International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

Feature Engineering for Insider Threat Detection in Healthcare Systems

Thasleem R[1], Dr. Sumathy Kingslin[2]

[1]Associate Professor, [2]Research Scholar, Department of Computer Science, Quaid-E-Millath Govt College for Women (A), Chennai, India

***Abstract*-- Insider threats represent a major security challenge for healthcare information systems due to the presence of privileged users and the highly sensitive nature of electronic health records (EHRs). Although artificial intelligence–based approaches have been widely applied for insider threat detection, their effectiveness is often limited by inadequate and non-domain-specific feature representation. This paper presents a comprehensive feature engineering strategy for insider threat detection in healthcare environments by systematically integrating behavioral, access pattern, temporal, statistical, and contextual features derived from system logs and EHR usage data. The proposed feature set is evaluated using multiple machine learning models to assess its impact on detection performance in terms of accuracy, precision, recall, F1-score, and AUC. Experimental results demonstrate that the engineered features significantly outperform baseline raw feature representations across all evaluation metrics. The findings highlight the critical role of domain-aware feature engineering in improving the reliability, robustness, and practical applicability of AI-based insider threat detection systems in healthcare settings.**

***Keywords*-- Insider Threat Detection, Healthcare Cybersecurity, Feature Engineering, Electronic Health Records, Machine Learning, Anomaly Detection, Access Control, User Behavior Analysis**

I. INTRODUCTION

Healthcare organizations increasingly depend on Electronic Health Record (EHR) systems to manage sensitive patient data. Despite the deployment of advanced cybersecurity measures such as firewalls, intrusion detection systems, and encryption, insider threats remain one of the most serious and least addressed security challenges in healthcare environments [1], [2]. Insider threats arise from authorized users, including clinicians and administrative staff, who misuse legitimate access privileges, either intentionally or unintentionally, resulting in privacy breaches and regulatory violations.

Unlike external attackers, insiders possess detailed knowledge of system workflows and access controls, enabling them to evade traditional security mechanisms [1].

Healthcare systems are particularly vulnerable due to complex access requirements, high staff turnover, and frequent emergency access scenarios, leading to a higher incidence of insider-related security events [2], [18]. Consequently, rule-based security mechanisms are often insufficient for accurately identifying malicious insider behavior.

Artificial intelligence (AI) and machine learning (ML) techniques have been widely explored for insider threat detection using system logs and user activity data [3], [4]. However, their performance is highly dependent on feature quality. Many existing approaches rely on raw or generic features that fail to capture subtle behavioral deviations and healthcare-specific context, resulting in high false positive rates [8].

Feature engineering is therefore essential for transforming raw logs into meaningful, domain-aware representations that enhance detection performance [14]. In healthcare environments, effective feature engineering must incorporate behavioral, temporal, access-based, and contextual information, such as role-based access deviation and abnormal record access patterns [18]. Nevertheless, systematic feature engineering strategies tailored for healthcare insider threat detection remain limited in existing research.

II. CONTRIBUTIONS OF THIS PAPER

To address the above challenges, this paper proposes a comprehensive feature engineering strategy specifically designed for insider threat detection in healthcare systems. The main contributions of this work are as follows:

1. A domain-aware feature categorization framework that integrates behavioral, access-based, temporal, statistical, and contextual features derived from healthcare system and EHR access logs.
2. A systematic feature extraction methodology tailored for healthcare environments, enabling effective representation of insider activity patterns.
3. An experimental evaluation of the proposed feature set using multiple machine learning models to assess its impact on detection performance.

4. A comparative analysis between baseline raw features and engineered features to demonstrate performance improvements across standard evaluation metrics.
5. Practical insights for healthcare cybersecurity researchers and practitioners on designing effective feature sets for reliable insider threat detection.

The proposed feature engineering strategy aims to enhance the accuracy, robustness, and interpretability of AI-based insider threat detection models, thereby improving the security and trustworthiness of healthcare information systems.

III. Related Work

Feature Usage In Insider Threat Detection

Early insider threat detection systems relied on rule-based and policy-driven mechanisms that monitored predefined violations. Although simple to implement, these approaches lacked adaptability and were ineffective against sophisticated insider behaviors, often failing to distinguish malicious actions from legitimate operational anomalies [1], [8].With advances in machine learning, researchers began using user activity logs, authentication records, file access histories, and device usage data to develop predictive detection models [3], [7]. Common features include login frequency, session duration, file access counts, and device usage patterns. However, these features provide only limited behavioral representation and lack contextual depth.

To improve detection performance, statistical features such as mean activity, variance, and entropy were introduced to capture behavioral variability [16], [17], along with temporal features such as after-hours and weekend access to identify timing anomalies [9]. More recent studies employ deep learning models, including LSTM networks, to learn sequential behavior patterns [4], [9]. Despite their potential, these models still depend on effective feature preprocessing and often lack interpretability.

In healthcare-focused studies, generic cybersecurity feature sets originally designed for enterprise environments are commonly reused [18]. Such features fail to capture clinical workflows, role-based access behavior, and patient data sensitivity, leading to reduced detection accuracy in healthcare settings.

Gap In Existing Works:

Despite notable progress in insider threat detection research, several important limitations remain:

Lack of healthcare-specific feature design: Most existing feature sets are not explicitly tailored to healthcare workflows, clinical roles, or patient-centric access requirements, limiting their applicability in healthcare environments [18].

Limited contextual understanding: Critical contextual attributes such as user role, department affiliation, and operational responsibilities are often ignored, leading to inaccurate interpretation of user behavior [7].

Unstructured featured selection: Features are frequently selected without systematic categorization or justification, making feature analysis and comparison difficult across studies [8].

High false alarm rates: Generic feature representations fail to distinguish legitimate medical access from malicious behavior, resulting in excessive false positives [3].

Minimal emphasis on Feature Engineering: Feature engineering is often treated as a secondary preprocessing step rather than a core research contribution, despite its significant impact on model performance [14].

These limitations highlight the necessity for a structured, domain-aware feature engineering strategy specifically designed for insider threat detection in healthcare systems.

IV. Feature Engineering Stratergy

This section presents the proposed feature engineering framework that transforms raw healthcare system logs into structured and meaningful representations for insider threat detection. The framework is designed to capture behavioral patterns, access characteristics, temporal anomalies, statistical trends, and contextual information relevant to healthcare environments, which have been shown to significantly influence insider threat detection performance [3], [14].

4.1 User Behavior Features

User behavior features describe how frequently and consistently users interact with healthcare systems and are widely used indicators in insider threat detection research [3], [8].

Login Frequency: Login frequency measures how often a user logs into the system within a defined time period. Significant deviations from a user's historical login behavior may indicate credential misuse, account sharing, or abnormal activity [3].

Session Duration: Session duration represents the length of time a user remains logged into the system. Extremely long sessions may suggest unattended access, while unusually short sessions may indicate automated or scripted behavior [7].

Failed Login Count: This feature counts unsuccessful authentication attempts. A high number of failed login attempts may signal unauthorized access attempts or compromised credentials [1].

4.2 Access Pattern Features

Access pattern features characterize how users interact with healthcare data and are critical for detecting insider misuse involving sensitive information [18].

EHR Record Access Count: This feature measures the number of patient records accessed by a user within a specific time window. Excessive access beyond operational requirements may indicate data harvesting or privacy violations [18].

Sensitive File Access: Sensitive file access tracks interactions with high-sensitivity medical records, such as psychiatric, oncology, or financial data. Unauthorized access to such records is considered a strong indicator of insider threat behavior in healthcare environments [18].

Access Time Deviation:

Access time deviation captures differences between current access times and a user's historical access patterns. Significant deviations may reflect abnormal or suspicious behavior [9].

4.3 Temporal Features

Temporal features analyze the timing characteristics of user activities and are effective in identifying insider threats that exploit unusual access times [9].

Access Outside Duty Hours:

This feature identifies system access occurring beyond a user's assigned working hours. Frequent after-hours access may indicate unauthorized activity, especially for non-shift-based staff [9].

Weekend Access Ratio:

Weekend access ratio measures the proportion of user activity occurring during weekends. Elevated weekend access may represent abnormal behavior in standard healthcare workflows [3].

4.4 Statistical Features

Statistical features summarize long-term user behavior trends and improve model stability and generalization [16].

Mean:

The mean represents average activity levels, such as the average number of accesses or session duration, providing a baseline behavioral profile.

Variance:

Variance measures fluctuations in user behavior over time. Higher variance may indicate inconsistent or anomalous activity patterns [16].

Entropy:

Entropy quantifies the unpredictability of user actions. Higher entropy values suggest irregular behavior and have been shown to be effective in insider threat detection [17].

4.5 Contextual Features

Contextual features incorporate healthcare-specific operational knowledge and play a crucial role in reducing false positives [18].

Role-Based Deviation:

Role-based deviation measures how much a user's behavior differs from typical behavior associated with their professional role (e.g., doctor, nurse, technician). Significant deviations may indicate misuse of privileges [18].

Department Mismatch:

This feature identifies access to patient records belonging to departments outside the user's assigned unit. Such mismatches often signal suspicious activity in healthcare environments where access is typically department-specific [18].

V. Dataset And Preprocessing

This study evaluates the proposed feature engineering strategy using the CERT insider threat dataset along with simulated healthcare system logs. The use of a publicly available benchmark dataset ensures reproducibility, while simulated healthcare logs provide domain relevance for EHR-based insider threat analysis [5], [6], [18].

5.1 CERT Insider Threat Dataset

The CERT insider threat dataset contains realistic synthetic organizational activity records, including login events, file access logs, device usage, web activity, and email communications [5], [6]. The dataset includes labeled normal and malicious insider behavior, making it suitable for supervised learning experiments. Although not healthcare-specific, selected attributes related to user behavior, access patterns, and temporal activity are mapped to healthcare operations such as EHR access and patient record usage [3], [7].

5.2 Simulated Healthcare Log Dataset

To capture healthcare-specific characteristics, simulated EHR access logs are generated to model real-world clinical workflows. The simulation includes multiple user roles (e.g., physicians, nurses, technicians, and administrators), department associations, access privileges, record sensitivity levels, and working schedules [18]. Malicious insider scenarios are introduced through abnormal access behaviors such as excessive record access, cross-department usage, sensitive file access, and after-hours activity, ensuring realistic representation of insider threat patterns.

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**

5.3 Data Preprocessing

Data preprocessing includes removal of duplicate entries, handling of missing values, filtering of incomplete records, and timestamp standardization to ensure data consistency [14]. Numerical features such as login frequency, access count, and session duration are normalized using min–max scaling to reduce bias caused by differing feature scales [14]. Categorical attributes, including user role, department, and access type, are transformed into numerical representations using label encoding and one-hot encoding techniques.

After preprocessing, the final dataset consists of structured feature vectors representing user activity instances, which serve as input to machine learning models for insider threat detection.

VI. EXPERIMENTAL EVALUATION

This section presents the experimental setup used to evaluate the effectiveness of the proposed feature engineering strategy for insider threat detection in healthcare systems. Multiple machine learning and deep learning models are employed to analyze the impact of engineered features on detection performance [3], [4].

6.1 Experimental Setup

Experiments are conducted using the preprocessed datasets described in Section 4. The data is divided into training and testing subsets to ensure fair evaluation, and experiments are repeated multiple times to improve result reliability [15]. Two feature configurations are evaluated:

- *Baseline features:* raw log attributes without structured feature engineering.
- *Proposed features:* engineered behavioral, access-based, temporal, statistical, and contextual features.

6.2 Classification Models

Several widely used models are selected for evaluation due to their effectiveness in insider threat and anomaly detection tasks.

Random Forest (RF): An ensemble learning method that combines multiple decision trees and is robust to noise and high-dimensional feature spaces [11].

Support Vector Machine (SVM): A margin-based classifier that performs well with complex feature distributions and limited training data [12].

XGBoost: A gradient boosting framework known for high predictive accuracy and its ability to capture non-linear feature interactions [13].

Long Short-Term Memory (LSTM): A recurrent neural network model used to capture sequential and temporal behavior patterns in user activity data [4], [9].

6.3 Performance Metrics

Model performance is evaluated using standard classification metrics [15]:

- Accuracy
- Precision
- Recall
- F1-score
- Area Under the ROC Curve (AUC)

These metrics provide a comprehensive assessment of detection effectiveness and reliability.

VII. Result And Discussion

7.1 Comparison with Baseline Features

Experimental results show that models trained using the proposed engineered features consistently outperform baseline models across all evaluation metrics.

Baseline models struggle to detect subtle insider behaviors, resulting in higher false positive and false negative rates [3]. In contrast, domain-aware engineered features significantly improve accuracy, precision, and recall, particularly for ensemble models such as Random Forest and XGBoost [11], [13].

Model Performance Comparison

Model	**Feature Type**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC**
Random Forest	Baseline	89.2	71.4	65.8	68.5	0.81
Random Forest	Proposed	**95.6**	**88.3**	**85.1**	**86.7**	**0.93**
SVM	Baseline	86.5	68.9	61.2	64.8	0.78
SVM	Proposed	**92.8**	**83.6**	**80.4**	**82.0**	**0.89**
XGBoost	Baseline	90.4	74.6	69.3	71.8	0.83
XGBoost	Proposed	**96.8**	**90.7**	**88.9**	**89.8**	**0.95**
LSTM	Baseline	88.1	70.2	66.5	68.3	0.80
LSTM	Proposed	**94.3**	**86.1**	**83.7**	**84.9**	**0.92**

7.2 Impact of Individual Feature Groups

The contribution of individual feature groups is analyzed by incrementally adding feature categories to the baseline model:

- *User behavior features* capture abnormal login and session patterns [3].
- *Access pattern features* effectively identify data misuse through excessive or sensitive record access [18].
- *Temporal features* enhance detection of after-hours and weekend access anomalies [9].
- *Statistical features* improve model stability by summarizing long-term behavior trends [16].
- *Contextual features* provide the greatest performance improvement by incorporating role-based and departmental information, significantly reducing false alarms [18].

These results confirm that integrating multiple feature perspectives leads to superior detection performance compared to using individual feature groups.

VIII. Conclusion And Future Work

This paper presented a structured, healthcare-oriented feature engineering strategy for insider threat detection. By transforming raw system logs into meaningful behavioral, access-based, temporal, statistical, and contextual representations, the proposed approach significantly enhances the performance of machine learning-based detection models [14], [18].

Experimental evaluation using benchmark and simulated healthcare datasets demonstrates notable improvements in accuracy, precision, recall, and robustness compared to baseline feature representations. The findings highlight the critical role of feature engineering in developing reliable and interpretable insider threat detection systems.

Future work will focus on incorporating real-world healthcare datasets, exploring adaptive and online feature learning techniques, and integrating explainable AI methods to improve transparency and trust in healthcare cybersecurity systems.

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 Mellon University.

[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] . 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] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[13] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” Proceedings of the 22nd ACM SIGKDD Conference, pp. 785–794, 2016.

[14] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann2011.

[15] N. Japkowicz and M. Shah, Evaluating Learning Algorithms: A Classification Perspective, Cambridge University Press, 2011.

[16] 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.

[17] 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.

[18] 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.