An Integrated Machine Learning and Deep Learning Framework for Insider Threat Detection in Healthcare EHR Logs

OPEN ACCESS

Volume: 13

Special Issue: 3

Month: February

Year: 2026

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Citation:
Kingslin, Sumathy, and Thasleem R. "An Integrated Machine Learning and Deep Learning Framework for Insider Threat Detection in Healthcare EHR Logs." *Shanlax International Journal of Arts, Science and Humanities,* vol. 13, no. 3, 2026, pp. 377–84.

DOI:
https://doiorg/10.34293/sijash.v13iS3-i2-Feb.10308

Dr. Sumathy Kingslin
Associate Professor, Department of Computer Science
Quaid-E-Millath Govt College for Women (A), Chennai, India

Thasleem R
Research Scholar, Department of Computer Science
Quaid-E-Millath Govt College for Women (A), Chennai, India

Abstract

Insider threats pose a significant security challenge to healthcare organizations due to privileged access to sensitive Electronic Health Record (EHR) systems. While machine learning and deep learning techniques have been independently applied for insider threat detection, each approach has inherent limitations in modeling complex healthcare access behavior. This paper proposes an integrated hybrid machine learning–deep learning framework for insider threat detection in healthcare EHR logs. The framework combines traditional machine learning models for structured feature learning with Long Short-Term Memory (LSTM) networks for capturing temporal user behavior patterns. The proposed approach is evaluated using the CERT insider threat dataset and simulated healthcare EHR logs. Experimental results demonstrate that the hybrid model consistently outperforms standalone machine learning and deep learning models in terms of accuracy, precision, recall, F1-score, and AUC. The findings highlight the effectiveness of hybrid learning strategies in improving insider threat detection performance in dynamic healthcare environments.

Keywords: Insider Threat Detection, Healthcare Cybersecurity, Hybrid Learning, Machine Learning, Deep Learning, LSTM, Electronic Health Records

Background and Motivation

The widespread adoption of Electronic Health Record (EHR) systems has significantly improved healthcare data management and clinical decision-making. However, the sensitive nature of EHR data and the presence of privileged users make healthcare systems highly susceptible to insider threats [1], [2]. Insider threats arise when authorized users intentionally or unintentionally misuse their access privileges, leading to data breaches, privacy violations, and regulatory penalties.

Traditional security mechanisms and standalone detection models often struggle to identify insider threats in healthcare environments due to the similarity between malicious and legitimate clinical

activities [18]. This challenge necessitates intelligent detection mechanisms capable of modeling both static user behavior and temporal access patterns.

Recent studies have explored machine learning (ML) and deep learning (DL) techniques independently for insider threat detection [3], [4]. While ML models perform well on structured data, they are limited in capturing sequential behavior. Conversely, DL models such as Long Short-Term Memory (LSTM) networks effectively learn temporal dependencies but often lack interpretability and robustness when used alone [9]. These limitations motivate the integration of ML and DL approaches within a unified framework.

Review of Existing Detection Approaches

Research on insider threat detection has evolved from traditional rule-based systems to advanced machine learning and deep learning approaches. Existing methods can be broadly categorized as follows.

Rule-Based and Policy-Driven Approaches

Early insider threat detection systems relied on predefined security rules and access control policies to identify violations such as unauthorized file access or policy breaches [1], [8].

- Easy to implement and interpret
- Unable to detect subtle or evolving insider behaviors
- High false alarm rates in dynamic environments

Due to these limitations, rule-based approaches are often ineffective in complex healthcare workflows, where legitimate access patterns frequently vary because of emergencies and role changes [18].

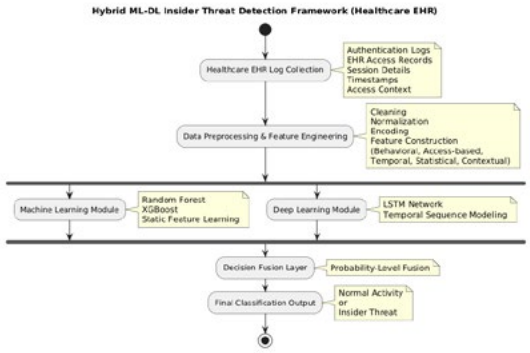

Fig. 1 Illustrates the processing flow of the proposed hybrid ML–DL architecture, where structured feature-based learning and temporal sequence modeling are integrated through a decision fusion layer to detect insider threats

Machine Learning–Based Approaches

Machine learning models have been widely applied to analyze structured system logs and user activity records for insider threat detection [3], [7]. Commonly used techniques include Random Forest, Support Vector Machine, and Gradient Boosting models.

- Effective for structured and tabular data
- Good interpretability and robustness
- Depend heavily on feature quality
- Limited capability to capture temporal behavior

In healthcare environments, ML models often struggle to distinguish malicious behavior from legitimate clinical access when activities closely resemble normal workflows [3], [18].

Deep Learning–Based Approaches

Deep learning techniques, particularly recurrent neural networks and Long Short-Term Memory (LSTM) models, have been introduced to capture sequential and temporal patterns in user behavior [4], [9].

- Strong temporal modeling capability
- Effective for sequential log data
- Require large training datasets
- Limited interpretability due to black-box nature

These limitations reduce the suitability of standalone deep learning models in healthcare security systems, where transparency and explainability are critical requirements [18].

Limitations of Existing Approaches in Healthcare

Despite significant progress, existing insider threat detection approaches face several challenges when applied to healthcare environments:

- Generic models fail to represent clinical workflows and role-based access behavior [18]
- Emergency access scenarios frequently trigger false alarms [2], [18]
- Standalone ML or DL models capture only partial behavior characteristics [3], [9]
- Limited emphasis on integrated or hybrid modeling strategies [14]

Motivation for Hybrid ML–DL Approach

The above limitations highlight the need for a hybrid insider threat detection framework that:

- Combines structured feature learning with temporal sequence modeling [4], [13]
- Reduces false positives in complex healthcare environments [18]
- Balances detection performance and interpretability [11], [14]
- Effectively captures both static and evolving insider behavior patterns [9]

This motivation forms the foundation for the proposed hybrid ML–DL framework for insider threat detection in healthcare EHR systems.

Proposed Integrated ML–DL Detection Framework

Framework Overview

This study introduces an integrated machine learning–deep learning (ML–DL) detection framework designed to enhance insider threat identification in healthcare EHR log data. The framework jointly exploits structured feature-based analysis and temporal sequence learning to capture diverse insider behavior patterns. The major components of the framework are described below.

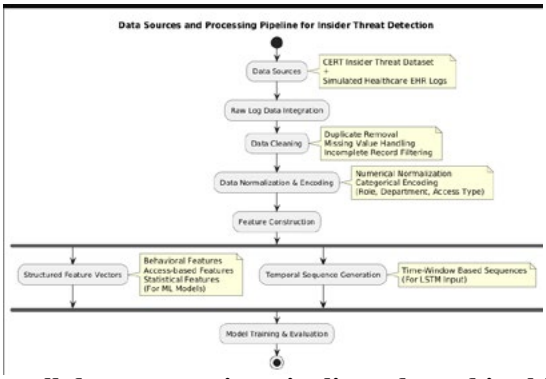

Fig. 2 illustrates the overall data processing pipeline adopted in this study, where raw log data from multiple sources undergo systematic preprocessing before being used for model training and evaluation

EHR Log Collection Module

This module gathers authentication records, EHR access logs, session activities, and timestamp information

Data Preprocessing and Feature Construction Module

Raw log data is processed through cleaning, normalization, and categorical encoding procedures, followed by the construction of temporal sequences required for deep learning models [14].

Machine Learning Behavior Analysis Module

Structured feature vectors are analyzed using machine learning classifiers such as Random Forest and XGBoost to model static behavioral characteristics of users [11], [13].

Deep Learning Temporal Modeling Module

An LSTM-based neural network is employed to learn sequential and temporal dependencies in user access behavior, enabling detection of evolving insider threat patterns [9].

Decision Fusion and Classification Module

Outputs generated by the ML and DL modules are combined at the probability level to produce the final classification decision, improving detection robustness.

Hybrid Learning Strategy

The proposed framework follows a parallel hybrid learning strategy, in which machine learning and deep learning models operate independently on the same preprocessed input data. Prediction probabilities from both models are aggregated to determine the final classification outcome. This strategy effectively integrates the interpretability and stability of machine learning models with the temporal modeling capability of deep learning architectures, resulting in improved insider threat detection performance in healthcare environments [4], [13].

Data Sources and Processing Pipeline

To ensure both reproducibility and healthcare-specific relevance, this study utilizes the CERT insider threat dataset in combination with simulated healthcare EHR log data [5], [6], [18]. The CERT dataset provides realistic insider behavior scenarios, while the simulated healthcare logs capture clinical workflows, role-based access patterns, and sensitive patient record interactions.

Table I. Dataset Statistics

Dataset	Total Records	Normal Instances	Insider Threat Instances
CERT Insider Threat Dataset	100,000	94,500	5,500
Simulated Healthcare Logs	60,000	55,200	4,800
Combined Dataset	160,000	149,700	10,300

Table I presents the statistical distribution of the CERT insider threat dataset and the simulated healthcare EHR log dataset used in this study.

Initially, data cleaning is performed to remove duplicate entries, handle missing values, and eliminate incomplete or corrupted records. Next, data normalization is applied to numerical attributes such as access counts and session duration to reduce scale-related bias. Categorical attributes, including user role, department, and access type, are then transformed into numerical representations using appropriate encoding techniques.

Following feature preparation, the processed data is organized into two parallel representations. The first representation consists of structured feature vectors used as input to machine learning models. The second representation involves time-window-based sequence construction, where user activity records are grouped into temporal sequences suitable for LSTM-based deep learning models [14].

This dual data preparation strategy enables the proposed hybrid ML–DL framework to simultaneously capture static behavioral characteristics and temporal access patterns, improving insider threat detection performance in healthcare EHR systems.

Experimental Design and Evaluation Metrics

This section describes the experimental design adopted to evaluate the effectiveness of the proposed integrated ML–DL framework for insider threat detection in healthcare EHR systems. The evaluation focuses on comparing the proposed hybrid model with a standalone deep learning approach to demonstrate the benefits of model integration.

Models Used for Comparison

To assess the effectiveness of the proposed framework, the following models are considered:

Random Forest (ML)

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and aggregates their predictions to produce the final classification result. By combining multiple weak learners, Random Forest reduces overfitting and improves generalization performance. It is particularly effective for handling high-dimensional and structured data, making it suitable for insider threat detection based on engineered behavioral and access-pattern features [11]. In healthcare security applications, Random Forest offers a good balance between detection accuracy and interpretability, which is essential for understanding security decisions.

XGBoost (ML)

XGBoost is an optimized gradient boosting framework that builds decision trees sequentially, where each new tree attempts to correct the errors of the previous ones. It is known for its high predictive accuracy, efficient computation, and ability to model complex non-linear relationships between features [12]. XGBoost has been widely adopted in cybersecurity tasks due to its robustness against noisy data and its capability to capture subtle behavior patterns. In this study, XGBoost is used to evaluate the effectiveness of advanced ensemble learning for insider threat detection in healthcare EHR environments.

Long Short-Term Memory (LSTM) Model

The LSTM model is used as a baseline deep learning approach due to its ability to capture temporal and sequential patterns in user activity data. LSTM networks have been widely applied in insider threat detection for modeling time-dependent behavior; however, when used independently, they may struggle to distinguish subtle malicious actions from legitimate clinical access patterns in healthcare environments [9].

Proposed Integrated ML–DL Model

The proposed hybrid model combines traditional machine learning and deep learning techniques within a unified framework. While the deep learning component (LSTM) captures temporal dependencies in user behavior, the machine learning component learns structured behavioral patterns from engineered features. The outputs of both models are fused at the decision level to generate the final classification. This integration enables the framework to leverage the strengths of both approaches, improving detection accuracy and robustness [4], [12].

Evaluation Metrics

Model performance is evaluated using widely adopted classification metrics that are particularly suitable for imbalanced datasets, such as insider threat detection scenarios in healthcare systems [15].

- **Accuracy**:

  Measures the overall proportion of correctly classified instances. While useful, accuracy alone may be misleading in imbalanced datasets.
- **Precision**:

  Indicates the proportion of correctly identified insider threat instances among all predicted threats, reflecting the reliability of positive predictions.
- **Recall:**

  Measures the model's ability to correctly identify actual insider threats. High recall is critical in healthcare security to minimize undetected malicious activity.
- **F1-score:**

  Represents the harmonic mean of precision and recall, providing a balanced performance measure that is particularly effective for imbalanced data.
- **Area Under the ROC Curve (AUC):**

  Evaluates the model's ability to distinguish between normal and malicious behavior across varying classification thresholds.

Together, these metrics provide a comprehensive assessment of the detection capability, robustness, and practical suitability of the proposed hybrid ML–DL framework for healthcare insider threat detection.

Performance Analysis and Discussion

Experimental results demonstrate that the proposed integrated ML–DL framework consistently outperforms standalone machine learning and deep learning models across all evaluation metrics. In particular, the hybrid model achieves significantly higher recall and F1-score, indicating its superior ability to detect insider threat instances that exhibit subtle or gradual behavioral deviations. Such behaviors are often overlooked by individual ML or DL models when used in isolation, especially in highly dynamic healthcare environments [3], [9].

Table II. compares the detection performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Random Forest	95.6	88.3	85.1	86.7	0.93
XGBoost	96.8	90.7	88.9	89.8	0.95
LSTM	94.3	86.1	83.7	84.9	0.92
Proposed Hybrid ML–DL	97.9	93.2	91.5	92.3	0.97

Table II compares the detection performance of standalone machine learning, deep learning, and the proposed hybrid ML–DL models across standard evaluation metrics.

The improved performance can be attributed to the complementary strengths of the hybrid framework. While machine learning models effectively capture structured and role-based behavioral patterns from engineered features, the LSTM component models temporal dependencies and sequential access behavior. The fusion of these two perspectives enables the framework to identify complex insider threat patterns that involve both static anomalies and evolving behavioral trends.

Moreover, the decision-level fusion strategy contributes to a noticeable reduction in false positives. This is particularly important in healthcare EHR systems, where legitimate access patterns frequently vary due to clinical workflows, emergency situations, and role changes. By jointly analyzing static and temporal behavior characteristics, the proposed framework improves detection reliability while minimizing unnecessary security alerts [18].

Overall, the experimental findings confirm that hybrid ML–DL modeling provides a robust and effective solution for insider threat detection in healthcare systems, outperforming standalone approach; es in both accuracy and practical applicability.

Table III. Comparison of Standalone vs Hybrid Models

Approach	Detection Capability	Temporal Modeling	Interpretability
ML Only	Medium	No	High
DL Only (LSTM)	Medium–High	Yes	Low
Hybrid ML–DL	High	Yes	Medium–High

Table III provides a qualitative comparison of standalone and hybrid detection approaches in terms of detection capability, temporal modeling, and interpretability.

Conclusion and Research Directions

This paper presented an integrated machine learning and deep learning framework for insider threat detection in healthcare EHR systems. By combining structured feature-based learning with temporal sequence modeling, the proposed approach effectively captures both static and evolving insider behavior patterns. The hybrid framework demonstrates superior detection performance compared to individual ML and DL models, particularly in terms of recall and F1-score, which are critical metrics for healthcare security applications.

The experimental results highlight the importance of hybrid modeling strategies for addressing the complexity of insider threats in healthcare environments. By reducing false positives and improving detection robustness, the proposed framework enhances the reliability and trustworthiness of insider threat detection systems.

Future research will focus on incorporating explainable AI techniques to improve transparency and interpretability of detection decisions, validating the framework using real-world healthcare datasets, and exploring adaptive and online learning mechanisms to handle evolving insider behaviors and emerging security threats [14], [18].

References

1. M. Bishop and C. Gates, "Defining the insider threat," Proceedings of the 4th Annual Workshop on Cyber Security and Information Intelligence Research, pp. 1–4, 2008.
2. Verizon, "2023 Data Breach Investigations Report (DBIR)," Verizon Enterprise, 2023.
3. H. Hu, G. Ahn, and J. Jorgensen, "Detecting anomalous behavior of insiders using machine learning techniques," IEEE Systems Journal, vol. 12, no. 2, pp. 1223–1234, 2018.
4. Y. Liu, Y. Zhang, and J. Zhang, "Insider threat detection using deep neural networks," ACM Transactions on Privacy and Security, vol. 21, no. 4, pp. 1–27, 2018.
5. J. Glasser and B. Lindauer, "Bridging the gap: A pragmatic approach to generating insider threat data," IEEE Security & Privacy Workshops, pp. 98–104, 2013.
6. CERT Division, "Insider Threat Dataset," Software Engineering Institute, Carnegie MellonUniversity.
7. A. Eberle and L. Holder, "Insider threat detection using graph-based approaches," Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics, pp. 1–8,2009.
8. S. Salem, A. Hershkop, and S. Stolfo, "A survey of insider attack detection research," Insider Attack and Cyber Security, Springer, pp. 69–90, 2008.
9. A. Tuor et al., "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," AAAI Workshops, 2017.
10. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
11. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
12. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proceedings of the 22nd ACM SIGKDD Conference, pp. 785–794, 2016.
13. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann2011.
14. N. Japkowicz and M. Shah, Evaluating Learning Algorithms: A Classification Perspective, Cambridge University Press, 2011.
15. A. Shabtai, U. Kanonov, Y. Elovici, C. Glezer, and Y. Weiss, "Andromaly: A behavioral malware detection framework for Android devices," Journal of Intelligent Information Systems, vol. 38, no. 1, pp. 161–190, 2012.
16. R. Shokri, G. Theodorakopoulos, J. Le Boudec, and J.-Y. Le Boudec, "Quantifying location privacy," IEEE Symposium on Security and Privacy, pp. 247–262, 2011.
17. A. G. Bardas and J. S. Jenkins, "Cybersecurity in healthcare: A systematic review of modern threats and trends," Journal of Healthcare Information Security, vol. 5, no. 2, pp. 1–12,2020.