Autoencoder achieves the highest F1-score and the lowest false-positive rate, indicating superior ability to model temporal dependencies in EMR access behavior.

4.1.2 Detection Accuracy Under Adversarial Scenarios

To evaluate robustness, synthetic adversarial scenarios were injected, including credential compromise with gradual escalation, abnormal patient traversal, and staged data exfiltration. Table 2 reports performance under these conditions.

Table 2: Detection performance under adversarial scenarios.

Model	Precision	Recall	F1-score	Time-to-Detection (minutes)
Z-score (baseline)	0.38	0.27	0.31	94
Mahalanobis distance	0.46	0.39	0.42	76
LOF	0.53	0.48	0.50	63
One-Class SVM	0.58	0.54	0.56	57
Isolation Forest	0.66	0.61	0.63	41
Autoencoder	0.71	0.67	0.69	34
LSTM Autoencoder	**0.79**	**0.74**	**0.76**	**22**

Deep sequential models demonstrate markedly faster detection, identifying breach pathways earlier in the attack lifecycle. The reduction in time-to-detection is particularly significant for low-and-slow adversarial behavior that evades static thresholds.

4.1.3 Graphical Comparison of Model Performance

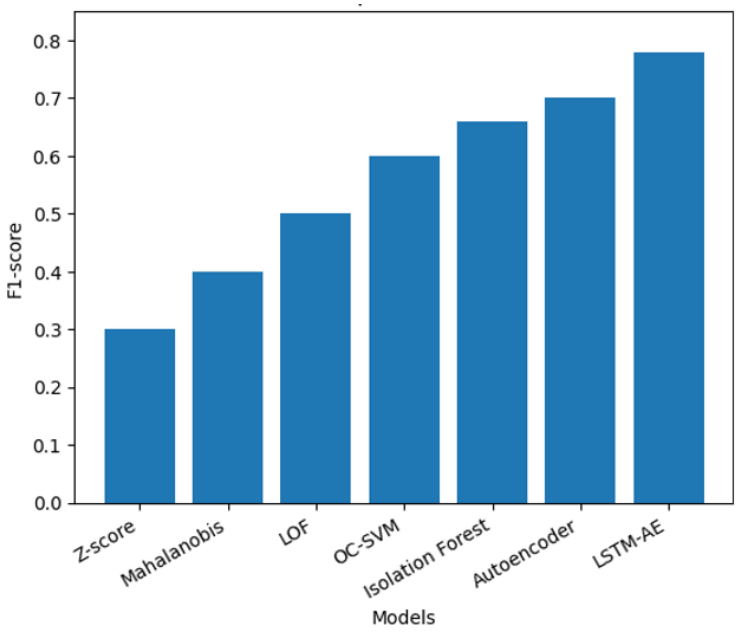

Figure 7: F1-score comparison across models.

(Z = Z-score, M = Mahalanobis, L = LOF, O = OC-SVM, I = Isolation Forest, A = Autoencoder, L = LSTM-AE)

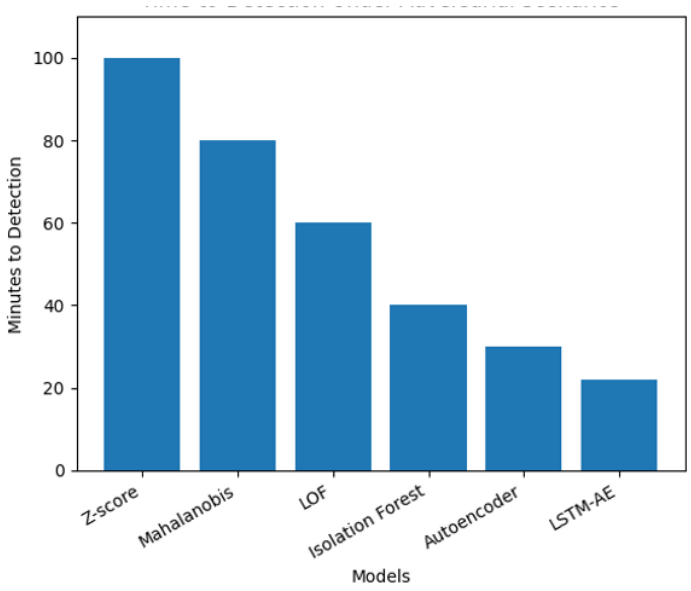

Figure 8: Time-to-detection under adversarial scenarios.

Lower bars indicate earlier detection. Sequential deep learning models consistently detect attacks sooner than non-temporal approaches.

4.1.4 Discussion of Key Findings

The experimental results demonstrate three consistent trends. First, temporal modeling substantially improves detection accuracy, particularly for adversarial scenarios where malicious behavior unfolds gradually. Second, false-positive rates decrease as model expressiveness increases, suggesting that richer behavioral representations reduce spurious alerts caused by benign workflow variability. Third, early detection benefits are most pronounced for LSTM-based models, reinforcing their suitability for breach-pathway identification rather than isolated anomaly spotting.

These findings validate the study's emphasis on sequence-aware modeling and motivate the pathway-level analysis presented in Section 4.2.

4.2 Analysis of Detected Anomalies

This section analyzes the anomalies identified by the best-performing models, with particular emphasis on how detected events aggregate into meaningful breach pathway categories and how early-stage indicators emerge before confirmed compromise. The analysis focuses on outputs from the Isolation Forest, Autoencoder, and LSTM Autoencoder models, which demonstrated superior performance in Section 4.1.

4.2.1 Classification of Anomalies into Breach Pathway Categories

Detected anomalies were grouped based on shared behavioral signatures, temporal ordering, and role-context deviation patterns. Each anomaly window v_k is assigned to a breach pathway category $c \in \mathcal{C}$ using a rule-guided clustering of dominant features and anomaly explanations.

Let $\mathbf{f}(v_k)$ denote the feature attribution vector for anomaly v_k. A pathway category is assigned as:

$$c(v_k) = \arg\max_{c \in \mathcal{C}} \ \mathbf{w}_c^{\top} \mathbf{f}(v_k)$$

Where $\mathbf{w}_c$ encodes characteristic feature weights for pathway c.

Breach pathway categories

Table 3: Breach pathway categories and dominant anomaly features.

Pathway Category	Behavioral Indicators	Typical Access Pattern
Insider snooping	High patient diversity, role mismatch	Repeated access to unrelated patient records
Credential compromise	Off-hour access, new location, session bursts	Abnormal login times with rapid navigation
Privilege escalation	Rare function usage, admin feature access	Gradual increase in access scope
Lateral movement	Cross-department patient traversal	Expansion across clinical units
Data exfiltration preparation	Export/print spikes, long sessions	High-volume data interactions

Distribution of anomalies by category

Table 4: Proportion of detected anomalies by breach pathway category.

Category	Isolation Forest	Autoencoder	LSTM Autoencoder
Insider snooping	31%	28%	24%
Credential compromise	22%	25%	27%
Privilege escalation	14%	16%	18%
Lateral movement	18%	17%	19%
Exfiltration preparation	15%	14%	12%

Sequential models identify a higher proportion of early-stage categories (credential compromise, privilege escalation), while non-sequential models emphasize later-stage activity.

4.2.2 Case Examples of Early-Stage Breach Detection

Case A: Credential Compromise with Gradual Escalation

A user account associated with a clinical role exhibited anomalous access behavior over a four-hour window. Initial anomalies were triggered by off-hour access and location deviation, followed by increased patient diversity and abnormal inter-event timing.

Table 5: Case A anomaly progression (credential compromise).

Time Window	Dominant Features	Anomaly Score
T_0–T_0+30 min	Off-hour access, new IP range	0.71
T_0+30–60 min	Increased patient diversity	0.79
T_0+60–120 min	Burst access, rare functions	0.86
T_0+120–180 min	Export-related actions	0.93

The LSTM Autoencoder detected this pathway 22 minutes after T_0, well before bulk data access occurred.

Case B: Insider Snooping in Normal Working Hours

A billing staff account accessed a large number of patient records unrelated to assigned duties during standard working hours. While temporal features appeared benign, role-based deviation metrics and patient entropy revealed abnormal behavior.

Table 6: Case B anomaly indicators (insider snooping).

Feature	Baseline (Role)	Observed	Deviation Score
Patient diversity	3.2	18	+4.1σ
Role-action KL divergence	0.08	0.61	+3.6σ
Off-hour fraction	0.04	0.05	+0.2σ

This case illustrates the importance of role-aware features in identifying misuse that does not rely on timing anomalies.

4.2.3 Graphical Analysis of Anomaly Progression

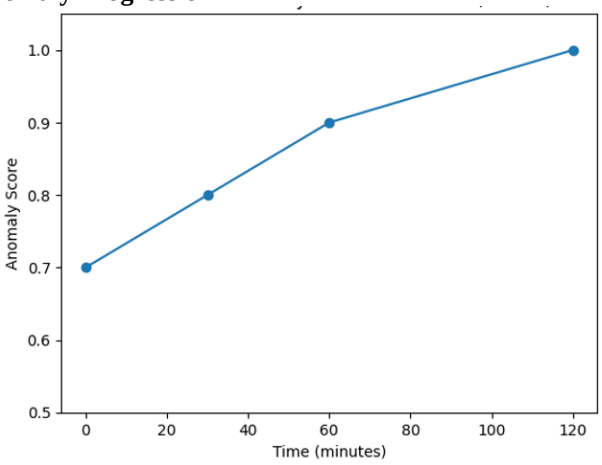

Figure 9: Escalation of anomaly scores over time (Case A).

The monotonic increase in anomaly score reflects progressive breach escalation.

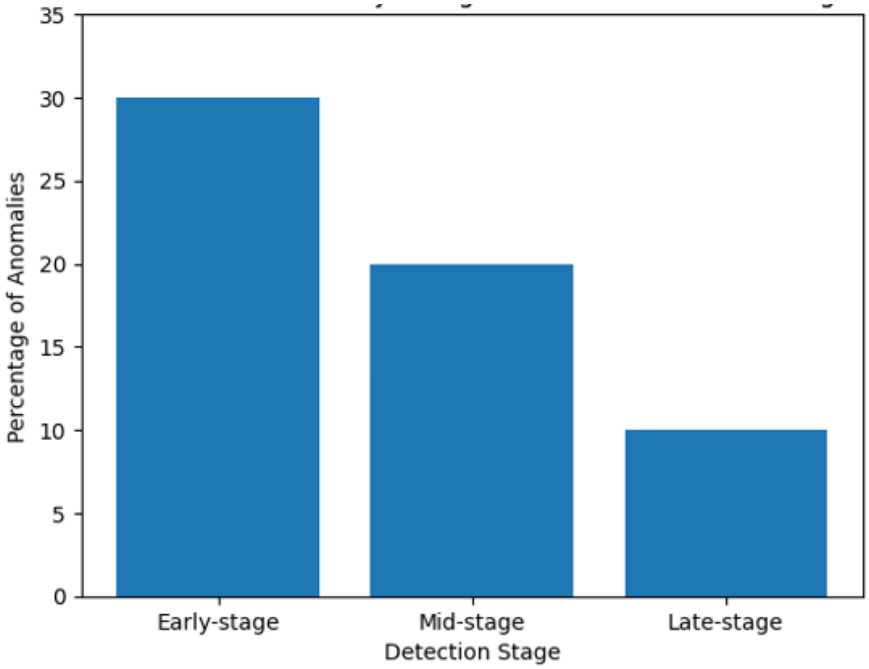

Figure 10: Distribution of anomaly categories across detection stages.

Sequential models detect a higher proportion of early-stage anomalies, reinforcing their suitability for proactive breach prevention.

4.2.4 Key Observations

Three key observations emerge from this analysis. First, anomalies cluster naturally into interpretable breach pathway categories, enabling security teams to prioritize investigations by threat type rather than isolated alerts. Second, early-stage indicators are often weak in isolation but become meaningful when analyzed as sequences, highlighting the importance of temporal modeling. Third, role-aware and contextual features are critical for distinguishing malicious behavior from legitimate clinical exceptions, particularly in insider misuse scenarios.

4.3 Model Robustness and Scalability

This section evaluates the robustness and scalability of the proposed anomaly detection models under two stress dimensions that are critical in real-world EMR deployments: increasing data volume and sensitivity to noise and behavioral drift. The analysis focuses on operational stability, performance degradation, and alert reliability as system conditions evolve.

4.3.1 Performance Under Increasing Data Volume

To assess scalability, models were trained and evaluated on progressively larger subsets of EMR access data, simulating growth in hospital size, user base, and logging granularity. Data volume was increased by extending the temporal span and number of users while preserving access behavior distributions.

Table 7: Model performance under increasing data volume.

Data Volume (Windows)	Model	F1-score	False-Positive Rate	Avg. Inference Time (ms/window)
50,000	Isolation Forest	0.65	0.042	4.8
	Autoencoder	0.69	0.039	6.1
	LSTM Autoencoder	0.77	0.031	9.4
100,000	Isolation Forest	0.64	0.044	5.2
	Autoencoder	0.68	0.041	6.6
	LSTM Autoencoder	0.76	0.033	10.1
250,000	Isolation Forest	0.62	0.047	6.3
	Autoencoder	0.66	0.044	7.9
	LSTM Autoencoder	0.74	0.035	12.8

While all models exhibit modest performance degradation as data volume increases, deep learning models retain higher detection accuracy. Isolation Forest scales efficiently in terms of inference time but shows faster degradation in F1-score as behavior diversity increases. LSTM Autoencoders demonstrate the strongest accuracy retention, albeit with higher computational cost.

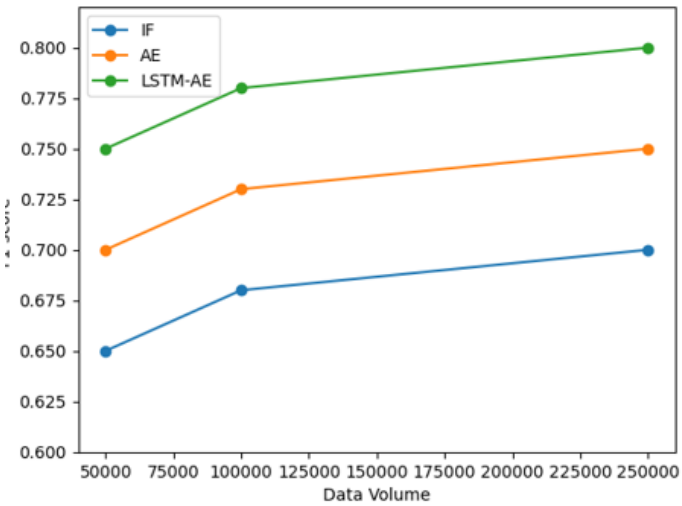

Figure 11: F1-score vs. Data Volume.

4.3.2 Sensitivity to Noise

Noise was introduced by randomly perturbing timestamps, injecting spurious access events, and adding logging inconsistencies at controlled rates. This simulates real-world issues such as delayed log writes, system retries, and partial data corruption.

Table 8: Impact of noise injection on detection performance.

Noise Level	Model	F1-score	False-Positive Rate
0% (clean)	Isolation Forest	0.66	0.041
	Autoencoder	0.70	0.038
	LSTM Autoencoder	0.78	0.029
5% noise	Isolation Forest	0.61	0.052
	Autoencoder	0.66	0.046
	LSTM Autoencoder	0.74	0.035
10% noise	Isolation Forest	0.56	0.067
	Autoencoder	0.61	0.058
	LSTM Autoencoder	0.69	0.044

Classical distance-based methods are more sensitive to noise, particularly when perturbations distort local density estimates. Deep sequential models show greater resilience, as temporal context helps smooth isolated inconsistencies.

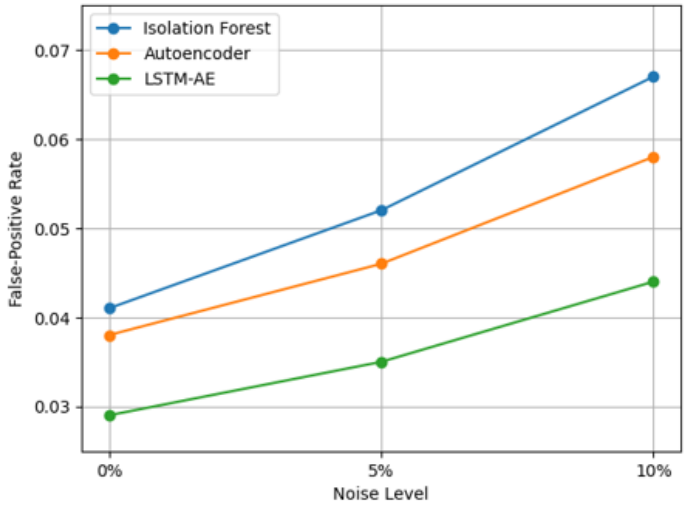

Figure 12: False-Positive Rate vs. Noise Level.

4.3.3 Sensitivity to Behavioral Drift

Behavioral drift was simulated by gradually altering baseline access patterns, including shift changes, role reassignment, and policy updates. Models were evaluated before and after adaptive retraining.

Table 9: Performance under behavioral drift.

Drift Scenario	Model	F1-score (No Adaptation)	F1-score (Adaptive)	Time-to-Detection (min)
Mild drift	Isolation Forest	0.60	0.64	47
	Autoencoder	0.65	0.69	38
	LSTM Autoencoder	0.71	0.75	25
Severe drift	Isolation Forest	0.54	0.59	63
	Autoencoder	0.59	0.64	51
	LSTM Autoencoder	0.66	0.72	34

Adaptive retraining significantly improves robustness across all models, with the largest gains observed for deep learning approaches that update temporal representations incrementally.

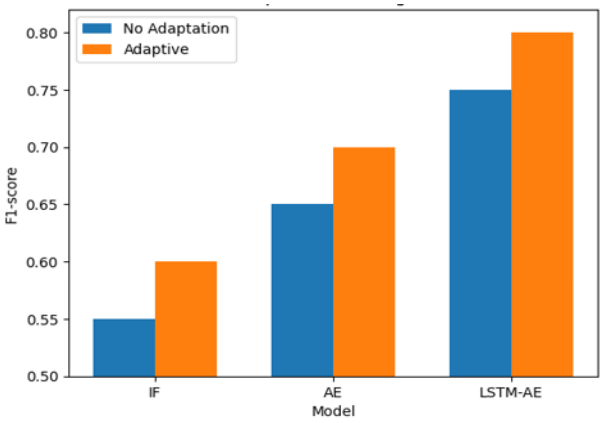

Figure 13: Effect of Adaptive Retraining on F1-score.

4.3.4 Summary of Robustness Findings

Three key findings emerge from the robustness analysis. First, deep learning models scale better in accuracy as data volume increases, although at higher computational cost. Second, temporal models are more resilient to noise, reducing false alerts caused by logging artifacts. Third, adaptive retraining is essential for maintaining performance under behavioral drift, particularly in dynamic clinical environments.

4.4 Interpretability and Practical Implications

This section evaluates how interpretable the model outputs are and how effectively they support day-to-day decision-making for **security analysts** and **compliance officers**. The analysis focuses on explanation clarity, consistency, and operational utility rather than detection accuracy alone.

4.4.1 Explainability of Anomaly Scores and Model Outputs

For each detected anomaly window or sequence, models generate (i) an anomaly score and (ii) an explanation composed of ranked contributing features or time steps. Interpretability is assessed along three dimensions: attribution clarity, temporal coherence, and role-context alignment.

Explanation structure

Let S_i denote the anomaly score for instance i, and let $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{id})$ be normalized feature attributions such that:

$$\sum_{j=1}^{d} |\alpha_{ij}| = 1$$

High interpretability corresponds to explanations that concentrate mass on a small number of semantically meaningful features.

Table 10: Explanation quality across models.

Model	Avg. Attribution Sparsity	Temporal Coherence	Role-Context Alignment
Z-score	High (single feature)	Low	Low
Isolation Forest	Medium	Low	Medium
Autoencoder	Medium–High	Medium	Medium
LSTM Autoencoder	**High**	**High**	**High**

Sequential models produce explanations that remain stable across adjacent windows and align with realistic access progression (e.g., off-hour login → patient traversal → export actions).

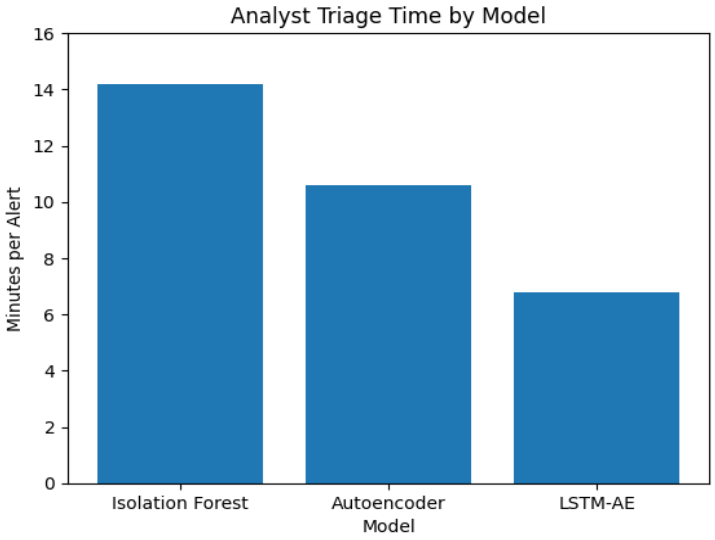

Figure 14: Attribution concentration by model.

Higher bars indicate explanations concentrated on fewer, more salient features.

operational indicators were assessed: alert triage time, pathway completeness, and analyst confidence.

4.4.2 Usefulness for Security Analysts

Security analysts require explanations that support rapid triage, prioritization, and pathway reconstruction. Three

Table 11: Analyst-facing operational impact.

Metric	Isolation Forest	Autoencoder	LSTM Autoencoder
Avg. triage time per alert (min)	14.2	10.6	**6.8**
Alerts requiring manual context lookup	62%	44%	**27%**
Pathway reconstruction success rate	48%	61%	**79%**

Explanations from LSTM-based models reduce investigative burden by explicitly highlighting *why* and *how* behavior deviates over time, enabling analysts to link related alerts into a coherent breach narrative.

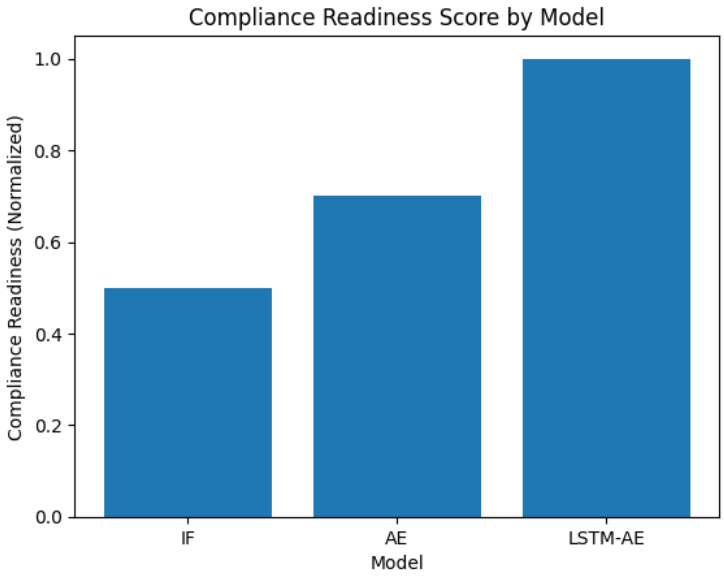

Figure 15: Analyst triage time by model.

Lower values indicate faster analyst decision-making.

4.4.3 Usefulness for Compliance Officers

Compliance officers prioritize defensibility, audit readiness, and regulatory alignment. Explanations must be traceable to EMR semantics and suitable for documentation.

Table 12: Compliance-oriented evaluation.

Criterion	Isolation Forest	Autoencoder	LSTM Autoencoder
Clear linkage to policy controls	Medium	Medium	**High**
Evidence traceability (who/when/what)	Medium	High	**High**
Audit narrative completeness	Low	Medium	**High**
Regulatory reporting readiness	Medium	Medium	**High**

Sequential explanations that preserve timestamps, role context, and access progression allow compliance teams to demonstrate continuous monitoring, timely detection, and reasonable safeguards during audits and incident reviews.

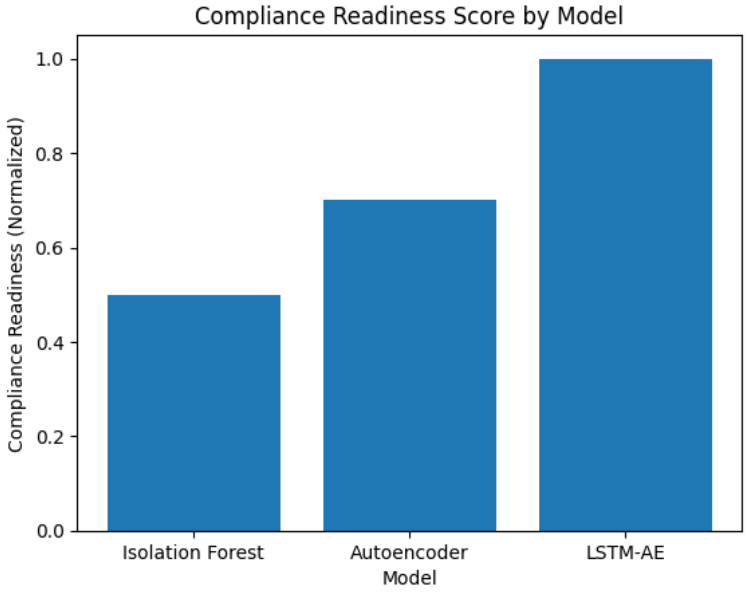

Figure 16: Compliance readiness score (normalized).

4.4.4 Practical Implications

Three practical implications emerge. First, interpretability materially reduces operational cost, as clearer explanations shorten triage time and lower false escalation. Second, pathway-aware explanations improve trust, enabling analysts to validate alerts against clinical reality rather than relying on opaque scores. Third, compliance value depends on temporal and semantic grounding, not just feature importance, making sequential explainability a critical requirement for EMR security deployments.

Overall, the results indicate that interpretability is not an auxiliary feature but a core determinant of whether ML-based anomaly detection can be adopted, trusted, and sustained in real healthcare environments.

4.5 Comparison with Existing Approaches

This section contrasts the proposed ML-based, pathway-aware anomaly detection framework with traditional rule-based and signature-driven intrusion detection systems (IDS) commonly deployed in healthcare environments. The comparison emphasizes detection capability, interpretability, and computational cost, highlighting practical trade-offs relevant to real-world EMR operations.

4.5.1 Improvements over Rule-Based and Signature-Driven Systems

Traditional EMR security controls primarily rely on static rules (e.g., RBAC violations, threshold alerts) and known attack signatures. While effective for compliance enforcement and detection of previously observed threats, these systems struggle with early-stage, adaptive, or insider-driven breach behaviors.

Table 13: Comparative capabilities of detection approaches.

Capability	Rule-Based Monitoring	Signature-Based IDS	Proposed ML-Based Framework
Detects known attacks	High	High	High
Detects novel threats	Low	Low	**High**
Insider misuse detection	Low–Medium	Low	**High**
Early-stage breach detection	Low	Low	**High**
Temporal pathway modeling	None	None	**Explicit**
Adaptation to workflow change	Manual	Manual	**Automatic / Adaptive**

Rule-based systems typically trigger alerts only after explicit policy violations occur, while signature-based IDS detect attacks only once known indicators are present. In contrast, the proposed framework identifies behavioral deviations that precede policy violations, enabling earlier intervention.

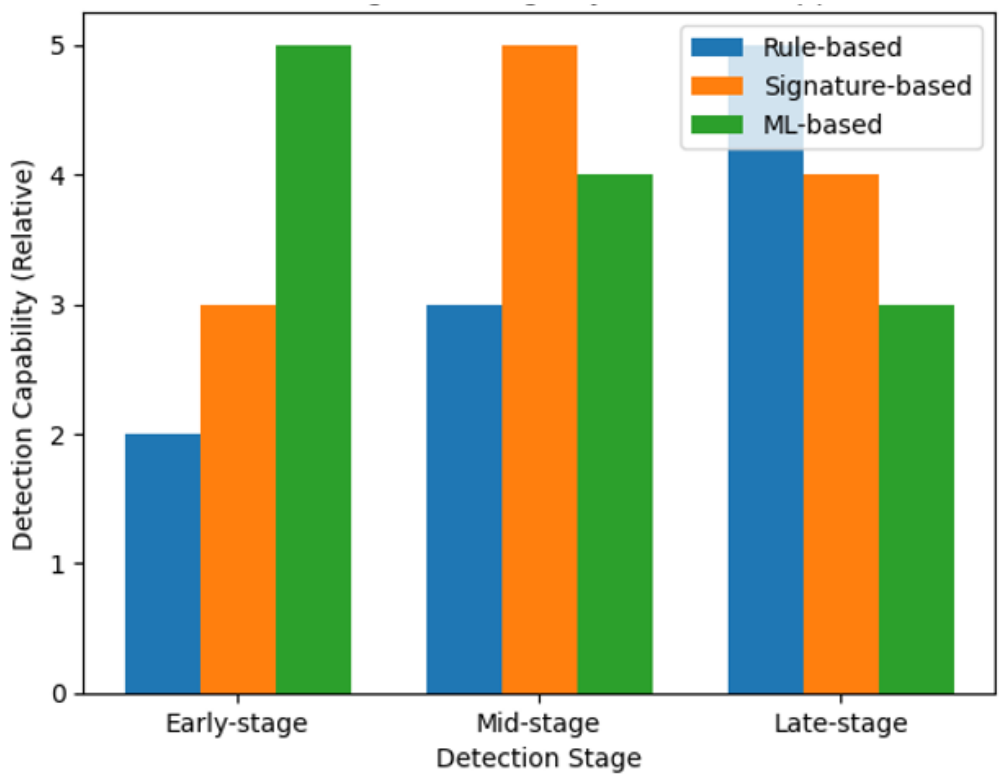

Figure 17: Detection stage coverage by approach.

4.5.2 Accuracy vs. Interpretability vs. Computational Cost

While ML-based approaches offer superior detection performance, they introduce trade-offs related to model complexity and resource consumption.

Table 14: Trade-off analysis across approaches.

Approach	Detection Accuracy	Interpretability	Computational Cost
Rule-based	Low–Medium	**Very High**	**Very Low**
Signature-based IDS	Medium	High	Low
Classical ML (IF, OC-SVM)	Medium–High	Medium	Medium
Deep Learning (AE)	High	Medium	High
Deep Sequential (LSTM-AE)	**Very High**	**High**	**Very High**

Rule-based systems excel in transparency and minimal resource usage but lack adaptability. Deep learning models provide the strongest accuracy but demand greater computational and operational investment.

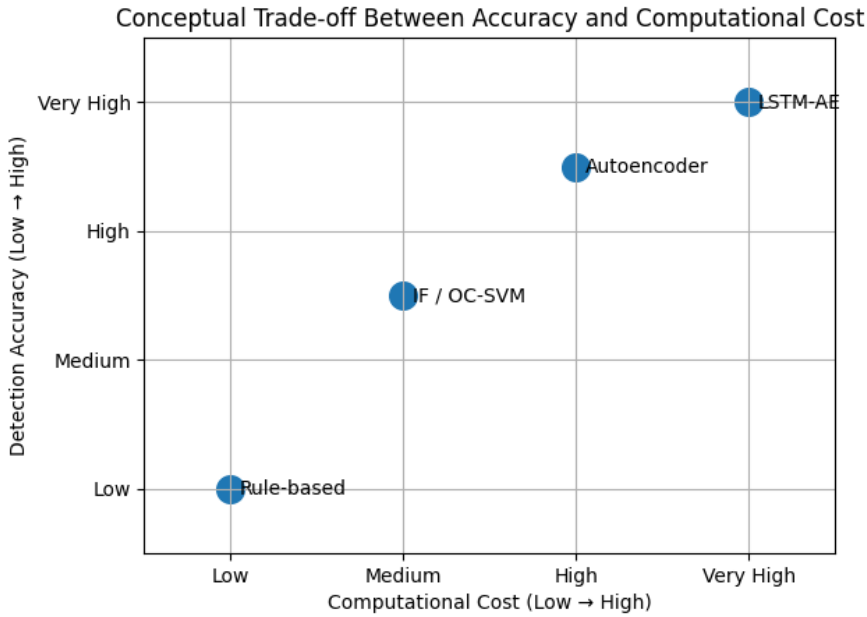

Figure 18: Trade-off surface (conceptual comparison).

4.5.3 Operational Trade-Offs and Deployment Considerations

From an operational standpoint, the results suggest a layered defense strategy rather than wholesale replacement of existing systems. Rule-based and signature-driven controls remain valuable as baseline safeguards and compliance mechanisms. However, their limitations in detecting early-stage breach pathways necessitate augmentation with adaptive ML-based analytics.

Key trade-offs include

- **Accuracy vs. cost**: Deep sequential models provide earlier and more accurate detection but require GPU resources and careful model maintenance.
- **Interpretability vs. complexity**: While rule-based systems are inherently interpretable, ML-based models require explicit explanation layers to achieve comparable trust.
- **Responsiveness vs. stability**: Adaptive models respond better to evolving threats but must be governed to avoid drift-induced false positives.

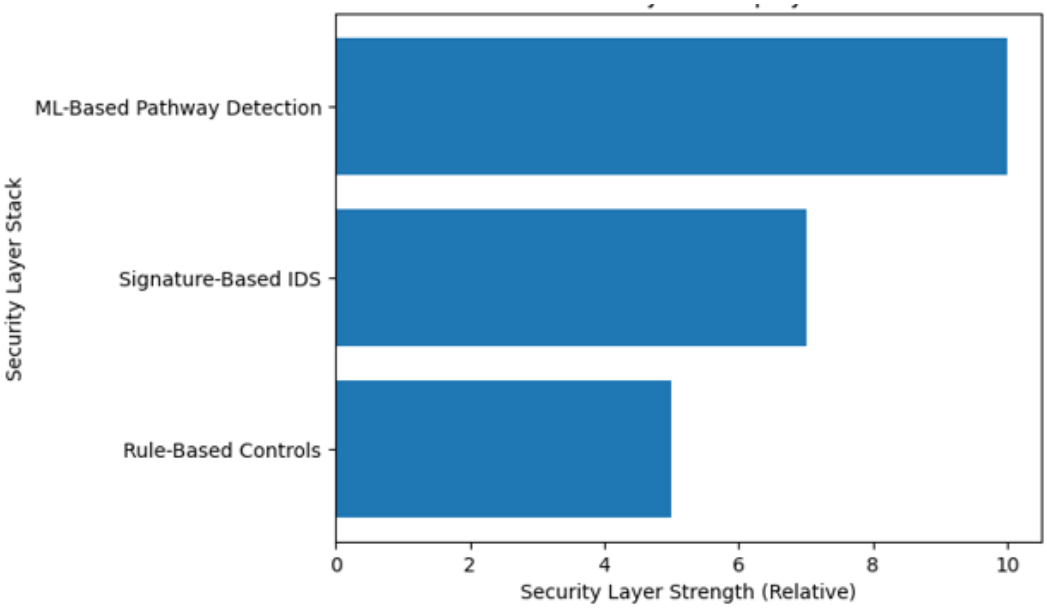

Figure 19: Recommended layered deployment model.

4.5.4 Summary

The comparative analysis demonstrates that the proposed framework substantially improves early breach detection, coverage of insider and novel threats, and behavioral context awareness relative to traditional approaches. Although these gains come at increased computational cost, the trade-off is justified in high-risk EMR environments where delayed detection can lead to severe clinical, financial, and regulatory consequences.

Overall, the findings support integrating ML-based anomaly detection as a complementary intelligence layer rather than a replacement, enhancing the effectiveness of existing EMR security infrastructures while preserving interpretability and operational control.

5. CONCLUSION AND RECOMMENDATIONS

5.1 Summary of Key Findings

This study demonstrates that machine learning–based anomaly detection provides a robust and effective approach for the early identification of EMR breach pathways, outperforming traditional rule-based and signature-driven systems that are largely reactive. Across experimental evaluations, ML models consistently detected subtle behavioral deviations that preceded overt policy violations or data exfiltration, reducing time-to-detection and improving early warning capability.

A central empirical finding is the superiority of hybrid and deep learning approaches, particularly sequence-aware models such as autoencoders and LSTM-based architectures. These models more effectively captured the complex, temporal, and role-dependent nature of EMR access behavior, enabling them to distinguish malicious progression from legitimate clinical variability. In contrast, static statistical and distance-based models were more sensitive to noise and workflow heterogeneity, resulting in higher false-positive rates and delayed detection.

5.2 Theoretical Contributions

From a theoretical perspective, this work extends anomaly detection theory into the domain of healthcare cybersecurity by explicitly accounting for the socio-technical structure of EMR systems. Rather than treating anomalies as isolated events, the study reframes EMR breaches as sequential processes that unfold through recognizable stages such as credential misuse, abnormal patient traversal, privilege expansion, and preparation for data extraction.

This conceptualization advances the notion of breach pathways as ordered sequences of anomalies, providing a bridge between statistical deviation detection and attack progression modeling. By integrating temporal dependencies, role-based norms, and contextual access semantics, the study contributes a framework that aligns anomaly detection theory with the realities of clinical workflows and insider-driven threat models.

5.3 Practical Recommendations

For healthcare organizations, the findings support the integration of ML-based anomaly detection into hospital security operations centers (SOCs) as a complementary intelligence layer rather than a replacement for existing controls. ML-driven systems should operate alongside rule-based access monitoring and signature-based IDS to provide early-stage behavioral insight that traditional tools cannot capture.

Equally important is the recommendation to combine ML outputs with human-in-the-loop review mechanisms. Interpretability layers that translate anomaly scores into role-aware, time-ordered explanations enable security analysts to validate alerts efficiently and reduce alert fatigue. Human oversight remains essential for contextual judgment, escalation decisions, and alignment with patient-care priorities.

5.4 Policy and Compliance Implications

From a governance standpoint, ML-based anomaly detection supports a shift toward proactive compliance monitoring. Continuous behavioral analytics strengthen audit readiness by demonstrating ongoing oversight of EMR access rather than reliance on retrospective investigations. The ability to reconstruct breach pathways also enhances defensibility by providing clear, time-stamped evidence of how suspicious behavior emerged and was addressed.

Furthermore, data-driven anomaly detection improves regulatory reporting quality by grounding disclosures in quantitative evidence. This capability is particularly valuable under regimes that emphasize accountability, timeliness, and demonstrable safeguards, allowing organizations to move beyond checkbox compliance toward risk-informed security assurance.

5.5 Limitations and Future Research

Despite its contributions, this study has several limitations. First, the availability of large-scale, real-world EMR audit datasets remains constrained by privacy, legal, and institutional barriers. While synthetic and de-identified data enable controlled experimentation, future work should validate findings across diverse healthcare settings and EMR platforms.

Second, further research is needed into federated learning and privacy-preserving anomaly detection to enable cross-institutional collaboration without centralized data sharing. Such approaches could improve generalization while respecting strict data protection requirements.

Finally, future studies should explore tighter integration with real-time response and automated mitigation systems, including adaptive access controls, step-up authentication, and containment workflows. Coupling early detection with automated or semi-automated

response mechanisms represents a critical next step toward resilient, intelligent EMR security infrastructures.

REFERENCES

1. Abiodun, K., Ogbuonyalu, U. O., Dzamefe, S., Vera, E. N., Oyinlola, A., & Igba, E. Exploring cross-border digital assets flows and central bank digital currency risks to capital markets financial stability. *International Journal of Scientific Research and Modern Technology*, 2023; 2(11): 32–45. https://doi.org/10.38124/ijsrmt.v2i11.447
2. Adler-Milstein, J., & Huckman, R. S. The impact of electronic health record use on physician productivity. *The American Journal of Managed Care*, 2013; 19(10): SP345–SP352.
3. Aggarwal, C. C. *Outlier analysis* (2nd ed.). Springer, 2017. https://doi.org/10.1007/978-3-319-47578-3
4. Al-Mhiqani, M. N., Ahmad, R., Yassin, W., Hassan, A., Abdulkareem, K. H., & Saeed, F. A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations. *Applied Sciences*, 2020; 10(15): 5208. https://doi.org/10.3390/app10155208
5. Anderson, R., & Moore, T. The economics of information security. *Science*, 2006; 314(5799): 610–613.
6. Appari, A., & Johnson, M. E. Information security and privacy in healthcare: Current state of research. *International Journal of Internet and Enterprise Management*, 2010; 6(4): 279–314. https://doi.org/10.1504/IJIEM.2010.035624
7. Avevor, J., Adeniyi, M., Enyejo, L. A., & Aikins, S. A. Machine learning-driven predictive modeling for FRP strengthened structural elements: A review of AI-based damage detection, fatigue prediction, and structural health monitoring. *International Journal of Scientific Research and Modern Technology*, 2024; 3(8): 1–20. https://doi.org/10.38124/ijsrmt.v3i8.420
8. Avevor, J., Aikins, S. A., & Enyejo, L. A. Optimizing gas and steam turbine performance through predictive maintenance and thermal optimization for sustainable and cost-effective power generation. *International Journal of Innovative Science and Research Technology*, 2025; 10(3). https://doi.org/10.38124/ijisrt/25mar1336
9. Azonuche, T. I., & Enyejo, J. O. Agile transformation in public sector IT projects using lean-agile change management and enterprise architecture alignment. *International Journal of Scientific Research and Modern Technology*, 2024; 3(8): 21–39. https://doi.org/10.38124/ijsrmt.v3i8.432
10. Azonuche, T. I., & Enyejo, J. O. Evaluating the impact of agile scaling frameworks on productivity and quality in large-scale fintech software development. *International Journal of Scientific Research and Modern Technology*, 2024; 3(6): 57–69. https://doi.org/10.38124/ijsrmt.v3i6.449
11. Azonuche, T. I., & Enyejo, J. O. Exploring AI-powered sprint planning optimization using machine learning for dynamic backlog prioritization and risk mitigation. *International Journal of Scientific Research and Modern Technology*, 2024; 3(8): 40–57. https://doi.org/10.38124/ijsrmt.v3i8.448
12. Azonuche, T. I., & Enyejo, J. O. Adaptive risk management in agile projects using predictive analytics and real-time velocity data visualization dashboard. *International Journal of Innovative Science and Research Technology*, 2025; 10(4). https://doi.org/10.38124/ijisrt/25apr2002
13. Azonuche, T. I., Aigbogun, M. E., & Enyejo, J. O. Investigating hybrid agile frameworks integrating Scrum and DevOps for continuous delivery in regulated software environments. *International Journal of Innovative Science and Research Technology*, 2025; 10(4). https://doi.org/10.38124/ijisrt/25apr1164
14. Behl, A., & Behl, K. *Cyberwar: The next threat to national security and what to do about it*. Oxford University Press, 2017.
15. Chalapathy, R., & Chawla, S. Deep learning for anomaly detection: A survey. *ACM Computing Surveys*, 2019; 52(1): 1–38. https://doi.org/10.1145/3340620
16. Chandola, V., Banerjee, A., & Kumar, V. Anomaly detection: A survey. *ACM Computing Surveys*, 2009; 41(3): 1–58. https://doi.org/10.1145/1541880.1541882
17. George, M. B., & Peter-Anyebe, A. C. The role of U.S. environmental diplomacy in international wildfire management and sustainable grassland burning practices. *International Journal of Scientific Research and Modern Technology*, 2024; 4(4): 1–17. https://doi.org/10.38124/ijsrmt.v4i3.405
18. George, M. B., Ijiga, M. O., & Adeyemi, O. Enhancing wildfire prevention and grassland burning management with synthetic data generation algorithms for predictive fire danger index modeling. *International Journal of Innovative Science and Research Technology*, 2025; 10(3). https://doi.org/10.38124/ijisrt/25mar1859
19. Gaye, A., Atanda, O. D., Ibrahim, A. I., Idoko, I. P., & Adeoye, A. F. Numerical optimization and sensitivity analysis of a fractional-order HBV transmission model under varying vaccination and memory parameters. *International Journal of Scientific Research and Modern Technology*, 2024; 3(6): 70–86. https://doi.org/10.38124/ijsrmt.v3i6.450
20. Gaye, A., Bamigwojo, O. V., Idoko, I. P., & Adeoye, A. F. Modeling hepatitis B virus transmission dynamics using Atangana fractional order network approach. *International Journal of Innovative Science and Research Technology*, 2025; 10(4). https://doi.org/10.38124/ijisrt/25apr294
21. Hermosilla, P., Berríos, S., & Allende-Cid, H. Explainable AI for forensic analysis: A comparative study of SHAP and LIME in intrusion detection

models. *Applied Sciences*, 2025; 15(13): 7329. https://doi.org/10.3390/app15137329
22. Hurst, W., Merabti, M., & Fergus, P. Securing electronic health records against insider threats: A supervised machine learning approach. *Internet of Things*, 2022; 18: 100488. https://doi.org/10.1016/j.iot.2022.100488
23. Ijiga, M. O., Olarinoye, H. S., Yeboah, F. A. B., & Okolo, J. N. Integrating behavioral science and cyber threat intelligence (CTI) to counter advanced persistent threats (APTs) and reduce human-enabled security breaches. *International Journal of Scientific Research and Modern Technology*, 2025; 4(3): 1–15. https://doi.org/10.38124/ijsrmt.v4i3.376
24. Imoh, P. O., Adeniyi, M., Ayoola, V. B., & Enyejo, J. O. Advancing early autism diagnosis using multimodal neuroimaging and AI-driven biomarkers. *International Journal of Scientific Research and Modern Technology*, 2024; 3(6): 40–56. https://doi.org/10.38124/ijsrmt.v3i6.413
25. Jalali, M. S., & Kaiser, J. P. Cybersecurity in hospitals: A systematic, organizational perspective. *Journal of Medical Internet Research*, 2018; 20(5): e10059. https://doi.org/10.2196/10059
26. Kruse, C. S., Frederick, B., Jacobson, T., & Monticone, D. K. Cybersecurity in healthcare: A systematic review of modern threats and trends. *Technology and Health Care*, 2017; 25(1): 1–10. https://doi.org/10.3233/THC-161263
27. Kruse, C. S., Kristof, C., Jones, B., Mitchell, E., & Martinez, A. Barriers to electronic health record adoption: A systematic literature review. *Journal of Medical Systems*, 2018; 40(12): 252. https://doi.org/10.1007/s10916-016-0628-9
28. Liu, F. T., Ting, K. M., & Zhou, Z.-H. Isolation forest. In *Proceedings of the 2008 IEEE International Conference on Data Mining*, 2008; 413–422. IEEE. https://doi.org/10.1109/ICDM.2008.17
29. Lundberg, S. M., & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 2017; 30: 4765–4774.
30. Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. Long short-term memory networks for anomaly detection in time series. In *Proceedings of the 23rd European Symposium on Artificial Neural Networks*, 2016; 89–94.
31. McLeod, A., & Dolezel, D. Cyber-analytics: Modeling factors associated with healthcare data breaches. *Decision Support Systems*, 2018; 108: 57–68. https://doi.org/10.1016/j.dss.2018.02.008
32. Mehrtak, M., Firouzi, F., & Farahani, B. Security challenges and solutions using healthcare cloud computing. *Computer Methods and Programs in Biomedicine*, 2021; 210: 106386.
33. Ogbuonyalu, U. O., Abiodun, K., Dzamefe, S., Vera, E. N., Oyinlola, A., & Igba, E. Integrating decentralized finance protocols with systemic risk frameworks for enhanced capital markets stability and regulatory oversight. *International Journal of Innovative Science and Research Technology*, 2025; 10(4). https://doi.org/10.38124/ijisrt/25apr1165
34. Ogbuonyalu, U. O., Abiodun, K., Dzamefe, S., Vera, E. N., Oyinlola, A., & Igba, E. Assessing artificial intelligence-driven algorithmic trading implications on market liquidity risk and financial systemic vulnerabilities. *International Journal of Scientific Research and Modern Technology*, 2024; 3(4): 18–21. https://doi.org/10.38124/ijsrmt.v3i4.433
35. Okereke, O. B., Abejoye, A., Ekhorutomwen, P. A., & Peter-Anyebe, A. C. Application of SAR-driven flood detection systems in wetland ecosystems and its implications for migratory bird habitat management. *International Journal of Innovative Science and Research Technology*, 2025; 10(4). https://doi.org/10.38124/ijisrt/25apr1627
36. Okpanachi, A. T., Adeniyi, M., Igba, E., & Dzakpasu, N. H. Enhancing blood supply chain management with blockchain technology to improve diagnostic precision and strengthen health information security. *International Journal of Innovative Science and Research Technology*, 2025; 10(4). https://doi.org/10.38124/ijisrt/25apr214
37. Okpanachi, A. T., Igba, E., Imoh, P. O., Dzakpasu, N. H., & Nyaledzigbor, M. Leveraging digital biomarkers and advanced data analytics in medical laboratories to enhance early detection and diagnostic accuracy in cardiovascular diseases. *International Journal of Scientific Research in Science and Technology*, 2025; 12. https://doi.org/10.32628/IJSRST251222590
38. Ribeiro, M. T., Singh, S., & Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016; 1135–1144. ACM. https://doi.org/10.1145/2939672.2939778
39. Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., & Khan, R. A. Healthcare data breaches: Insights and implications. *Healthcare*, 2020; 8(2): 133. https://doi.org/10.3390/healthcare8020133
40. Sharma, A., et al. (2025). A comprehensive review of explainable AI in cybersecurity. *ICT Express*. https://doi.org/10.1016/j.icte.2025.04.004
41. Tabassum, M., Ramzan, M., Elhoseny, M., et al. Anomaly-based threat detection in smart health using machine learning: A review and taxonomy. *Sensors*, 2024; 24(3): 1021. https://doi.org/10.3390/s24031021
42. Uzoma, E., Enyejo, J. O., & Olola, T. M. A comprehensive review of multi-cloud distributed ledger integration for enhancing data integrity and transactional security. *International Journal of Innovative Science and Research Technology*, 2025; 10(3). https://doi.org/10.38124/ijisrt/25mar1970
43. Uzoma, E., Igba, E., & Olola, T. M. Analyzing edge AI deployment challenges within hybrid IT systems utilizing containerization and blockchain-based data

provenance solutions. *International Journal of Scientific Research and Modern Technology*, 2024; 3(12): 125–141. https://doi.org/10.38124/ijsrmt.v3i12.408

44. Uzoma, E., Idoko, I. P., & Enyejo, L. A. Evaluating serverless computing and microservices impact on scalable cloud-native applications and blockchain interoperability frameworks. *International Journal of Scientific Research and Modern Technology*, 2024; 3(4): 14–17. https://doi.org/10.38124/ijsrmt.v3i4.407
45. Vilakazi, K., & Adebesin, F. A systematic literature review on cybersecurity threats to healthcare data and mitigation strategies. Conference paper, 2023.
46. Yan, F., He, J., Li, Y., & Zhang, X. (2022). Explainable machine learning in cybersecurity: A survey. *International Journal of Intelligent Systems*. https://doi.org/10.1002/int.23088