VARSTVOSLOVJE
Journal of Criminal Justice and Security
year 2026
volume 28
pp. 1–24

The Cert Dataset Decade: A Systematic Review of Methodological Evolution and Performance Bias

Marko Jurišić, Igor Tomičić

Purpose:

The purpose of this paper is to identify methodological biases and limitations in machine learning–based insider threat detection using the Computer Emergency Response Team [CERT] dataset, in order to guide the development of more realistic, robust, and operationally relevant detection approaches.

Design/Methods/Approach:

The objectives are achieved through a systematic literature analysis of 131 peer-reviewed studies published between 2013 and 2025 that apply machine learning to insider threat detection using the CERT dataset, employing a Preferred Reporting Items for Systematic Reviews and Meta-Analyses [PRISMA]-guided selection process and a structured comparative framework to examine dataset versions, feature engineering strategies, model architectures, and evaluation metrics from a methodological and empirical perspective.

Findings:

The analysis shows that most studies rely on the less realistic CERT v4.2 dataset, resulting in inflated performance that does not generalize to operational settings. It also finds that feature engineering is a stronger determinant of detection performance than model complexity, while inconsistent evaluation practices hinder meaningful comparison across studies.

Research Limitations / Implications:

The study is limited by its reliance on published research using a single synthetic dataset, which constrains generalization to real-world environments.

Practical Implications:

The findings indicate that practitioners should be cautious when adopting models validated on simplified benchmark settings, and instead prioritize solutions tested under extreme class imbalance. Emphasis should be placed on robust feature engineering, unsupervised or hybrid detection approaches, and evaluation metrics.

Originality/Value:

This paper provides the first large-scale, methodologically focused analysis of insider threat detection research that explicitly exposes performance inflation caused by dataset version bias and evaluation inconsistency, offering concrete, evidence-based guidance for improving the realism, comparability, and operational value of future studies in the field.

Keywords: insider threat detection, CERT dataset, machine learning, anomaly detection, dataset bias, evaluation metrics

UDC: 004.056

Desetletje nabora podatkov CERT: sistematični pregled metodološkega razvoja in pristranskosti zmogljivosti

Namen prispevka:

Namen prispevka je opredeliti metodološke pristranskosti in omejitve pri odkrivanju notranjih groženj na osnovi strojnega učenja z uporabo nabora podatkov CERT, da bi usmerili razvoj bolj realističnih, robustnih in operativno uporabnih pristopov za zaznavanje.

Metode:

Cilji so doseženi s sistematično analizo literature 131 recenziranih študij, objavljenih med letoma 2013 in 2025, ki uporabljajo strojno učenje za odkrivanje notranjih groženj na podlagi nabora podatkov CERT. Uporabljena sta bila postopek izbora po smernicah Prednostne postavke poročanja za sistematične preglede in metaanalize (angl. PRISMA – *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) ter strukturiran primerjalni okvir za proučevanje različic nabora podatkov, strategij značilnosti inženiringa, arhitektur modelov in evalvacijskih metrik z metodološkega in empiričnega vidika.

Ugotovitve:

Analiza kaže, da se večina študij zanaša na manj realističen nabor podatkov CERT v4.2, kar vodi do precenjenih rezultatov zmogljivosti, ki se ne posplošujejo na operativna okolja. Poleg tega ugotavlja, da je značilnost inženiringa pomembnejši dejavnik uspešnosti zaznavanja kot kompleksnost modelov, medtem ko nedosledne evalvacijske prakse otežujejo smiselno primerjavo med študijami.

Omejitve/uporabnost raziskave:

Študija je omejena zaradi zanašanja na objavljeno literaturo, ki uporablja en sam sintetični nabor podatkov, kar omejuje posploševanje na resnična okolja.

Praktična uporabnost:

Ugotovitve kažejo, da bi morali biti praktiki previdni pri uvajanju modelov, validiranih na poenostavljenih referenčnih okoljih, ter namesto tega dajati prednost rešitvam, preizkušenim v pogojih izrazite neuravnoteženosti razredov. Poudarek bi moral biti na značilnosti robustnega inženiringa, nenadzorovanih ali

hibridnih pristopih zaznavanja ter evalvacijskih metrikah.

Izvirnost/pomembnost prispevka:

Prispevek predstavlja prvo obsežno, metodološko usmerjeno analizo raziskav na področju odkrivanja notranjih groženj, ki izrecno razkriva precenjenost rezultatov zmogljivosti zaradi pristranskosti različic naborov podatkov in nedoslednosti evalvacije ter ponuja konkretna, na dokazih temelječa priporočila za izboljšanje realističnosti, primerljivosti in operativne vrednosti prihodnjih raziskav na tem področju.

Ključne besede: odkrivanje notranjih groženj, nabor podatkov, CERT, strojno učenje, zaznavanje anomalij, pristranskost nabora podatkov, evalvacijske metrike

UDK: 004.056

1 INTRODUCTION

Insider threat detection remains a critical challenge in cybersecurity due to the difficulty of distinguishing malicious intent from authorized user behavior. Since real-world datasets are scarce due to privacy concerns or trade secrets, synthetic datasets have become the standard alternative. The most widely used synthetic benchmark is the Computer Emergency Response Team [CERT] dataset, developed by the Software Engineering Institute at Carnegie Mellon University, in partnership with ExactData, LLC, and under sponsorship from DARPA I2O (Glasser & Lindauer, 2013).

In operational environments, insider threats are typically handled within the mandate of Computer Emergency Response Teams [CERTs], which are increasingly required to address not only external incidents, but also misuse, data exfiltration, and policy violations originating from within the organization. Unlike perimeter-focused attacks, insider incidents unfold gradually, often blending into legitimate activity streams, and evading traditional rule-based monitoring deployed by CERTs.

This places a growing burden on CERTs to adopt behavior-based and anomaly-driven detection mechanisms capable of identifying subtle deviations over extended periods, under extreme class imbalance. Consequently, the CERT dataset has become a de facto experimental proxy for evaluating whether machine learning-based approaches can realistically support CERT operations in detecting internal threats, triaging alerts, and prioritizing analyst effort.

Simulating logs from a large organization over 500 days, the CERT dataset captures logon events, HTTP activity, file transfers, device usage, and emails as well as user psychometric profiles (big five' personality traits), and decoy files. Various insider threat scenarios (Table 1) were injected into this background traffic. A critical distinction exists between dataset versions: version 4.2 features 1000 users and 70 insiders (a dense, easier classification problem), whereas the newer version 6.2 features 4000 users and only 5 insiders (a realistic, high-imbalance problem).

Despite being released in 2013, over a decade ago, CERT remains the primary testbed for new ML approaches. This review systematically analyzes 131 papers to answer four key questions, guiding our work similarly to previous research (Jurišić et al., 2023):

1. Which dataset version is predominantly used?
2. How has feature engineering evolved?
3. Which machine learning methods are most effective?
4. What evaluation metrics are standard?

Table 1: CERT Threat Scenarios (v6.2)

Scen.	Description
1	User logs in after hours, uses removable drives, and uploads data to Wikileaks before leaving.
2	User surfs job sites, solicits employment from competitors, and steals data via thumb drive.
3	Sysadmin uses a keylogger to steal supervisor credentials and sends an alarming mass email.
4	User logs into another machine, searches for files, and emails them to a home address over 3 months.
5	Layoff victim uploads documents to Dropbox for personal gain.

2 SEARCH AND SELECTION METHODOLOGY

We conducted a systematic search on Scopus and Web of Science (WoS) covering the period from the dataset's release (2013) to September 2025. The search query targeted the intersection of the dataset citation (Glasser & Lindauer, 2013) and keywords including *anomaly detection*, *machine learning*, *deep learning*, and *security*.

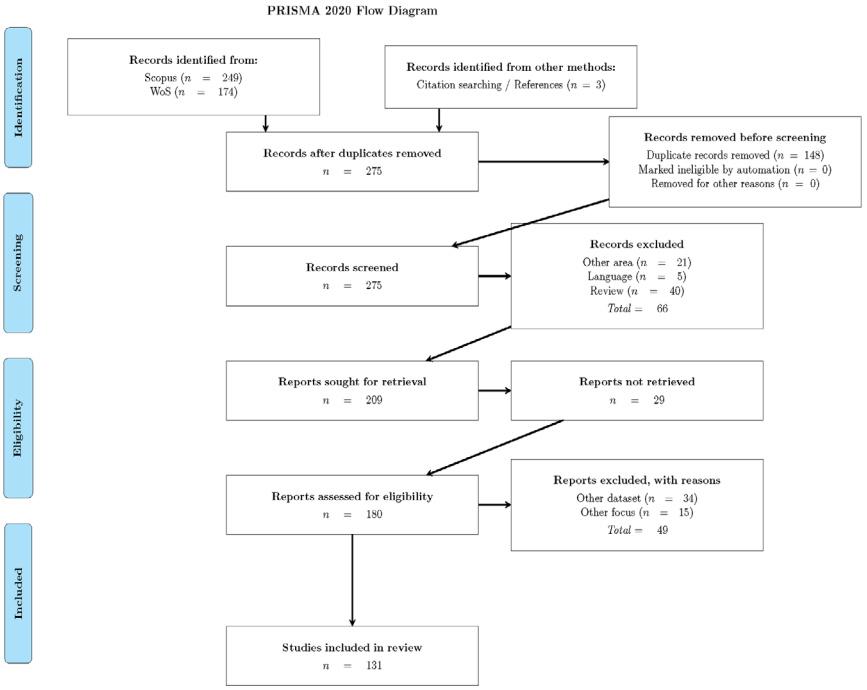

Figure 1: PRISMA 2020 flow diagram of the selection process. The search initially yielded 275 records, resulting in 131 included studies after screening and eligibility checks.

2.1 PRISMA FLOW

The initial search yielded 275 unique records. Following PRISMA 2020 guidelines (Page et al., 2021):

1. Screening: 66 records were removed (reviews, non-English, off-topic).
2. Retrieval: 29 papers were inaccessible.
3. Eligibility: 49 papers were excluded after full-text reading (did not use at all CERT for model training).
4. Included: A final set of 131 studies was selected for this review (see Fig. 1).

Interest in the dataset remains robust, with citation counts indicating a steady stream of new methodologies applied to CERT every year, either as the only dataset, or as one of the benchmarks (Fig. 2).

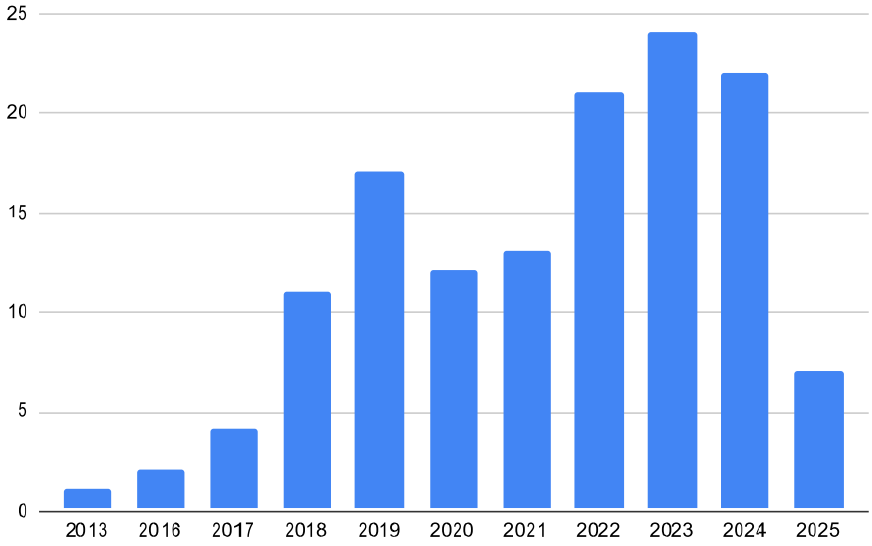

Figure 2: Citations per year of the original CERT dataset paper (Glasser & Lindauer, 2013).

3 SURVEY OF MODELS AND TECHNIQUES

The literature demonstrates a clear methodological progression from static classifiers to sequence-aware deep learning and graph-based approaches, and, recently, LLM approaches.

There are many possible approaches to group and categorize different machine learning methods and approaches, e.g. supervised and unsupervised methods, historical and recent methods, singular models vs. ensemble and hybrid approaches. In this context, the machine learning methods used are organized into broad groups based on similar approaches and characteristics. It is important to note that these categories are not strictly mutually exclusive, as many models possess characteristics of multiple groups. This pragmatic grouping is intended to provide a clear framework for comparing the diverse approaches applied to the CERT dataset.

Our review categorizes these approaches based on their primary architectural logic. It is important to note that many authors use multiple methods, either to establish baselines and validate the performance gains of the proposed approach or they employ ensembles, where the model's output is based on several classifiers. One recent trend is the increase of hybrid approaches, where the authors use output of one method as input for the next one, e.g. using LSTM for extraction of sequence embeddings and CNN for final classification, highlighting the complementary strengths of these techniques. Table 2 categorizes the reviewed studies.

Table 2: Overview of Methods Applied to the CERT Dataset

Model Category	Method	References
Linear Models	Linear & Logistic Regression (LR)	Jiang et al., 2019b, Jiang et al., 2019a, Ferreira et al., 2019, D. Le and Nur Zincir-Heywood, 2019, D. Le et al., 2020, D. Le and Zincir-Heywood, 2021b, Bharathi and Balasubramanian, 2022, C. Wang and Zhu, 2022, X. Li, Li, et al., 2023, X. Li, Li, Li, Cai, and Li, 2023, Abdallah et al., 2024, D. Le et al., 2019, Dong et al., 2025
Tree-Based Models	Random Forest (RF)	Jiang et al., 2019b, Jiang et al., 2019a, Ferreira et al., 2019, D. Le and Nur Zincir-Heywood, 2019, Rauf et al., 2019, D. Le et al., 2020, D. Le and ZincirHeywood, 2021b, Al-Shehari and Alsowail, 2021, D. C. Le et al., 2021, Naicker and van Niekerk, 2021, C. Wang and Zhu, 2022, Al-Shehari and Alsowail, 2023, Mittal and Garg, 2023, Peccatiello et al., 2023, X. Li, Li, et al., 2023, Manoharan et al., 2023, Al-Shehari et al., 2023, Al-Shehari, Rosaci, et al., 2024, Abdallah et al., 2024, Garchery and Granitzer, 2019, Bharathi and Balasubramanian, 2022, D. Le et al., 2019, Dong et al., 2025, C. Zheng et al., 2022, Corradini et al., 2025, Zou et al., 2020, Elisa et al., 2023, Feng et al., 2025
	XGBoost	F. Yuan et al., 2020, D. Le et al., 2020, Al-Shehari and Alsowail, 2023, Al-Shehari et al., 2023, Al-Shehari, Rosaci, et al., 2024, Abdallah et al., 2024, Jovanovic et al., 2024, Mladenovic et al., 2024, Dong et al., 2025, Yi and Tian, 2024, C. Zheng et al., 2022, Zou et al., 2020
	Other Tree-based	D. Le and Zincir-Heywood, 2018, Bharathi and Balasubramanian, 2022, Kumpf et al., 2024, F. Yuan et al., 2020, Mladenovic et al., 2024, Elisa et al., 2023
Instance-Based and	Support Vector Machine (SVM)	Palomares et al., 2017, Haidar and Gaber, 2018, A. Liu et al., 2018, Jiang et al., 2019b, Jiang et al., 2019a, Rauf et al., 2019, Aldairi et al., 2019, Bartoszewski et al., 2021, Padmavathi et al., 2022, Han et al., 2022, Molina et al., 2022, X. Li, Li, et al., 2023, Bin Sarhan and Altwaijry, 2023, Abdallah et al., 2024, Song et al., 2024, X. Zhu et al., 2024, Kan et al., 2023, Zhou et al., 2022, D.-W. Kim et al., 2023, Mehmood et al., 2024, Sridhar, 2025, Sun et al., 2021
Kernel Methods	k-Nearest Neighbors (k-NN)	Bose et al., 2017, Al-Shehari and Alsowail, 2021, Naicker and van Niekerk, 2021, Bharathi and Balasubramanian, 2022, Al-Shehari and Alsowail, 2023, Al-Shehari et al., 2023, Al-Shehari, Rosaci, et al., 2024, Abdallah et al., 2024
Probabilistic Models	Bayesian Methods	Roberts et al., 2016, D. Le et al., 2018, Al-Shehari and Alsowail, 2021, D. Le and Zincir-Heywood, 2021b, Mittal and Garg, 2023, Manoharan et al., 2023, Elisa et al., 2023, Bertrand et al., 2023
	Markovian Models	Rashid et al., 2016, D. Le and Zincir-Heywood, 2018, Dahmane and Foucher, 2018, Saaudi et al., 2019, Ye et al., 2020, Bartoszewski et al., 2021, Alshehri, 2022, Yang et al., 2022, Song et al., 2024, B. Zhang et al., 2024
	Other Probabilistic	Kan et al., 2023, P. Zheng et al.,2021

Neural Networks	LSTM	Tuor et al., 2017a, Tuor et al., 2017b, Saaudi et al., 2018, F. Yuan et al., 2018, Matterer and Lejeune, 2018, Paul and Mishra, 2020, F. Yuan et al., 2020, Y. Wei et al., 2021, C. Zhang et al., 2021, He et al., 2021, Han et al., 2022, Zuo et al., 2022, Alshehri, 2022, Molina et al., 2022, C. Li et al., 2022, Alshehri et al., 2023, Song et al., 2024, Shanmugapriya et al., 2024, X. Zhu et al., 2024, Tiwary et al., 2024, Sridhar, 2025, Vinay et al., 2024, Ding et al., 2024, Cheng et al., 2024, F. Xiao et al., 2025, B. Zhang et al., 2024, Chen and Pao, 2024, BaghalizadehMoghadam et al., 2024, C. Zheng et al., 2022, Sivakrishna et al., 2025, D. Zhu et al., 2022, Sun et al., 2021, Vinay et al., 2023, X. Li et al., 2024
	CNN	Saaudi et al., 2018, F. Yuan et al., 2018, A. Liu et al., 2018, Jiang et al., 2018, Jiang et al., 2019b, Zhou et al., 2022, Shanmugapriya et al., 2024, Tiwary et al., 2024, Ye et al., 2025, AlShehari, Kadrie, et al., 2024, Ding et al., 2024, Dong et al., 2025, Chen and Pao, 2024, Ge et al., 2022, D. Zhu et al., 2022, algabri2025mitd
	Autoencoder	L. Liu et al., 2018, D. C. Le and Zincir-Heywood, 2020, D. Le and Zincir-Heywood, 2021b, L.-P. Yuan et al., 2021, D. Le and Zincir-Heywood, 2021a, D. Li et al., 2022, X. Wang et al., 2023, Y.-F. Wang et al., 2023, Sridhar, 2025, Jang et al., 2020, Corradini et al., 2025, Sivakrishna et al., 2025
	Other NNs	Garchery and Granitzer, 2019, Ferreira et al., 2019, D. Le and Nur Zincir-Heywood, 2019, L. Liu et al., 2019, L. Liu et al., 2020, S. Yuan et al., 2020, D. Le et al., 2020, Gayathri et al., 2020, Zerhoudi et al., 2020, D. Li et al., 2021, Naicker and van Niekerk, 2021, Vinay et al., 2022, Bharathi and Balasubramanian, 2022, C. Li et al., 2022, J. Xiao et al., 2023, Mouyart et al., 2023, Song et al., 2024, Abdallah et al., 2024, Zhou et al., 2022, D. Le et al., 2019, Qi, An, et al., 2025, W. Liu et al., 2025, Bertrand et al., 2024, M. Zhang et al., 2024, Kotb et al., 2025, Chen and Pao, 2024, C. Zheng et al., 2022, Corradini et al., 2025, Sivakrishna et al., 2025, R. Wei et al., 2019
Anomaly Detection	Isolation Forest (IF)	Haidar and Gaber, 2018, Garchery and Granitzer, 2019, Aldairi et al., 2019, D. C. Le and Zincir-Heywood, 2020, S. Yuan et al., 2020, Bartoszewski et al., 2021, D. Le and Zincir-Heywood, 2021a, Han et al., 2022, Molina et al., 2022, Peccatiello et al., 2023, Bin Sarhan and Altwaijry, 2023, Al-Shehari et al., 2023, Song et al., 2024, X. Zhu et al., 2024, Kan et al., 2023, Zhou et al., 2022, Mehmood et al., 2024, Sridhar, 2025, Fei et al., 2025, Yi and Tian, 2024, Corradini et al., 2025, Sun et al., 2021
and Clustering	Local Outlier Factor (LOF)	Nicolaou et al., 2020, Bartoszewski et al., 2021, X. Li, Li, et al., 2023, Rauf et al., 2023, X. Li, Li, Li, Cai, and Li, 2023, Al-Shehari, Rosaci, et al., 2024, Z. Wei et al., 2024, Kan et al., 2023, Mehmood et al., 2024, Yi and Tian, 2024, X. Li et al., 2024
	Clustering	D. Le and Zincir-Heywood, 2018, Haidar and Gaber, 2018, J. Kim et al., 2019, Garchery and Granitzer, 2019

Graph-Based	Graph Neural Networks	Jiang et al., 2019a, X. Li, Li, et al., 2023, X. Li, Li, Li, Cai, and Li, 2023, Y.-F. Wang et al., 2023, Ding et al., 2024, F. Xiao et al., 2025, Fei et al., 2025, L. Zheng et al., 2025, X. Li et al., 2024, Qi, Yan, et al., 2025
Methods	Other Graph-based	Jiang et al., 2019b, Das Bhattacharjee et al., 2017, C. Wang and Zhu, 2022, X. Wang et al., 2023, Fei and Zhou, 2024, B. Zhang et al., 2024
Other Methods	Miscellaneous	D. Le et al., 2018, Igbe and Saadawi, 2018, D. Le et al., 2019, Orizio et al., 2020, Chung et al., 2020, Alhajjar and Bradley, 2022, Elisa et al., 2023, Arendt et al., 2018, Jyosthna and Reddy, 2021

3.1 Tree-Based and Linear Models

Tree-based models, particularly Random Forest (RF) and eXtreme Gradient Boosting (XGBoost), are pervasive in the literature, serving as both high-performing primary classifiers and robust baselines (Al-Shehari & Alsowail, 2023; D. Le et al., 2020). Their popularity stems from their ability to handle the tabular nature of aggregated log features (e.g., daily counts of emails or file accesses). Recent work has favored gradient boosting (XGBoost) due to its effectiveness with imbalanced data when paired with sampling techniques like SMOTE (Abdallah et al., 2024; Zou et al., 2020). Studies consistently show that while Deep Learning offers potential, well-tuned tree ensembles often yield comparable results with significantly lower computational cost and higher interpretability (Le & Nur Zincir-Heywood, 2019).

3.2 Probabilistic and Markovian Models

Early research heavily utilized Hidden Markov Models (HMMs) to capture the sequential nature of user actions, treating logs as a timeline of events (Rashid et al., 2016; Saaudi et al., 2019). While HMMs provide interpretability, they often struggle with the long-term dependencies inherent in prolonged insider attacks. Consequently, recent studies primarily use HMMs as benchmarks for LSTMs. Bayesian methods, including Naive Bayes and Bayesian Gaussian Mixture Models, continue to appear as efficient baselines or components in ensemble systems (Bertrand et al., 2023).

3.3 Recurrent Neural Networks

The introduction of Long Short-Term Memory [LSTM] networks marked a paradigm shift in CERT research (Tuor et al., 2017a). LSTMs address the vanishing gradient problem of standard RNNs, allowing models to learn long-range dependencies in user behavior sequences (e.g., correlating a file download on Monday with data exfiltration on Friday).

Key trends in this category include:

- Sequence Construction: Most studies transform user logs into daily sequences or use sliding windows (e.g., 50 events) to capture temporal context (Saaudi et al., 2019; Tuor et al., 2017a).
- Hybrid Architectures: Combining Convolutional Neural Networks [CNNs] (to extract local feature patterns) with LSTMs (for temporal sequencing) to improve detection rates (Saaudi et al., 2018; F. Yuan et al., 2018).
- Autoencoders: Utilizing LSTM-based autoencoders for unsupervised anomaly detection. These models learn to reconstruct "normal" sequences; high reconstruction errors flag potential threats (Paul & Mishra, 2020; F. Yuan et al., 2020).
- Attention Mechanisms: Recent models incorporate attention layers to focus on specific, high-risk events within a long sequence (He et al., 2021; Zuo et al., 2022).

3.4 Convolutional Neural Networks

While typically used for image processing, CNNs have been adapted for insider threat detection by transforming log data into 2D images (e.g., mapping activity frequencies to pixel intensities) (A. Liu et al., 2018). More commonly, 1D-CNNs are used in hybrid models to extract features from text-based logs (emails, URLs) before passing them to sequential classifiers (Jiang et al., 2018; Ye et al., 2025).

3.5 Unsupervised Anomaly Detection

Given the scarcity of labeled insider data in the real world, unsupervised methods are critical. Isolation Forest (IF) and Local Outlier Factor (LOF) are the standard-bearers here. IF, in particular, is frequently cited as a state-of-the-art baseline, often outperforming complex supervised models when training data is limited or unrepresentative (Al-Shehari et al., 2023; Peccatiello et al., 2023). These methods are increasingly used in ensemble frameworks to reduce false positives (Le & Zincir-Heywood, 2021a).

3.6 Graph-Based Methods

A significant recent development is the shift from analyzing isolated user timelines to modeling the *relationships* between entities (users, devices, files) using graphs. Graph Neural Networks (GNNs) and Graph Convolutional Networks (GCNs) are employed to detect lateral movement and structural anomalies that sequential models might miss (Jiang et al., 2019a; X. Li, Li, et al., 2023). A significant recent development is the shift from homogeneous graphs (user-user relations) to heterogeneous graphs that model the complex interaction between diverse entities (users, devices, files, IP addresses). Methods such as Log2Graph (Fei et al., 2025) and heterogeneous graph embeddings (Zheng et al., 2022) demonstrate that capturing the structural context of an event (e.g., a user accessing a file they have no organizational relationship with) significantly outperforms purely sequential baselines.

3.7 Emerging Methods: Large Language Models and Generative Artificial intelligence

The most recent frontier involves Large Language Models (LLMs). Studies like (Zhang et al., 2024) treat log entries as a language, fine-tuning models like LLaMA to detect semantic anomalies in user behavior. Additionally, Generative Adversarial Networks (GANs) are being used to augment the minority class (insiders) to improve training stability (Yuan et al., 2020).

4 DISCUSSION

4.1 The "Performance Bubble": Version 4.2 vs 6.2

Our analysis reveals a critical bias in the field: the dominance of CERT version 4.2. As shown in Fig. 3, nearly half of all reviewed papers rely on version 4.2, while less than a quarter utilize the more realistic version 6.2.

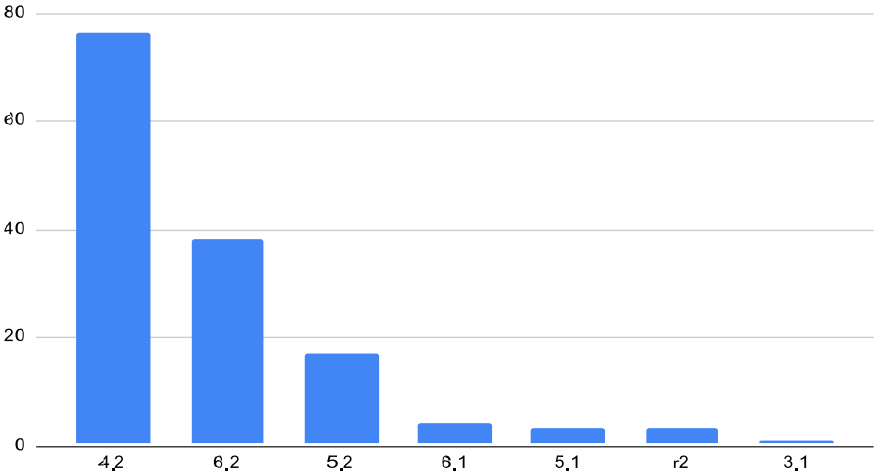

Figure 3: Distribution of CERT dataset versions used in reviewed studies.

This preference is problematic. Version 4.2 contains 70 insiders among 1000 users (1:14 ratio), creating a "dense needle" problem that is significantly easier to solve than the "needle in a haystack" scenario of version 6.2 (5 insiders among 4000 users, 1:800 ratio). Models optimized on 4.2 often report inflated F1-scores that do not generalize to operational environments where threats are extremely rare. This creates a "performance bubble," where published results overstate the practical efficacy of proposed methods.

4.2 Feature Engineering Trends

Feature engineering remains the primary differentiator between successful and unsuccessful models. We observe three dominant approaches in the literature:

- Temporal & Statistical Aggregation: Day-based aggregation is the standard, deriving counts (e.g., number of emails), means, and variances

over daily windows. However, multigranularity approaches (combining session, day, and week views) are proving more robust against "low and slow" attacks (Le et al., 2020).

- Semantic Encoding: There is a decisive move away from simple one-hot encoding toward treating log entries as text. Recent approaches use NLP techniques like TF-IDF or embeddings (Word2Vec, BERT) to capture the contextual relationship between actions (e.g., the similarity between "upload to Dropbox" and "upload to Google Drive") (Bertrand et al., 2023).
- Psychometric Integration: Several studies explicitly integrate the 'Big Five' personality traits provided in the dataset to weight the anomaly scores of users with risk-prone personality profiles.

4.3 Metrics and Standardization

The review highlights a lack of standardized evaluation. While most studies now correctly avoid Accuracy (which is misleading in imbalanced datasets) in favor of AUC-ROC and F1-score, there is no consistency in train/test splits or negative sampling ratios. This fragmentation makes it nearly impossible to determine the true State-of-the-Art (SOTA). Furthermore, operational metrics like *Detection Delay* and *Budget-based Recall* (how many alerts an analyst must review to find an insider) are rarely reported, despite being crucial for real-world adoption.

5 CONCLUSION

This systematic review examined more than a decade of insider threat detection research based on the CERT dataset, revealing substantial methodological progress alongside persistent structural limitations. The field has notably advanced from static classifiers to sequence-aware deep learning models and graph-based representations, but reported performance gains are frequently inflated by dataset version bias, unrealistic class distributions, and inconsistent evaluation practices.

From an operational perspective, these findings have direct implications for CERTs and security operations teams. Models validated primarily on CERT v4.2 provide limited assurance of effectiveness in real-world environments, where insider incidents are rare, ambiguous, and costly to investigate. Without standardized evaluation protocols and metrics that reflect analyst workload (such as detection delay or alert budget efficiency), research outcomes remain weakly coupled to CERT decision-making processes.

To improve both scientific rigor and practical relevance, future research should prioritize evaluation under extreme imbalance conditions, emphasize unsupervised and hybrid detection strategies, and integrate explainability mechanisms that support human-in-the-loop analysis. Cross-version validation and transparent reporting of experimental assumptions should be treated as baseline requirements. Only through such alignment can insider threat detection research meaningfully contribute to CERT capabilities and close the gap between benchmark performance and operational utility.

REFERENCES

Abdallah, H. E. -E., Abd-Elkader, H. H., Mohamed, K. K., Abd-Elmoniem, M., El-Assal, N. W., Mohamed, S. M, Said, S. A., & Salem, S. A. (2024). Performance evaluation framework for insider threat detection using machine learning. *2024 Intelligent Methods, Systems, and Applications (IMSA)*, 1–6. https://doi.org/10.1109/IMSA61967.2024.10652829

Aldairi, M., Karimi, L., & Joshi, J. (2019). A trust aware unsupervised learning approach for insider threat detection. *2019 IEEE 20th international conference on information reuse and integration for data science*, 89–98. https://doi.org/10.1109/IRI.2019.00027

Alhajjar, E., & Bradley, T. (2022). Survival analysis for insider threat: Detecting insider threat incidents using survival analysis techniques. *Computational and mathematical organization theory, 28*(4), 335–351. https://link.springer.com/article/10.1007/s10588-021-09341-0

Al-Shehari, T., Al-Razgan, M., Alfakih, T., Alsowail, R., & Pandiaraj, S. (2023). Insider threat detection model using anomaly-based isolation forest algorithm. *IEEE Access*, *11*, 118170–118185. https://doi.org/10.1109/ACCESS.2023.3326750

Al-Shehari, T., & Alsowail, R. (2023). Random resampling algorithms for addressing the imbalanced dataset classes in insider threat detection. *International Journal of Information Security*, *22*(3), 611–629. https://doi.org/10.1007/s10207-022-00651-1

Al-Shehari, T., Kadrie, M., Al-Mhiqani, M. N., Alfakih, T., Alsalman, H., Uddin, M., Ullah, S. S., & Dandoush, A. (2024). Comparative evaluation of data imbalance addressing techniques for CNN-based insider threat detection. *Scientific Reports, 14*, Article 24715. https://doi.org/10.1038/s41598-024-73510-9

Al-Shehari, T., Rosaci, D., Al-Razgan, M., Alfakih, T., Kadrie, M., Afzal, H., & Nawaz, R. (2024). Enhancing insider threat detection in imbalanced cybersecurity settings using the densitybased local outlier factor algorithm. *IEEE Access*, *12*, 34820–34834. https://doi.org/10.1109/ACCESS.2024.3373694

Al-Shehari, T., & Alsowail, R. A. (2021). An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. *Entropy*, *23*(10), Article 1258. https://doi.org/10.3390/e23101258

Alshehri, A. (2022). Relational deep learning detection with multi-sequence representation for insider threats. *International Journal of Advanced Computer Science and Applications*, *13*(5), 758–765. https://doi.org/10.14569/IJACSA.2022.0130587

Alshehri, A., Khan, N., Alowayr, A., & Alghamdi, M. Y. (2023). Cyberattack detection framework using machine learning and user behavior analytics. *Computer Systems Science and Engineering*, *44*(2), 1679–1689. https://doi.org/10.32604/csse.2023.026526

Arendt, D. L., Franklin, L. R., Yang, F., Brisbois, B. R., & LaMothe, R. R. (2018). Crush Your Data with ViC2ES Then CHISSL Away. *2018 IEEE Symposium on Visualization for Cyber Security (VizSec)*, 1–8. https://doi.org/10.1109/VIZSEC.2018.8709212

Baghalizadeh-Moghadam, N., Neal, C., Cuppens, F., & Boulahia-Cuppens, N. (2024). NLP and Neural Networks for Insider Threat Detection. *2024 IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2010–2018. https://doi.org/10.1109/TrustCom63139.2024.00279

Bartoszewski, F., Just, M., Lones, M., & Mandrychenko, O. (2021). Anomaly detection for insider threats: An objective comparison of machine learning models and ensembles. *IFIP Advances in Information and Communication Technology, 625*, 367–381. https://doi.org/10.1007/978-3-030-78120-0_24

Bertrand, S., Germain, P., & Tawbi, N. (2024). Unsupervised Insider Threat Detection Using Multi-Head Self-Attention Mechanisms. *4th Intelligent Cybersecurity Conference (ICSC)*, 228–236. https://doi.org/10.1109/ICSC63108.2024.10895186

Bertrand, S., Desharnais, J., & Tawbi, N. (2023). Unsupervised User-Based Insider Threat Detection Using Bayesian Gaussian Mixture Models. *20th Annual International Conference on Privacy, Security and Trust (PST)*, 1-10. https://doi.org/10.1109/PST58708.2023.10320169

Bharathi, S. T., & Balasubramanian, C. (2022). Non-trusted user classification-comparative analysis of machine and deep learning approaches. *International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 316–324. https://doi.org/10.1109/ICAISS55157.2022.10010811

Bin Sarhan, B., & Altwaijry, N. (2023). Insider threat detection using machine learning approach. *Applied Sciences*, *13*(1), Article 259. https://doi.org/10.3390/app13010259

Bose, B., Avasarala, B., Tirthapura, S., Chung, Y. -Y., & Steiner, D. (2017). Detecting insider threats using radish: A system for real-time anomaly detection in heterogeneous data streams. *IEEE Systems Journal*, *11*(2), 471–482. https://doi.org/10.1109/JSYST.2016.2558507

Chen, C. -C., & Pao, H. -K. (2024). Reconsider time series analysis for insider threat detection. *2024 IEEE International Conference on Big Data (BigData)*, 1558–1565. https://doi.org/10.1109/BigData62323.2024.10825829

Cheng, H., Xu, D., Yuan, S., & Wu, X. (2024). Achieving counterfactual explanation for sequence anomaly detection. *Machine Learning and Knowledge Discovery in Databases. Research Track and Demo Track*, 19– 35. https://doi.org/10.1007/978-3-031-70371-3_2

Chung, M.-H., Chignell, M., Wang, L., Jovicic, A., & Raman, A. (2020). Interactive machine learning for data exfiltration detection: Active learning with human expertise. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 280–287. https://doi.org/10.1109/SMC42975.2020.9282831

Corradini, E., Chen, W., & Cauteruccio, F. (2025). Ioe-graphformer: A graph transformer-based framework for anomaly detection in internet of everything. *21st International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSSIoT)*, 1009–1014. https://dx.doi.org/10.1109/DCOSS-IoT65416.2025.00152

Dahmane, M., & Foucher, S. (2018). Combating insider threats by user profiling from activity logging data. *1st International Conference on Data Intelligence and Security (ICDIS)*, 194–199. https://doi.org/10.1109/ICDIS.2018.00039

Das Bhattacharjee, S., Yuan, J., Jiaqi, Z., & Tan, Y. -P. (2017). Context-aware graph-based analysis for detecting anomalous activities. *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 1021–1026. https://doi.org/10.1109/ICME.2017.8019421

Ding, J., Qian, P., Ma, J., Wang, Z., Lu, Y., & Xie, X. (2024). Detect insider threat with associated session graph. *Electronics, 13*(24), Article 4885. https://doi.org/10.3390/electronics13244885

Dong, J., Wei, J., Hu, X., Dong, Z., Chen, F., Hu, X., & Qi, J. (2025). DDCC: Synergizing Denoising Diffusion Probabilistic Models and Curriculum-Based Complexity Control for Insider Threat Detection. *IEEE Transactions on Information Forensics and Security IEEE Transactions on Industrial Informatics, 21*(11), 8351–8361 https://doi.org/10.1109/TII.2025.3574417

Elisa, N., Yang, L., Chao, F., Naik, N., & Boongoen, T. (2023). A Secure and Privacy-Preserving E-Government Framework Using Blockchain and Artificial Immunity. *IEEE Access*, *11*, 8773–8789. https://doi.org/10.1109/ACCESS.2023.3239814

Fei, K., & Zhou, J. (2024). An insider threat investigation method by graph analysis with log texts. *2024 3rd International Conference on Networks, Communications and Information Technology*, 19–23. https://doi.org/10.3233/JCS-230092

Fei, K., Zhou, J., Su, L., Wang, W., & Chen, Y. (2025). Log2graph: A graph convolution neural network based method for insider threat detection. *Journal of Computer Security, 33*(1), 37–56. https://doi.org/10.3233/JCS-23009

Feng, W., Cao, Y., Chen, Y., Wang, Y., Hu, N., Jia, Y., & Gu, Z. (2025). Multi-granularity user anomalous behavior detection. *Applied Sciences, 15*(1), Article 128. https://doi.org/10.3390/app15010128

Ferreira, P., Le, D. C., & Zincir-Heywood, N. (2019). Exploring Feature Normalization and Temporal Information for Machine Learning Based Insider Threat Detection. *15th International Conference on Network and Service Management (CNSM)*, 1–7. https://doi.org/10.23919/CNSM46954.2019.9012708

Garchery, M., & Granitzer, M. (2019). Identifying and Clustering Users for Unsupervised Intrusion Detection in Corporate Audit Sessions. *2019 IEEE International Conference on Cognitive Computing (ICCC)*, 19–27. https://doi.org/10.1109/ICCC.2019.00016

Gayathri, R., Sajjanhar, A., & Xiang, Y. (2020). Image-Based Feature Representation for Insider Threat Classification. *Applied Sciences*, *10*(14), Article 4945. https://doi.org/10.3390/app10144945

Ge, D., Zhong, S., & Chen, K. (2022). Multi-source data fusion for insider threat detection using residual networks. *3rd International Conference on Electronics, Communications and Information Technology (CECIT)*, 359–366. https://doi.org/10.1109/CECIT58139.2022.00069

Glasser, J., & Lindauer, B. (2013). Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data. *2013 IEEE Security and Privacy Workshops*, 98–104. https://doi.org/10.1109/SPW.2013.37

Haidar, D., & Gaber, M. (2018). Adaptive One-Class Ensemble-based Anomaly Detection: An Application to Insider Threats. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–9. https://doi.org/10.1109/IJCNN.2018.8489107

Han, X., Xu, D., Yuan, S., & Wu, X. (2022). Few-shot Anomaly Detection and Classification Through Reinforced Data Selection. *2022 IEEE International Conference on Data Mining (ICDM)*, 963–968. https://doi.org/10.1109/ICDM54844.2022.00115

He, W., Wu, X., Wu, J., Xie, X., Qiu, L., & Sun, L. (2021). Insider Threat Detection Based on User Historical Behavior and Attention Mechanism. *IEEE Sixth International Conference on Data Science in Cyberspace (DSC)*, 564–569. https://doi.org/10.1109/DSC53577.2021.00089

Igbe, O., & Saadawi, T. (2018). Insider Threat Detection using an Artificial Immune system Algorithm. *9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 297–302. https://doi.org/10.1109/UEMCON.2018.8796583

Jang, M., Ryu, Y., Kim, J. -S., & Cho, M. (2020). Against Insider Threats with Hybrid Anomaly Detection with Local-Feature Autoencoder and Global Statistics (LAGS). *IEICE Transactions on Information and Systems, 103*(4), 888–891. https://doi.org/10.1587/transinf.2019EDL8180

Jiang, J., Chen, J., Gu, T., Choo, K. -K. R., Liu, C., Yu, M., Huang, W., & Mohapatra, P. (2019). Anomaly Detection with Graph Convolutional Networks for Insider Threat and Fraud Detection. 2019 IEEE Military Communications Conference (MILCOM), 109–114. https://doi.org/10.1109/MILCOM47813.2019.9020760

Jiang, J., Chen, J., Gu, T., Choo, K. -K. R., Liu, C., Yu, M., Huang, W., & Mohapatra, P. (2019). Warder: Online Insider Threat Detection System Using Multi-Feature Modeling and Graph-Based Correlation. *2019 IEEE Military Communications Conference (MILCOM)*, 1–6. https://doi.org/10.1109/MILCOM47813.2019.9020931

Jiang, J., Chen, J., Choo, K. -K. R., Liu, K., Liu, C., Yu, M., & Mohapatra, P. (2018). Prediction and Detection of Malicious Insiders' Motivation Based on Sentiment Profile on Webpages and Emails. *2018 IEEE military communications conference (MILCOM)*, 1–6. https://doi.org/10.1109/MILCOM.2018.8599790

Jovanovic, L., Kaljevic, J., Zivkovic, M., Bacanin, N., Antonijevic, M., & Cajic, M. (2024). Insider Threat Identification From Accessed Website Content Optimized by Modified Metaheuristic. *2024 International Conference on Circuit, Systems and Communication (ICCSC)*, 1–6. https://doi.org/10.1109/ICCSC62074.2024.10617256

Jurišić, M., Tomičić, I., & Grd, P. (2023). User behavior analysis for detecting compromised user accounts: A review paper. *Cybernetics and Information Technologies*, *23*(3), 102–113. https://doi.org/10.2478/cait-2023-0027

Jyosthna, P. M., & Reddy, K. T. (2021). Threat Analysis using N-median Outlier Detection Method with Deviation Score. *International Journal of Advanced Computer Science and Applications*, *12*(8), 568–575. https://doi.org/10.14569/IJACSA.2021.0120866

Kan, X., Fan, Y., Zheng, J., Kudreyko, A., Chi, C.-H., Song, W., & Tregubova, A. (2023). User-level malicious behavior analysis model based on the NMF-GMM algorithm and ensemble strategy. *Nonlinear Dynamics, 111*(22), 21391–21408. https://doi.org/10.1007/s11071-023-08954-1

Kim, D. -W., Shin, G. -Y., & Han, M. -M. (2023). Anomaly Detection Based on Discrete Wavelet Transformation for Insider Threat Classification. *Computer Systems Science and Engineering*, *46*(1), 153–164. https://doi.org/10.32604/csse.2023.034589

Kim, J., Park, M., Kim, H., Cho, S., & Kang, P. (2019). Insider threat detection based on user behavior modeling and anomaly detection algorithms. *Applied Sciences*, *9*(19), Article 4018. https://doi.org/10.3390/app9194018

Kotb, H. M., Gaber, T., Aljanah, S., Zawbaa, H. M., & Alkhathami, M. (2025). A novel deep synthesis-based insider intrusion detection (DS-IID) model for malicious insiders and AI-generated threats. *Scientific Reports, 15*, Article 207. https://doi.org/10.1038/s41598-024-84673-w

Kumpf, K., Protic, M., Jovanovic, L., Cajic, M., Zivkovic, M., & Bacanin, N. (2024). Insider Threat Detection Using Bidirectional Encoder Representations From Transformers and Optimized AdaBoost Classifier. *2024 International Conference on Circuit, Systems and Communication (ICCSC)*, 1–6. https://doi.org/10.1109/ICCSC62074.2024.10616526

Le, D. C., & Nur Zincir-Heywood, A. (2019). Machine learning based Insider Threat Modelling and Detection. *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 1–6. https://ieeexplore.ieee.org/abstract/document/8717892

Le, D. C., Nur Zincir-Heywood, A., Khanchi, S., & Heywood, M. (2018). Benchmarking evolutionary computation approaches to insider threat detection. *GECCO '18: Proceedings of the Genetic and Evolutionary Computation Conference*, 1286–1293. https://doi.org/10.1145/3205455.3205612

Le, D. C., & Zincir-Heywood, A. (2018). Evaluating Insider Threat Detection Workflow Using Supervised and Unsupervised Learning. *2018 IEEE Security and Privacy Workshops (SPW)*, 270–275. https://doi.org/10.1109/SPW.2018.00043

Le, D. C., Zincir-Heywood, A., & Heywood, M. I. (2019). Dynamic Insider Threat Detection Based on Adaptable Genetic Programming. *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2579–2586. https://doi.org/10.1109/SSCI44817.2019.9003134

Le, D. C., & Zincir-Heywood, N. (2021a). Anomaly Detection for Insider Threats Using Unsupervised Ensembles. *IEEE Transactions on Network and Service Management*, *18*(2), 1152–1164. https://doi.org/10.1109/TNSM.2021.3071928

Le, D. C., & Zincir-Heywood, N. (2021b). Exploring anomalous behaviour detection and classification for insider threat identification. *International Journal of Network Management*, *31*(4), Article 2109. https://doi.org/10.1002/nem.2109

Le, D. C., Zincir-Heywood, N., & Heywood, M. (2020). Analyzing data granularity levels for insider threat detection using machine learning. *IEEE Transactions on Network and Service Management*, *17*(1), 30–44. https://doi.org/10.1109/TNSM.2020.2967721

Le, D. C., & Zincir-Heywood, N. (2020). Exploring Adversarial Properties of Insider Threat Detection. *2020 IEEE Conference on Communications and Network Security (CNS)*, 1–9. https://doi.org/10.1109/CNS48642.2020.9162254

Le, D. C., Zincir-Heywood, N., & Heywood, M. (2021). Training regime influences to semisupervised learning for insider threat detection. *2021 IEEE Security and Privacy Workshops (SPW)*, 13–18. https://doi.org/10.1109/SPW53761.2021.00010

Li, C., Li, F., Yu, M., Guo, Y., Wen, Y., & Li, Z. (2022). Insider Threat Detection Using Generative Adversarial Graph Attention Networks. *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2680–2685. https://doi.org/10.1109/GLOBECOM48099.2022.10001207

Li, D., Yang, L., Zhang, H., Wang, X., & Ma, L. (2022). Memory-augmented insider threat detection with temporal-spatial fusion. *Security and Communication Networks*, *2022*, 6418420. https: //doi.org/10.1155/2022/6418420

Li, D., Yang, L., Zhang, H., Wang, X., Ma, L., & Xiao, J. (2021). Image-based insider threat detection via geometric transformation. *Security and Communication Networks*, *2021*(1), Article 1777536. https://doi.org/10.1155/2021/1777536

Li, X., Li, L., Li, X., Cai, B., & Li, B. (2023). TGCN-DA: A Temporal Graph Convolutional Network with Data Augmentation for High Accuracy Insider Threat Detection. *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 1249–1257. https://doi.org/10.1109/TrustCom60117.2023.00170

Li, X., Li, X., Jia, J., Li, L., Yuan, J., Gao, Y., & Yu, S. (2023). A High Accuracy and Adaptive Anomaly Detection Model With Dual-Domain Graph Convolutional Network for Insider Threat Detection. *IEEE Transactions on Information Forensics and Security*, *18*, 1638–1652. https://doi.org/10.1109/TIFS.2023.3245413

Li, X., Li, L., Li, X., Cai, B., Jia, J., Gao, Y., & Yu, S. (2024). GMFITD: Graph Meta-Learning for Effective Few-Shot Insider Threat Detection, *IEEE Transactions on Information Forensics and Security, 19*, 7161–7175. https://doi.org/10.1109/TIFS.2024.3430106

Liu, A., Du, X., & Wang, N. (2018). Recognition of access control role based on convolutional neural network. 2018 *IEEE 4th International Conference on Computer and Communications (ICCC)*, 2069–2074. https://www.semanticscholar.org/paper/Recognition-of-Access-Control-Role-Based-on-Neural-Liu-Du/456bb1dc1164283781a1684d4752256c2ff5c98a

Liu, L., Chen, C., Zhang, J., De Vel, O., & Xiang, Y. (2019). Insider Threat Identification Using the Simultaneous Neural Learning of Multi-Source Logs. *IEEE Access, 7*, 183162–183176. https://doi.org/10.1109/ACCESS.2019.2957055

Liu, L., Chen, C., Zhang, J., De Vel, O., & Xiang, Y. (2020). Doc2vec-Based Insider Threat Detection through Behaviour Analysis of Multi-source Security Logs. *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 301–309. https://doi.org/10.1109/TrustCom50675.2020.00050

Liu, L., De Vel, O., Chen, C., Zhang, J., & Xiang, Y. (2018). Anomaly-Based Insider Threat Detection Using Deep Autoencoders. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 39–48. https://doi.org/10.1109/ICDMW.2018.00014

Liu, W., Gao, P., Zhang, H., Li, K., Yang, W., Wei, X., & Shu, J. (2025). Attributing Stealth Cyberattacks via Temporal Probabilistic Graph Neural Networks. *Journal of Computer Information Systems*, 1–15. https://doi.org/10.1080/08874417.2024.2446955

Manoharan, P., Yin, J., Wang, H. & Wang, Y.-H. (2024). Insider threat detection using supervised machine learning algorithms. *Telecommunication Systems, 87*(4), 899–915. https://doi.org/10.1007/s11235-023-01085-3

Matterer, J., & Lejeune, D. (2018). Peer group metadata-informed lstm ensembles for insider threat detection. *Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, 62–67. https://cdn.aaai.org/ocs/17698/17698-77703-1-PB.pdf

Mehmood, R., Singh, P., & Jeffery, Z. (2024). Unsupervised learning for insider threat prediction: A behavioral analysis approach.*2024 IEEE International Conference on Big Data (Big Data)*, 1–6. https://doi.org/10.1109/SIN63213.2024.10871807

Mittal, A., & Garg, U. (2023). Prediction and detection of insider threat detection using emails: A comparision. *2023 2nd international conference on electrical, electronics, information and communication technologies*. (*ICEEICT*). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10157297

Mladenovic, D., Antonijevic, M., Jovanovic, L., Simic, V., Zivkovic, M., Bacanin, N., Zivkovic, T., & Perisic, J. (2024). Sentiment classification for insider threat identification using metaheuristic optimized machine learning classifiers. *Scientific Reports*, *14*(1), Article 25731. https://doi.org/10.1038/s41598-024-77240-w

Molina, A., Goncalves, V., De Sousa, R., Pividal, M., Meneguette, R., & Filho, G. (2022). A lightweight unsupervised learning architecture to enhance user behavior anomaly detection. *2022 IEEE Latin-American Conference on Communications (LATINCOM)*. https://doi.org/10.1109/LATINCOM56090.2022.10000477

Mouyart, M., Medeiros Machado, G., & Jun, J.-Y. (2023). A Multi-Agent Intrusion Detection System Optimized by a Deep Reinforcement Learning Approach with a Dataset Enlarged Using a Generative Model to Reduce the Bias Effect. *Journal of Sensor and Actuator Networks*, *12*(5). https://doi.org/10.3390/jsan12050068

Naicker, T., & van Niekerk, B. (2021).Machine learning for insider threat detection. In: *3rd european conference on the impact of artificial intelligence and robotics, eciair 2021*, 122– 131. DOI: 10.34190/EAIR.21.036

Nicolaou, A., Shiaeles, S., & Savage, N. (2020). Mitigating insider threats using bio-inspired models. *Applied Sciences (Switzerland)*, *10*(15), Article 5046. https://doi.org/10.3390/app10155046

Orizio, R., Vuppala, S., Basagiannis, S., & Provan, G. (2020). Towards an explainable approach for insider threat detection: Constraint network learning. *International Conference on Intelligent Data Science Technologies and Applications. (IDSTA)*2, 42–49. https://doi.org/10.1109/IDSTA50958.2020.9264049

Padmavathi, G., Shanmugapriya, D., & Asha, S. (2022). A framework to detect the malicious insider threat in cloud environment using supervised learning methods. *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*, 354–358. https://doi.org/10.23919/INDIACom54597.2022.9763205

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.) 372*(71). https://doi.org/10.1136/bmj.n71

Palomares, I., Kalutarage, H., Huang, Y., Miller, P., McCausland, R., & McWilliams, G. (2017). A fuzzy multicriteria aggregation method for data analytics: Application to insider threat monitoring. *IFSA-SCIS 2017: Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems*, 1-6. https://doi.org/10.1109/IFSA-SCIS.2017.8023360

Paul, S., & Mishra, S. (2020). LAC: LSTM AUTOENCODER with Community for Insider Threat Detection. *ICBDR '20: Proceedings of the 4th International Conference on Big Data Research*, 71–77. https://doi.org/10.1145/3445945.3445958

Peccatiello, R. B., Gondim, J. J. C., & Garcia, L. P. F. (2023). Applying One-Class Algorithms for Data Stream-Based Insider Threat Detection. *IEEE Access*, 70560–70573. https://doi.org/10.1109/ACCESS.2023.3293825

Qi, Y., An, N., Wang, Z., Yao, Y., Zhang, C., Zhu, Y., & Lu, Z. (2025). SiamEHGT: An Evolving Heterogeneous Graph Transformer for Insider Threat Detection based on Siamese Architecture. *28th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2354-2359. https://doi.org/10.1109/CSCWD64889.2025.11033609

Qi, Y., Yan, C., Wang, Z., Zhang, C., Liu, S., Lu, Z., & Jiang, B. (2025). ATHITD: Attention-based temporal heterogeneous graph neural network for insider threat detection. *Computers & Security, 157*, Article104185. https://doi.org/10.1016/j.cose.2025.104587

Rashid, T., Agrafiotis, I., & Nurse, J. R. (2016). A new take on detecting insider threats: Exploring the use of hidden markov models. *Proceedings of the 8th ACM CCS International workshop on managing insider security threats*, 47–56. https://doi.org/10.1145/2995959.2995964

Rauf, U., Shehab, M., Qamar, N., & Sameen, S. (2019). Bio-inspired approach to thwart against insider threats: An access control policy regulation framework. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 289*, 39–57. https://doi.org/10.1007/978-3-030-24202-2_4

Rauf, U., Wei, Z., & Mohsen, F. (2023). Employee Watcher: A Machine Learning-based Hybrid Insider Threat Detection Framework. *2023 7th Cyber Security in Networking Conference (CSNet)*, 39–45. https://doi.org/10.1109/CSNet59123.2023.10339777

Roberts, S., Holodnak, J., Nguyen, T., Yuditskaya, S., Milosavljevic, M., & Streilein, W. (2016). A model-based approach to predicting the performance of insider threat detection systems. *Proceedings of the 2016 IEEE Symposium on Security and Privacy Workshops (SPW)*, 314–323. https://doi.org/10.1109/SPW.2016.14

Saaudi, A., Al-Ibadi, Z., Tong, Y., & Farkas, C. (2018). Insider threats detection using CNN–LSTM model. *Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 94–99. https://doi.org/10.1109/CSCI46756.2018.00025

Saaudi, A., Tong, Y., & Farkas, C. (2019). Probabilistic Graphical Model on Detecting Insiders: Modeling with SGD-HMM. *Proceedings of the 5th International Conference on Information Systems Security and Privacy*, 461–470. https://doi.org/10.5220/0007404004610470

Shanmugapriya, D., Dhanya, C., Asha, S., Padmavathi, G., & Suthisini, D. (2024). Cloud insider threat detection using deep learning models. *11th International Conference on Computing for Sustainable Global Development (INDIACom 2024)*, 434–438. https://doi.org/10.23919/INDIACom61295.2024.10498767

Sivakrishna, A. M., Mohan, R., Nair, N. S., & Rohini, V. (2025). Temporal Quantum Neural Networks for Insider Threat Detection. *6th International Conference on Recent Advances in Information Technology (RAIT)*, 1–6. https://doi.org/10.1109/RAIT65068.2025.11089064

Song, S., Gao, N., Zhang, Y., & Ma, C. (2024). BRITD: Behavior rhythm insider threat detection with time awareness and user adaptation. *Cybersecurity, 7*, Article 2. https://doi.org/10.1186/s42400-023-00190-9

Sridhar, A. P. (2025). Unauthorized Deep Learning Techniques for Identifying Insider Risks in Standardized Cybersecurity Databases. *2025 6th International Conference on Signal Processing and Communication (IC3)*, 1178-1183. https://doi.org/10.1109/IC363308.2025.10957272

Sun, D., Liu, M., Li, M., Shi, Z., Liu, P., & Wang, X. (2021). DeepMIT: A Novel Malicious Insider Threat Detection Framework based on Recurrent Neural Network. *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 335–341. https://doi.org/10.1109/CSCWD49262.2021.9437887

Tiwary, P., Madhubalan, A., Gautam, A., & Darji, R. (2024). Attention to Patterns is all you need for Insider threat detection. *2024 International Conference on Artificial Intelligence, Metaverse and Cybersecurity (ICAMAC)*, 1–6. https://doi.org/10.1109/ICAMAC62387.2024.10828982

Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017). Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams. *AI for Cyber Security Workshop at AAAI 2017*, 224–234. https://doi.org/10.48550/arXiv.1710.00811

Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017). Predicting User Roles from Computer Logs Using Recurrent Neural Networks. *The Thirty-First AAAI Conference on Artificial Intelligence*, 4993–4994. https://doi.org/10.1609/aaai.v31i1.11069

Vinay, M., Yuan, S., & Wu, X. (2024). Contrastive Learning for Fraud Detection from Noisy Labels. *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 1421–1434. https://doi.org/10.1109/ICDE60146.2024.00117

Vinay, M., Yuan, S., & Wu, X. (2022). Contrastive Learning for Insider Threat Detection. *Database Systems for Advanced Applications (DASFAA)*, 395–403. https://doi.org/10.1007/978-3-031-00123-9_32

Vinay, M. S., Yuan, S., & Wu, X. (2023). Robust Fraud Detection via Supervised Contrastive Learning. *2023 IEEE International Conference on Big Data (BigData)*, 1279–1288. https://doi.org/10.48550/arXiv.2308.10055

Wang, C., & Zhu, H. (2022). Wrongdoing Monitor: A Graph-Based Behavioral Anomaly Detection in Cyber Security. *IEEE Transactions on Information Forensics and Security*, *17*(5), 2703– 2718. https://doi.org/10.1109/TIFS.2022.3191493

Wang, X., Jiang, J., Wang, Y., Lv, Q., & Wang, L. (2023). UAG: User Action Graph Based on System Logs for Insider Threat Detection. *2023 IEEE Symposium on Computers and Communications (ISCC)*, 1027–1032. https://doi.org/10.1109/ISCC58397.2023.10218139

Wang, Y.-F., Guo, Y.-B., & Fang, C. (2023). A semantic-based method for analysing unknown malicious behaviours via hyper-spherical variational auto-encoders. *IET Information Security, 17*(2), 244–254. https://doi.org/10.1049/ise2.12088

Wei, R., Cai, L., Yu, A., & Meng, D. (2019). AGE: Authentication Graph Embedding for Detecting Anomalous Login Activities. *Information and Communications Security (ICICS)*, 341–356. https://doi.org/10.1007/978-3-030-41579-2_20

Wei, Y., Chow, K.-P., & Yiu, S.-M. (2021). Insider threat prediction based on unsupervised anomaly detection scheme for proactive forensic investigation. *Forensic Science International: Digital Investigation*, 2666-2817. https://doi.org/10.1016/j.fsidi.2021.301126

Wei, Z., Rauf, U., & Mohsen, F. (2024). E-Watcher: Insider threat monitoring and detection for enhanced security. *Ann. Telecommun*, 819–831. https://doi.org/10.1007/s12243-024-01023-7

Xiao, F., Chen, S., Chen, S., Ma, Y., He, H., & Yang, J. (2025). Sentinel: Insider threat detection based on multi-timescale user behavior interaction graph learning. *IEEE Transactions on Network Science and Engineering, 12*(2), 774–790. https://doi.org/10.1109/TNSE.2024.3519155

Xiao, J., Yang, L., Zhong, F., Wang, X., Chen, H., & Li, D. (2023). Robust Anomaly-Based Insider Threat Detection Using Graph Neural Network. *IEEE Transactions on Network and Service Management, 20*(3), 3717–3733. https://doi.org/10.1109/TNSM.2022.3222635

Yang, W., Gao, P., Huang, H., Wei, X., Zhang, H., & Qu, Z. (2022). Advanced Persistent Threat Detection in Smart Grid Clouds Using Spatiotemporal Context-Aware Graph Embedding. *GLOBECOM 2022 - 2022 IEEE Global Communications*, 534–540. https://doi.org/10.1109/GLOBECOM48099.2022.10001486

Ye, X., Cui, H., Luo, F., Wang, J., Xiong, X., Zhang, W., Yu, J., & Zhao, W. (2025). Daily insider threat detection with hybrid TCN transformer architecture. *Scientific Reports, 15*(1), Article 28590. https://doi.org/10.1038/s41598-025-12063-x

Ye, X., Hong, S.-S., & Han, M.-M. (2020). Feature Engineering Method Using Double-Layer Hidden Markov Model for Insider Threat Detection. *International Journal of Fuzzy Logic and Intelligent Systems*, *20*(1), 17–25. https://doi.org/10.5391/IJFIS.2020.20.1.17

Yi, J., & Tian, Y. (2024). Insider Threat Detection Model Enhancement Using Hybrid Algorithms between Unsupervised and Supervised Learning. *Electronics*, *13*(5), Article 973. https://doi.org/10.3390/electronics13050973

Yuan, F., Cao, Y., Shang, Y., Liu, Y., Tan, J., & Fang, B. (2018). Insider threat detection with deep neural network. *Computational Science (ICCS)*, 43–54. https://doi.org/10.1007/978-3-319-93698-7_4

Yuan, F., Shang, Y., Liu, Y., Cao, Y., & Tan, J. (2020). Data Augmentation for Insider Threat Detection with GAN. *20 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, 632–638. https://doi.org/10.1109/ICTAI50040.2020.00102

Yuan, L.-P., Choo, E., Yu, T., Khalil, I., & Zhu, S. (2021). Time-Window Based Group-Behavior Supported Method for Accurate Detection of Anomalous Users. *51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 250–262. https://doi.org/10.1109/DSN48987.2021.00038

Yuan, S., Zheng, P., Wu, X., & Tong, H. (2020). Few-shot Insider Threat Detection. *CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2289–2292. https://doi.org/10.1145/3340531.3412161

Zerhoudi, S., Granitzer, M., & Garchery, M. (2020). Improving Intrusion Detection Systems using Zero-Shot Recognition via Graph Embeddings. *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 790–797. https://doi.org/10.1109/COMPSAC48688.2020.0-165

Zhang, B., Ma, Y., Ge, Q., Wu, Q., Liu, J., & Zhang, W. (2024). An Insider Risk Detection of EMS in Energy Storage Systems Based on Random Walks and Heterogeneous Graph Embedding. *2024 14th Asian Control Conference (ASCC)*, 1161–1166. https://doi.org/10.1109/BigDataSecurityHPSCIDS54978.2022.00013

Zhang, C., Wang, S., Zhan, D., Yu, T., Wang, T., & Yin, M. (2021). Detecting Insider Threat from Behavioral Logs Based on Ensemble and Self-Supervised Learning. *Security and Communication Networks 2021*(1), Article 4148441. https://doi.org/10.1155/2021/4148441

Zhang, M., Liang, X., Tian, F., Yang, Y., Yu, H., & Li, B. (2024). LLM4ITD: Insider Threat Detection with Fine-Tuned Large Language Models. *2024 International Conference on Interactive Intelligent Systems and Techniques (IIST)*, 236–241. https://doi.org/10.1109/IIST62526.2024.00017

Zheng, C., Hu, W., Li, T., Liu, X., Zhang, J., & Wang, L. (2022). An Insider Threat Detection Method Based on Heterogeneous Graph Embedding. *2022 IEEE 8th Intl Conference on Big Data Security on Cloud (BigDataSecurity)*, 11–16. https://doi.org/10.1109/BigDataSecurityHPSCIDS54978.2022.00013

Zheng, L., Birge, J., Wu, H., Zhang, Y., & He, J. (2025). Cluster Aware Graph Anomaly Detection. *WWW '25: The ACM Web Conference 2025*, 1771–1782. https://doi.org/10.1145/3696410.3714575

Zheng, P., Yuan, S., & Wu, X. (2021). Using Dirichlet Marked Hawkes Processes for Insider Threat Detection. *Digital Threats: Research and Practice (DTRAP)*, *3*(1), 1–19. https://doi.org/10.1145/3457908

Zhou, S., Wang, L., Yang, J., & Zhan, P. (2022). SITD: Insider Threat Detection Using Siamese Architecture on Imbalanced Data. *IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 245–250. https://doi.org/10.1109/CSCWD54268.2022.9776049

Zhu, D., Huang, X., Li, N., Sun, H., Liu, M., & Liu, J. (2022). RAP-Net: A Resource Access Pattern Network for Insider Threat Detection. *2022 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN55064.2022.9892183

Zhu, X., Dong, J., Qi, J., Zhou, Z., Dong, Z., Sun, Y., & Wang, M. (2024). An Adversarial Autoencoder Based Unsupervised Insider Threat Detection Scheme for Multisource Logs. *IEEE Transactions on Industrial Informatics*, *20*(9), 10954–10965. https://doi.org/10.1109/TII.2024.3393491

Zou, S., Sun, H., Xu, G., & Quan, R. (2020). Ensemble Strategy for Insider Threat Detection from User Activity Logs. *Computers, Materials & Continua*, *65*(2), 1321–1334. https://doi.org/10.32604/cmc.2020.09649

Zuo, X., Yan, F., Hou, B., Chen, Z., & Guo, Y. (2022). Insider threat detection model of power system based on lstm-attention. *UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science*, *84*(2), 319–336. https://doi.org/10.1088/1742-6596/1994/1/012021

ABOUT THE AUTHORS

Marko Jurišić, doctoral candidate, Faculty of Organization and Informatics, University of Zagreb. Email: marko.jurisic@foi.unizg.hr

Igor Tomičić, Associate Professor, Faculty of Organization and Informatics, University of Zagreb. Email: igor.tomicic@foi.unizg.hr