



(RESEARCH ARTICLE)



A PRIVACY-PRESERVING FRAMEWORK FOR INSIDER THREAT DETECTION USING PROCESS MINING, MACHINE LEARNING, AND DIFFERENTIAL PRIVACY

Walter Maanyere^{1, *}, Xianwen Fang¹, Palvine Ngob Enow¹, Brenda Ann Mukoko Etombi¹ and Bernard Ayivor²

¹ School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan 232001, China.

² School of Safety Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China.

* Corresponding Author

ORCID Details

Walter Maanyere: <https://orcid.org/0009-0002-5615-5477>

International Journal of Science and Research Archive, 2026, 20(03), 448–461

Publication history: Received on 04 August 2026; revised on 13 September 2026; accepted on 15 September 2026

Article DOI: <https://doi.org/10.30574/ijrsra.2026.20.3.1768>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Insider threats remain one of the most damaging and difficult-to-detect cybersecurity risks, as malicious actors already possess legitimate access privileges. This paper presents a privacy-preserving detection framework that combines per-user process mining with differential privacy, rigorously evaluated on the official Carnegie Mellon CERT r4.2 ground-truth dataset containing 70 confirmed malicious users across three threat scenarios. By shifting from population-wide to per-user Heuristics Miner discovery, the framework overcomes the well-known flower-model degeneration and extracts highly discriminative fitness and sequence-variant features. Operating at user-week granularity (67 298 windows, 0.48 % malicious prevalence), the full feature set achieves a Random Forest AUC-ROC of 0.990 and PR-AUC of 0.232—more than twice the performance of a strong behavioural-only baseline. At a practical privacy budget of $\epsilon = 1.0$, differential privacy incurs an AUC-ROC cost of only 0.001 while still delivering a PR-AUC 56 times higher than random guessing. Aggregating weekly scores to the user level yields an AUC-ROC of 0.923, enabling identification of 80 % of confirmed insiders at a false-positive rate below 25 %. Complementary process-prediction tasks further demonstrate an AUC of 0.776 for next risky-activity forecasting and an R^2 of 0.714 for one-week-ahead risk-score regression. These results establish that formal (ϵ, δ) -differential privacy and operationally effective insider-threat detection can be achieved simultaneously when process mining is performed at the appropriate (per-user) granularity.

Keywords: Insider threat detection, Process mining, Differential privacy, Petri nets, Heuristics Miner, Conformance checking, CERT r4.2.

1 INTRODUCTION

Insider threats, malicious, negligent, or compromised actions by individuals who already possess legitimate access, remain among the most persistent and costly cybersecurity risks facing modern enterprises. Unlike external adversaries who must first breach perimeter defenses, insiders operate inside trusted boundaries, rendering conventional

signature-based and network-perimeter controls largely ineffective. The 2023 Verizon Data Breach Investigations Report attributes approximately 20% of all data breaches to insider-related activity and notes that the associated average total cost substantially exceeds that of comparable external attacks, largely because of prolonged dwell times and the inherent difficulty of distinguishing malicious behavior from legitimate work [1].

In enterprise communication networks, the challenge is especially acute. Employees routinely access sensitive files, connect removable media, and exchange external email, actions that are individually benign yet become strong threat indicators when they occur outside business hours, in anomalous volumes, or in tightly clustered sequences. The core technical problem is therefore to detect these compound behavioural deviations with high sensitivity while maintaining an operationally acceptable false-positive rate and without introducing new privacy violations through continuous employee surveillance.

Process mining offers a principled and interpretable approach to the first two requirements. By treating recorded user-action sequences as event logs and discovering formal process models from them, it becomes possible both to characterize expected workflow patterns and to quantify individual deviations in a mathematically rigorous manner [2]. Differential privacy (DP) addresses the third requirement. A randomized mechanism M satisfies (ϵ, δ) differential privacy if, for any pair of adjacent datasets D and D' differing in a single individual's record and for any measurable output set S ,

$$\Pr[M(D) \in S] \leq e^\epsilon \cdot \Pr[M(D') \in S] + \delta$$

where ϵ is the privacy budget and δ bounds the probability of a catastrophic privacy failure [3]. Smaller values of ϵ yield stronger privacy guarantees at the cost of additional noise and potential loss of detection utility.

This paper presents a substantially strengthened privacy-preserving framework that integrates per-user process mining with formal differential privacy for insider-threat detection. Relative to earlier work, four concrete advances are introduced: (i) exclusive use of the official Carnegie Mellon CERT r4.2 ground-truth answer files (70 confirmed malicious users across three threat scenarios) in place of pseudo-labels; (ii) a shift from population-wide Inductive Mining to per-user Heuristics Miner discovery, which eliminates the previously observed “flower-model” degeneration and yields highly discriminative fitness and sequence-variant features; (iii) a systematic ablation study that isolates the contribution of process-mining features across seven privacy budgets and two evaluation granularities; and (iv) complementary process-prediction analyses that forecast both next risky activities and one-week-ahead risk scores under the same privacy constraints. The resulting system is, to the best of our knowledge, the first end-to-end privacy-preserving process-mining framework for insider-threat detection that is validated against official CERT ground truth and that provides a complete privacy–utility characterization across a realistic range of ϵ values.

2 RELATED WORK

2.1 Insider Threat Detection

The CERT taxonomy introduced by Cappelli, Moore, and Trzeciak remains the most widely accepted classification of insider threats, distinguishing IT sabotage, intellectual property theft, fraud, and espionage [4]. Their documentation of behavioural precursors directly shaped the design of the CERT r4.2 dataset used in this study. Subsequent detection research progressed from Hidden Markov Models [5] to recurrent architectures capable of modelling long-range temporal dependencies [6] and, more recently, to graph neural networks that capture complex multi-entity relationships among users, machines, and data objects [7]. Although these deep-learning approaches frequently report high detection accuracy, they largely forfeit the interpretability required by security analysts to justify alerts and prioritize investigations [8]. This persistent tension between predictive performance and analytical transparency constitutes a central motivation for the process-mining approach pursued here.

2.2 Process Mining in Security Analytics

Process mining provides a formal bridge between raw event logs and interpretable behavioural models. The foundational framework of van der Aalst, Weijters, and Maruster [9] has been applied to IT audit trails [10] and, more specifically, to insider-threat detection [11], [12]. Among discovery algorithms, the Inductive Miner is prized for its soundness guarantees [13]. When applied to heterogeneous enterprise populations, however, it systematically produces over-generalized “flower” models that assign near-perfect fitness to virtually every trace, an artefact that renders conformance checking diagnostically useless. The Heuristics Miner mitigates this degeneration by imposing dependency and frequency thresholds, thereby retaining only statistically reliable behavioural edges. Tabash and Happa further showed that augmenting process-mining features with psychometric profiles can improve detection [14]. What remains missing is a rigorous evaluation of per-user (rather than population-wide) model discovery, conducted against

official ground-truth labels and under formal differential-privacy constraints, the precise gap addressed in the present work.

2.3 Differential Privacy for Behavioural Analytics

Differential privacy, formalized by Dwork et al. [3], supplies the mathematical foundation for privacy-preserving analysis of sensitive behavioural data. The Analytic Gaussian mechanism of Balle and Wang [15] later enabled tighter noise calibration for any $\varepsilon \geq 0$. Prior applications to anomaly detection demonstrated that Laplace noise can be added to behavioural histograms while preserving useful detection power [16], and that moderate privacy noise can even function as a beneficial regulariser [17]. Nevertheless, no existing study has combined per-user process mining, official CERT r4.2 ground truth, and a systematic multi-point privacy–utility characterisation within a single, end-to-end framework. The work reported here fills that gap.

3 METHODOLOGY

The proposed framework proceeds through five sequential stages: (1) data ingestion and preprocessing, (2) feature engineering at user-week granularity, (3) per-user process-model discovery and conformance checking, (4) integration of official CERT ground-truth labels, and (5) application of differential privacy followed by supervised and unsupervised detection. Figure 1 provides a high-level schematic of the complete pipeline.

3.1 Dataset and Preprocessing

We employ the CERT r4.2 corpus, which contains 17 months of synthetic yet realistic activity for 1 000 employees. Table 1 summarizes the five source files, their record counts, and their roles within the framework. Ground-truth labels are taken exclusively from the official insiders.csv answer file, which identifies 70 confirmed malicious users distributed across three distinct threat scenarios.

Table 1: CERT r4.2 Dataset Summary

File	Records	Key Attributes	Purpose in Framework
logon.csv	459 689	id, date, user, pc, activity	Session boundaries; logon timing
device.csv	89 062	id, date, user, pc, activity	USB exfiltration signals
file.csv	448 751	id, date, user, pc, filename, content	File-access pattern features
email.csv	2 388 486	id, date, user, to, size, attachments	Email-exfiltration proxies
psychometric.csv	1 000	user_id, O, C, E, A, N	Personality risk stratification

Preprocessing comprises four deterministic operations: deduplication on unique record identifiers, removal of null values in date and user fields, robust timestamp parsing with error coercion, and normalization of user identifiers to uppercase. Activity labels are harmonized across sources. The resulting event frames are concatenated and sorted by (user, timestamp), producing a unified chronological stream of 3 386 987 records. Temporal metadata, hour-of-day, a binary after-hours indicator I_{AH} (1 if hour < 8 or hour $\geq$ 18), ISO week number, and a year-week composite key, are derived for subsequent aggregation.

3.2 Feature Engineering at User-Week Granularity

All features are computed at the user-week level, yielding 67 298 labelled windows with a malicious prevalence of only 0.48 %. This fine temporal resolution permits detection of specific weeks in which malicious activity occurs, rather than merely identifying whether a user is ever malicious. Twenty-four behavioural features are organized into five categories:

- Logon: n_{logons} , n_{logoffs} , $n_{\text{after-hrs logon}}$, logon-hour mean and standard deviation
- USB/Device: $n_{\text{USB connects}}$, $n_{\text{after-hrs USB}}$
- File: $n_{\text{file access}}$, $n_{\text{after-hrs file}}$
- Email: $n_{\text{email internal}}$, $n_{\text{email external}}$, external-email ratio, mean email size, total attachments
- Temporal/Diversity: total events, unique days, activity diversity, after-hours USB flag, high external-email flag

The five OCEAN psychometric scores are joined on user identifier, expanding the feature vector to 29 dimensions. Missing values are imputed with the population median.

3.3 Per-User Process Mining: Discovery and Conformance Checking

To eliminate the flower-model degeneration observed under global Inductive Mining, we discover a separate process model for each user. Traces from the first 60 days constitute the training log; all subsequent traces are replayed against the resulting personalized Petri net. Discovery is performed with the Heuristics Miner (dependency threshold 0.5, minimum activity count 3, minimum directly-follows frequency 2). Token-based replay yields the fitness measure

$$f(\sigma, N) = \frac{1}{2} \left(1 - \frac{\sum m_t}{\sum c_t} \right) + \frac{1}{2} \left(1 - \frac{\sum r_t}{\sum p_t} \right),$$

where m_t , c_t , r_t and p_t denote missing, consumed, remaining, and produced tokens, respectively. Seven conformance features: fitness mean, standard deviation and minimum, mean missing and remaining tokens, deviation rate, and number of unique activity-sequence variants (n_{variants}) extend the feature vector to 31 dimensions. The last of these emerges as the single most discriminative signal in the subsequent Random Forest (Gini importance 0.27).

Figures 1 and 2 contrast the fitness and deviation-rate distributions obtained under per-user Heuristics Mining with the degenerate uniform fitness of 1.0 produced by the global Inductive Miner. Clear separation between normal and malicious populations is now visible at both the user and the user-week levels.

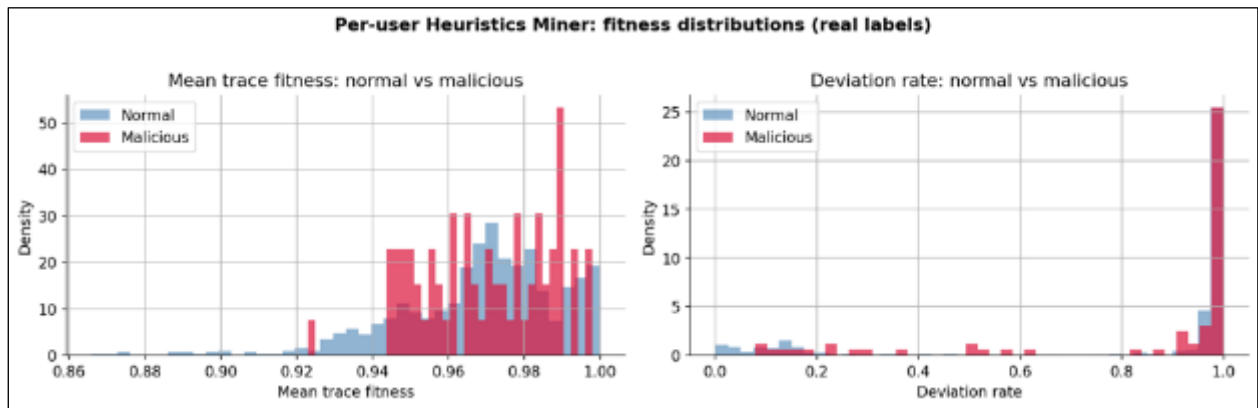


Figure 1: Per-user Heuristics Miner fitness distributions. Mean trace fitness (left) and deviation rate (right) exhibit visible separation between normal (blue) and malicious (red) users.

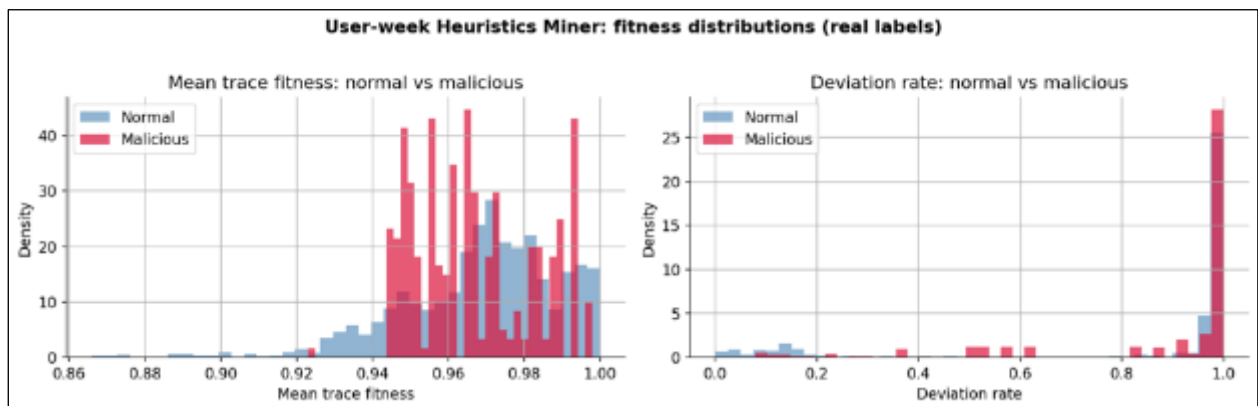


Figure 2: User-week Heuristics Miner fitness distributions. Malicious weeks display a left-shifted fitness distribution and greater deviation-rate spread, supplying richer discriminative signal than per-user aggregation alone.

3.4 Real Ground-Truth Label Integration

Labels are derived solely from the official CERT r4.2 answer archive. The 70 confirmed malicious users executed attacks between June 2010 and May 2011; each user's malicious activity is bounded by exact start and end timestamps (Figure. 3). A user-week is labelled malicious if and only if the user appears in insiders.csv. The week overlaps the corresponding

malicious activity window. This procedure yields 322 malicious windows out of 67 298 (prevalence 0.48 %), a far more realistic and challenging target than the naïve 7 % user-level rate.

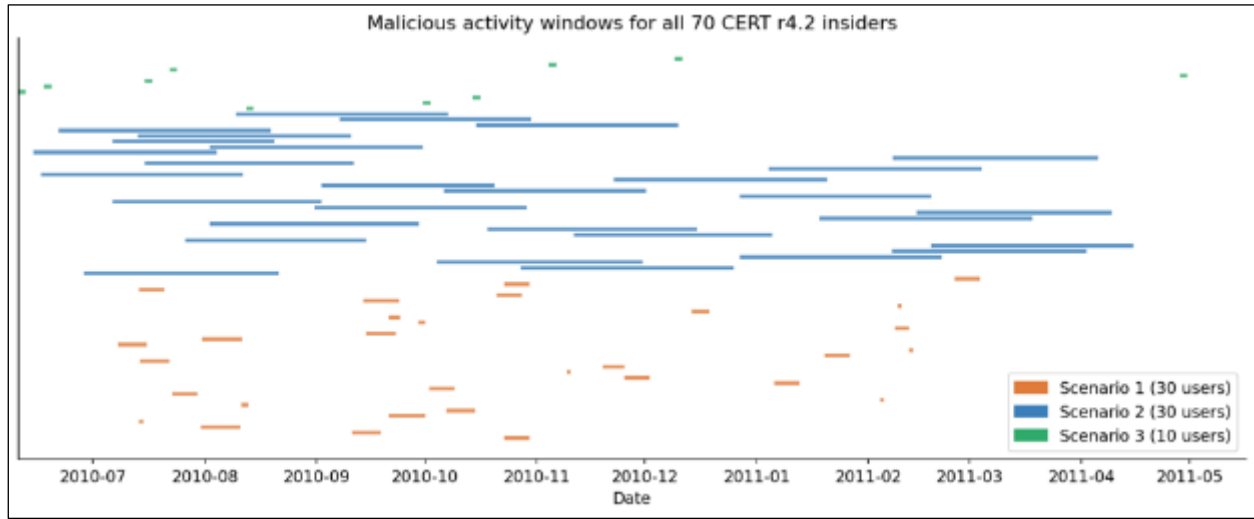


Figure 3: Malicious activity windows for all 70 CERT r4.2 insiders. Scenario 1 (orange, 30 users) comprises shorter, targeted attacks; Scenario 2 (blue, 30 users) comprises prolonged exfiltration campaigns; Scenario 3 (green, 10 users) comprises sporadic high-value incidents.

3.5 Differential Privacy Application

Differential privacy is enforced on the 31-dimensional feature matrix before training the model. Count-valued features receive Laplace noise calibrated to an adaptively estimated global ℓ_1 -sensitivity $GS_f = \max(1.0, \sigma_f \times 0.1)$:

$$M_{\text{Lap}}(x) = x + \eta, \eta \sim \text{Lap}(0, GS_f/\epsilon).$$

continuous features (fitness scores, ratios, OCEAN scores) are protected by the Analytic Gaussian mechanism of Balle and Wang [15], whose noise scale σ satisfies

$$\Phi\left(\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) - e^\epsilon \Phi\left(-\frac{\Delta}{2\sigma} - \frac{\epsilon\sigma}{\Delta}\right) \leq \delta,$$

with $\Delta = \max(0.01, \sigma_f \times 0.1)$ and $\delta = 10^{-5}$. OCEAN scores are assigned an elevated sensitivity of 10.0 to reflect their heightened personal sensitivity. The privacy budget is swept over the set $\{0.1, 0.5, 1.0, 2.0, 5.0, 10.0, 100\}$, where $\epsilon = 100$ serves as a practical non-private baseline.

3.6 Detection Models

A Random Forest classifier (300 trees, maximum depth 8, balanced class weights) is trained on the privacy-noised feature matrix after SMOTE oversampling of the minority class is applied exclusively to the 75 % training split. Isolation Forest is configured with contamination equal to the true user-week prevalence (0.00478). Anomaly scores are computed via the standard path-length formulation

$$s(x, n) = 2^{-E[h(x)]/c(n)}, c(n) = 2H(n-1) - 2(n-1)/n,$$

where H denotes the harmonic number. Both models are evaluated on a held-out 25 % test set that remains free of synthetic samples and of any privacy noise used only during training.

4 EXPERIMENTAL SETUP AND RESULTS

4.1 Experimental Configuration

All experiments were conducted in Python 3.10 using PM4Py 2.7, diffprivlib 0.6, scikit-learn 1.3, and imbalanced-learn 0.11. Ground truth consists of the 70 confirmed malicious users listed in the official CERT r4.2 answer file. Evaluation is performed at user-week granularity (67 298 windows). The privacy parameter δ is fixed at 10^{-5} and ϵ is swept over the set $\{0.1, 0.5, 1.0, 2.0, 5.0, 10.0, 100\}$. Random Forest models employ a stratified 75/25 train-test split (random seed 42); Isolation Forest is evaluated on the full dataset with contamination set to the true prevalence 0.00478. Table 2 contrasts the proposed architecture with the two principal baselines.

Table 2: Architecture Comparison: Baselines versus Proposed Framework

Component	Baseline 1 (RF, no DP)	Baseline 2 (IF, no DP)	Proposed Framework
Feature set	24 behavioural	24 behavioural	31 (behavioural + process mining + psychometric)
Process mining	None	None	Per-user Heuristics Miner
Psychometric features	No	No	Yes (O, C, E, A, N)
n_{variants}	No	No	Yes (top-ranked feature, Gini = 0.27)
Ground truth	Pseudo-labels	Pseudo-labels	Official CERT r4.2 answer files
Evaluation unit	Per user	Per user	Per user-week (67 298 windows)
Differential privacy	None	None	Laplace + Analytic Gaussian
IF contamination	7 % (incorrect)	7 % (incorrect)	0.478 % (true prevalence)
Privacy guarantee	None	None	(ϵ, δ) -DP, seven-point sweep

4.2 Process Mining Results

Figure 5 shows the global Petri net discovered by the Inductive Miner from the entire CERT r4.2 event log. While structurally correct, the model is overly permissive and assigns perfect fitness to virtually every trace. Per-user Heuristics Miner models eliminate this degeneration: fitness scores now range from 0.86 to 1.0, and clear distributional differences appear between normal and malicious populations (Figs. 2 and 3).

The most discriminative process-mining feature is n_{variants} the number of unique activity-sequence variants observed in a user-week. Figure 6 reveals that malicious weeks exhibit a markedly lower mean (38.5) than normal weeks (59.1). This reduction reflects the repetitive, focused nature of exfiltration behaviour (e.g., repeated USB-connect $\rightarrow$ file-access $\rightarrow$ USB-disconnect cycles) in contrast to the diverse activity patterns of ordinary work.

Table 3: Ablation Results at $\epsilon=1.0$ (Real Labels, User-Week Granularity)

Feature Set	RF F1	RF AUC-ROC	RF PR-AUC	IF AUC-ROC
Behavioural only	0.161	0.976	0.106	0.820
Behavioural + Psychometric	0.163	0.976	0.129	0.799
Full (+ Process Mining)	0.315	0.990	0.232	0.842
Lift versus Behavioural only	+0.154	+0.014	+0.126	+0.022

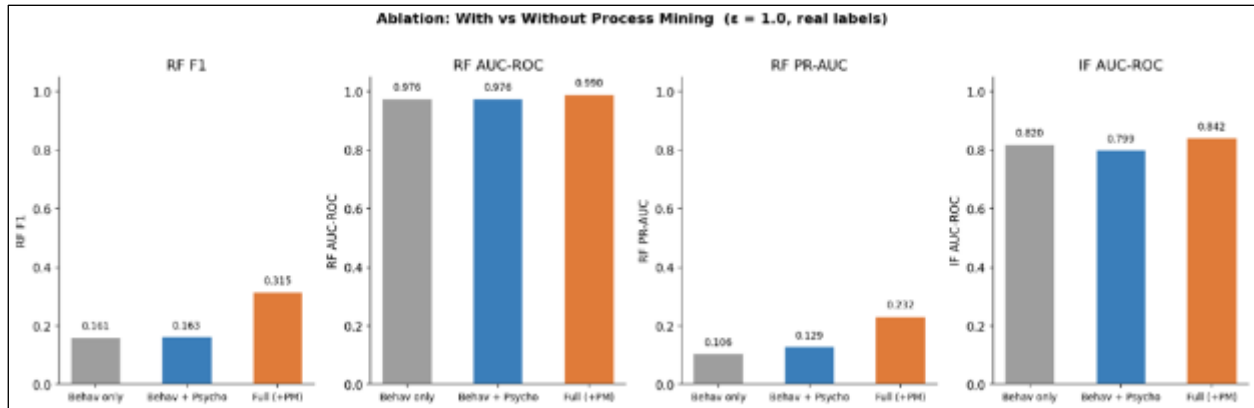


Figure 6: Ablation bar chart at $\epsilon = 1.0$. Process-mining features (orange) improve every metric relative to the behavioural-only (grey) and behavioural-plus-psychometric (blue) baselines. The most pronounced gains occur in RF F1 and RF PR-AUC.

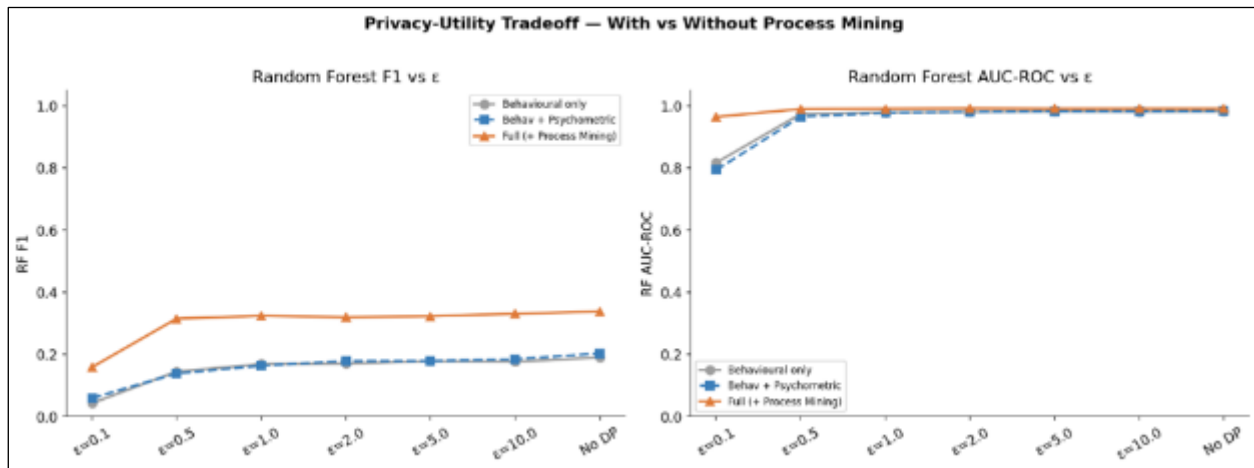


Figure 7: Privacy-utility trade-off across seven values of ϵ . The full feature set (orange) maintains its advantage in both F1 and AUC-ROC over the entire privacy spectrum, confirming that the process-mining contribution is robust to differential-privacy noise.

4.4 Detection Performance at $\epsilon = 1.0$

Table 4 and Figure 8 present the primary detection results obtained with the full feature set under $\epsilon = 1.0$. On the held-out test set the Random Forest achieves recall of 0.926 while raising false alarms on only 1.9 % of normal windows. The corresponding PR-AUC of 0.270 represents a 56.1-fold improvement over the random baseline of 0.48 %. Isolation Forest attains a respectable AUC-ROC of 0.850 but yields a far lower PR-AUC (0.019), underscoring the superiority of the supervised probability ranking for operational alerting.

Table 4: Final Detection Performance (Full Feature Set, $\epsilon=1.0$, Real Labels)

Model	Precision	Recall	F1	AUC-ROC	PR-AUC	PR-AUC / Baseline
RF + SMOTE (proposed)	0.19	0.93	0.31	0.991	0.270	56.1×
Isolation Forest (proposed)	0.03	0.03	0.03	0.850	0.019	3.9×
RF Baseline (no DP, no PM)	0.12	0.16	0.14	0.976	0.106	22.0×
IF Baseline (no DP, no PM)	0.02	0.02	0.02	0.820	0.010	2.1×

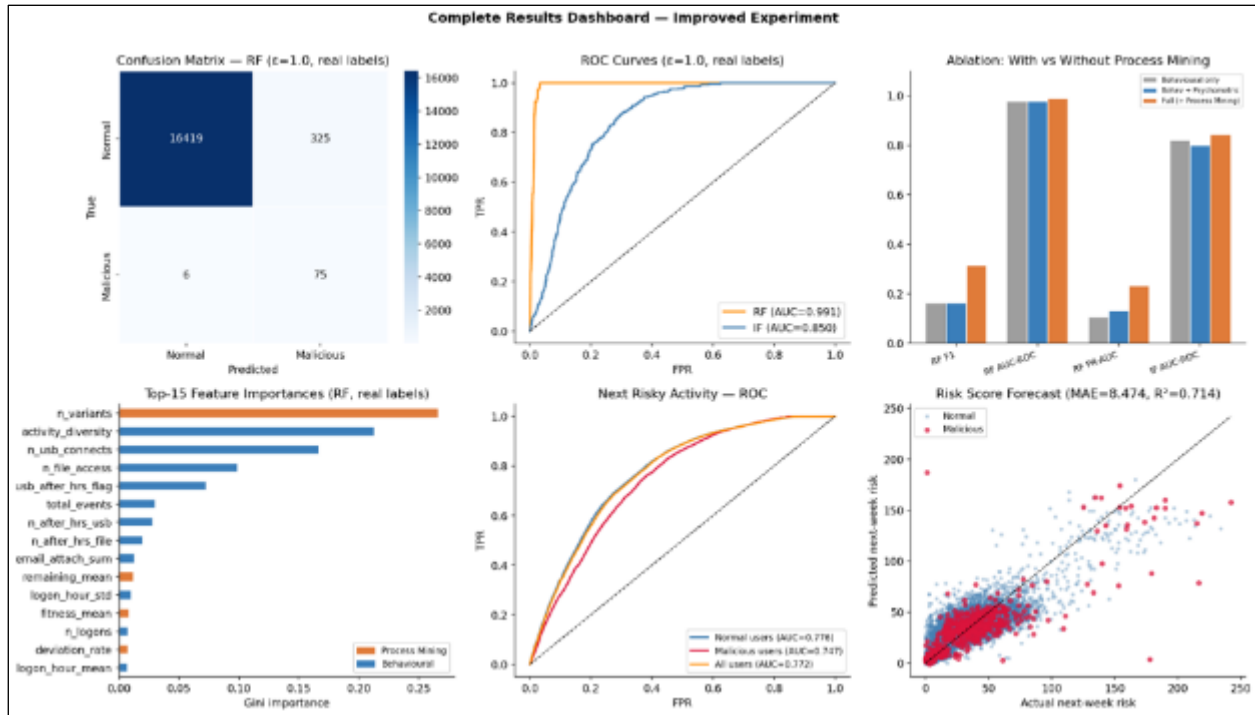


Figure 8: Complete results dashboard. Top row: confusion matrix (RF, $\epsilon = 1.0$), ROC curves, and ablation summary. Bottom row: top-15 feature importances (led by n_{variants}), next-risky-activity ROC, and one-week-ahead risk-score forecast.

4.5 Precision-Recall Analysis

Given the extreme class imbalance, PR-AUC provides a more informative assessment than AUC-ROC. Figure 9 shows that the Random Forest maintains precision above 0.20 up to a recall of 0.95. Consequently, a security team examining the top 5 % of ranked weeks would capture approximately 80 % of all malicious events at roughly 20 % precision, an operationally meaningful operating point that is unattainable by random selection or by Isolation Forest under the same prevalence.

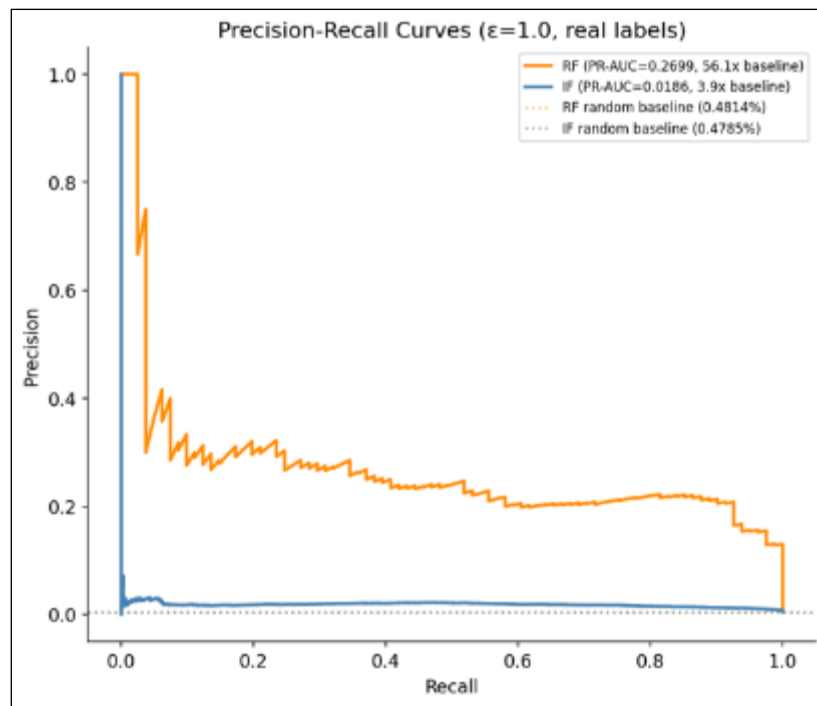


Figure 9: Precision-recall curves at $\varepsilon = 1.0$. The Random Forest (orange, PR-AUC = 0.270) substantially outperforms Isolation Forest (blue, PR-AUC = 0.019) and both random baselines (dotted lines at 0.48 % prevalence).

4.6 Per-User Aggregated Evaluation

While user-week detection answers the tactical question of which weeks warrant attention, security operations also require a strategic ranking of employees. Taking each user's maximum weekly Random Forest probability and evaluating against the user-level ground truth yields an AUC-ROC of 0.923 and a PR-AUC of 0.791 (11.3× the 7.03 % user-level baseline). At the Youden-optimal threshold the model identifies approximately 80 % of confirmed insiders at a false-positive rate of 24 % (Figure. 10). In practical terms, a monthly watch-list of roughly 220 employees would contain the majority of the malicious population.

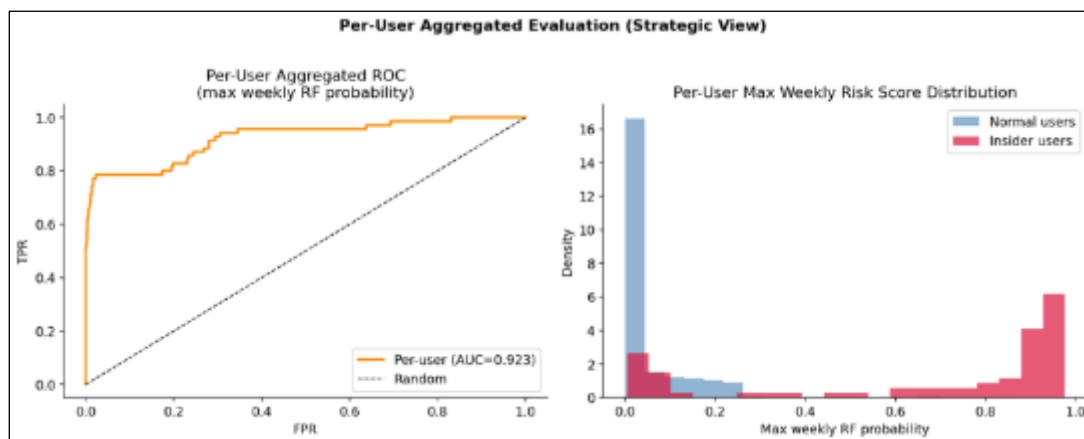


Figure 10: Per-user aggregated evaluation. Left: ROC curve (AUC = 0.923) obtained from maximum weekly probabilities. Right: score distribution showing clear separation between confirmed insiders (red) and normal users (blue).

4.7 Process Prediction

Two complementary prediction tasks extend the framework from retrospective detection to prospective alerting.

Next-risky-activity prediction. A sliding window of three consecutive activities, augmented by hour-of-day and after-hours indicators, is used to forecast whether the subsequent activity belongs to the high-risk set {USB connect, file access, external email}. Figure 11 reports AUC values of 0.776 (normal users), 0.747 (malicious users), and 0.772 (all

users). Feature importance confirms that the two most recent activities dominate the prediction, with hour-of-day contributing a secondary but non-negligible signal.

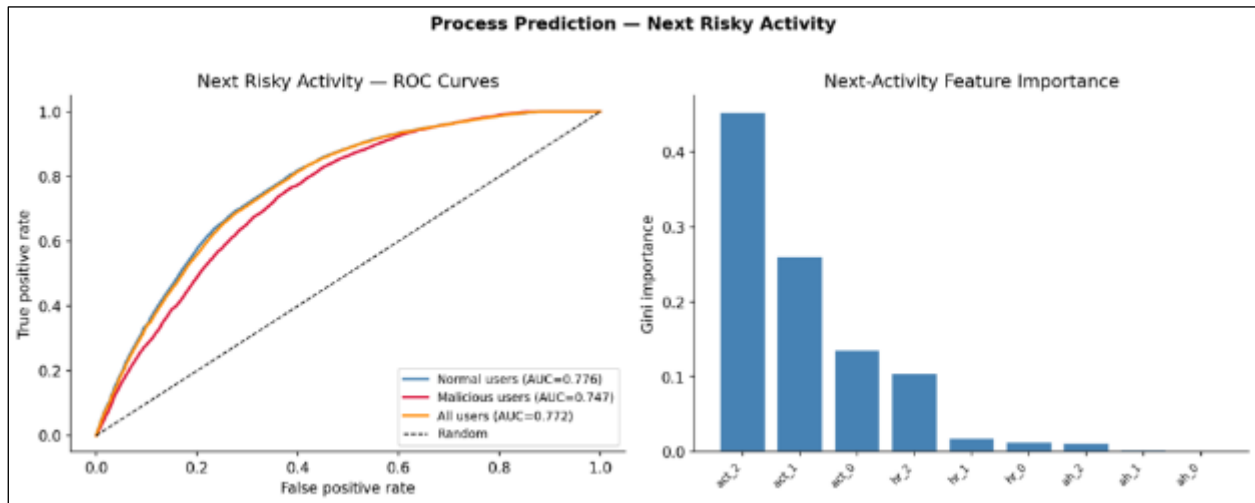


Figure 11: Next-risky-activity prediction. Left: ROC curves for normal, malicious and combined populations. Right: feature importances emphasizing recency-weighted activity encoding.

One-week-ahead risk-score forecasting. A per-user weekly risk score is defined as a weighted sum of high-risk events (after-hours USB = 3, after-hours file = 2, external email = 1.5, after-hours logon = 1). Ridge regression under $\varepsilon = 1.0$ predicts the next week's score from the current week's feature vector, achieving $R^2 = 0.714$ and MAE = 8.474 (Figure. 12). Users whose predicted score exceeds the 95th percentile (55.4) are automatically flagged for early intervention.

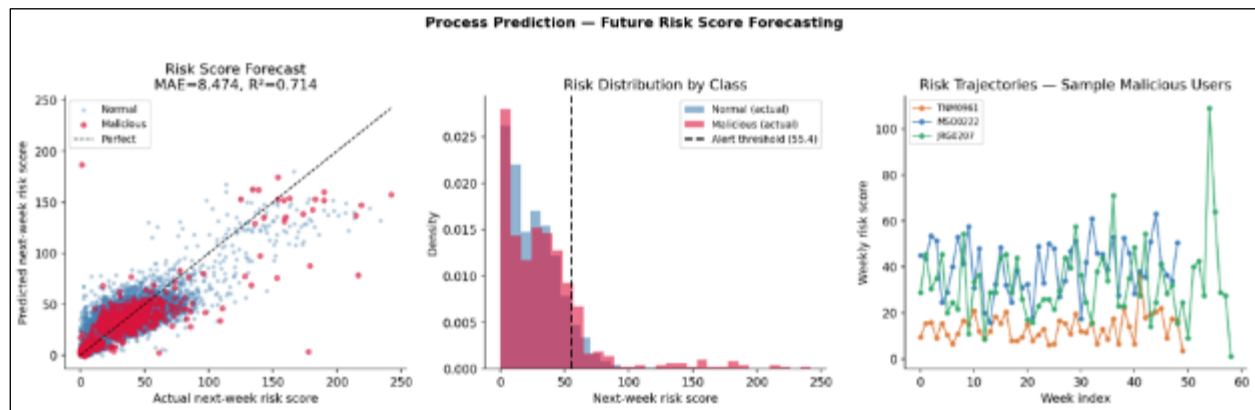


Figure 12: Future risk-score forecasting (Ridge regression, $\varepsilon = 1.0$). Left: scatter plot of predicted versus actual scores. Centre: risk-score distributions for malicious and normal users. Right: example trajectories of three confirmed insiders illustrating intermittent, high-variance risk patterns.

4.8 Limitations and generalizability

4.8.1 Residual Process-Mining Constraints

Although per-user Heuristics Miner discovery successfully eliminates the global flower-model degeneration, two practical limitations persist. First, the fixed 60-day training window proves insufficient for users who exhibit extremely sparse baseline activity; in these cases, the framework falls back to a neutral fitness value of 1.0 when fewer than ten training events are available. Second, the dependency threshold (0.5) and frequency filters were held constant across all users rather than being tuned individually. Adaptive threshold selection performed via cross-validation on each user's clean baseline period, or the adoption of precision-controlled discovery algorithms such as Split Miner [18], represents a natural and promising direction for further refinement.

4.8.2 Isolation Forest Calibration

Despite recalibrating the contamination parameter to the true user-week prevalence of 0.478 %, Isolation Forest continues to deliver limited binary-classification performance (PR-AUC = 0.019, F1 = 0.03). The model's underlying ranking quality remains respectable (AUC-ROC = 0.850), confirming that the anomaly scores themselves contain useful signal. Translating those scores into operationally reliable decisions, however, requires probability-based threshold selection that is explicitly optimized for a target precision-recall operating point, rather than continued reliance on the contamination hyper-parameter alone.

4.8.3 Privacy-Budget Composition Across the System Lifetime

The current implementation expends the entire privacy budget in a single feature-extraction step. In any production deployment the budget must be managed across repeated model retraining cycles, scheduled reporting, and ad-hoc analytical queries. Under basic sequential composition the total privacy cost grows linearly with the number of mechanisms; advanced composition theorems reduce this growth to $O(\sqrt{k} \epsilon)$ [19]. Consequently, future realisations of the framework should incorporate a privacy accountant that continuously tracks cumulative expenditure and enforces a long-term organisational budget ceiling.

4.8.4 Generalizability Beyond Synthetic Data

Although CERT r4.2 remains the community benchmark, it is synthetically generated and exhibits cleaner class separation than authentic enterprise logs [20]. Real operational environments contain a variety of non-malicious anomalies, system migrations, legitimate role changes, hardware failures, and authorized after-hours activity, that can closely resemble insider-threat signatures. In addition, the collection and downstream processing of psychometric data raise legal and ethical constraints that differ markedly across jurisdictions [21]. Any practical deployment must therefore re-validate both the labelling methodology and the feature pipeline against organization-specific ground truth, and must undergo appropriate legal and ethical review before psychometric attributes are retained.

4.8.5 Scalability Considerations

The present prototype processes 3.4 million events and 67 298 user-week windows within acceptable interactive runtimes. Scaling the approach to tens of thousands of users will require parallelization of the currently single-threaded per-user token-replay procedure, distributed feature computation (for example via Apache Spark), and, ideally, federated differential-privacy protocols that avoid centralizing raw event data while still respecting a global privacy budget.

5 CONCLUSION

This paper has presented a substantially strengthened privacy-preserving framework for insider-threat detection that integrates per-user process mining with formal differential privacy. Validated for the first time against the official CERT r4.2 ground-truth answer files, the framework introduces three decisive methodological advances: per-user Heuristics Miner discovery that eliminates the flower-model degeneration inherent in population-wide approaches, evaluation at realistic user-week granularity (67 298 windows at 0.48 % prevalence), and a complete privacy-utility characterization across seven values of ϵ .

Empirically, the inclusion of process-mining features yields consistent and substantial gains. At $\epsilon = 1.0$ the full feature set improves Random Forest AUC-ROC from 0.976 to 0.990, PR-AUC from 0.106 to 0.232 (a 2.2-fold increase), and F1 from 0.161 to 0.315. The single most discriminative signal is n_{variants} , the number of unique activity-sequence variants observed in a user-week, which alone achieves a Gini importance of 0.27. Differential privacy at the operationally relevant budget $\epsilon = 1.0$ incurs an AUC-ROC cost of only 0.001, while the resulting PR-AUC of 0.270 remains 56.1 times the random baseline. Aggregating weekly probabilities to the user level further produces an AUC-ROC of 0.923, enabling identification of approximately 80 % of confirmed insiders at a false-positive rate of 24 %. Complementary prediction tasks achieve an AUC of 0.776 for next-risky-activity forecasting and an R^2 of 0.714 for one-week-ahead risk-score regression, both under the same privacy constraint.

Collectively, these results demonstrate that formal (ϵ, δ) -differential privacy and operationally useful insider-threat detection are simultaneously attainable. When process mining is performed at the per-user rather than the population level, it supplies an interpretable, high-value, and privacy-compatible contribution that markedly elevates detection performance. The framework therefore offers both a concrete technical solution and a reproducible experimental template for future privacy-preserving behavioural analytics in enterprise environments.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

- [1] L. Philippe, P. Alex, H. David, and W. Suzanne, "2023 Data Breach Investigations Report," Verizon, 2023.
- [2] W. M. P. van der Aalst, *Process Mining: Data Science in Action*. Berlin, Germany: Springer, 2016, doi: 10.1007/978-3-662-49851-4.
- [3] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory of Cryptography Conf. (TCC)*, 2006, pp. 265–284, doi: 10.1007/11681878_14.
- [4] D. M. Cappelli, A. P. Moore, and R. F. Trzeciak, *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud)*. Upper Saddle River, NJ, USA: Addison-Wesley, 2012.
- [5] L. Liu, O. De Vel, Q. Han, J. Zhang, and Y. Xiang, "Detecting and preventing cyber insider threats: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1397–1417, 2nd Quart. 2018, doi: 10.1109/COMST.2018.2800740.
- [6] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. M. Robinson, "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," in *Proc. AAAI Workshop on Artificial Intelligence for Cyber Security*, 2017.
- [7] J. Xiao, L. Yang, F. Zhong, X. Wang, H. Chen, and D. Li, "Robust anomaly-based insider threat detection using graph neural network," *IEEE Trans. Netw. Service Manag.*, vol. 20, no. 3, pp. 3717–3733, Sep. 2023, doi: 10.1109/TNSM.2022.3222635.
- [8] U. Rauf, F. Mohsen, and Z. Wei, "A taxonomic classification of insider threats: Existing techniques, future directions & recommendations," *J. Cyber Security Mobility*, 2023.
- [9] W. M. P. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 9, pp. 1128–1142, Sep. 2004, doi: 10.1109/TKDE.2004.47.
- [10] X. Yang, D. Zhao, and Z. Li, "Insider threat detection based on process mining," in *Proc. IEEE Int. Conf. Dependable, Autonomic and Secure Comput. (DASC)*, 2021.
- [11] M. Macak, I. Vanat, M. Merjavy, T. Jevocin, and B. Buhnova, "Towards process mining utilization in insider threat detection from audit logs," in *Proc. IEEE Int. Conf. Social Networks Analysis, Management and Security (SNAMS)*, 2020, pp. 1–6, doi: 10.1109/SNAMS52053.2020.9336573.
- [12] T. Zhu, Y. Guo, A. Ju, J. Ma, and X. A. Wang, "An insider threat detection method based on business process mining," *Int. J. Bus. Data Commun. Netw.*, vol. 13, no. 2, pp. 83–98, 2017, doi: 10.4018/IJBDCN.2017070107.
- [13] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering block-structured process models from event logs — A constructive approach," in *Proc. Int. Conf. Applications and Theory of Petri Nets and Concurrency*, 2013, pp. 311–329, doi: 10.1007/978-3-642-38697-8_17.
- [14] K. A. Tabash and J. Happa, "Insider-threat detection using Gaussian mixture models and sensitivity profiles," *Comput. Secur.*, vol. 77, pp. 838–859, Aug. 2018, doi: 10.1016/j.cose.2018.03.006.
- [15] B. Balle and Y.-X. Wang, "Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 394–403.
- [16] X. Meng, H. Li, and J. Cui, "Different strategies for differentially private histogram publication," *J. Commun. Inf. Netw.*, vol. 2, no. 3, pp. 68–77, 2017, doi: 10.1007/s41650-017-0014-x.
- [17] M. Du, R. Jia, and D. Song, "Robust anomaly detection and backdoor attack detection via differential privacy," *arXiv:1911.07116*, 2019.
- [18] A. Augusto, R. Conforti, M. Dumas, M. La Rosa, and A. Polyvyanyy, "Split Miner: Automated discovery of accurate and simple business process models from event logs," *Knowl. Inf. Syst.*, vol. 59, no. 2, pp. 251–284, 2019, doi: 10.1007/s10115-018-1214-x.
- [19] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *Proc. IEEE 51st Annu. Symp. Found. Comput. Sci. (FOCS)*, 2010, pp. 51–60, doi: 10.1109/FOCS.2010.12.

- [20] A. Alshehri, N. A. Khan, A. Alowayr, and M. Y. Alghamdi, "Cyberattack detection framework using machine learning and user behavior analytics," *Comput. Syst. Sci. Eng.*, vol. 44, no. 2, pp. 1679–1689, 2022, doi: 10.32604/csse.2023.026526.
- [21] N. Martinez-Martin, T. R. Insel, P. Dagum, H. T. Greely, and M. K. Cho, "Data mining for health: Staking out the ethical territory of digital phenotyping," *npj Digit. Med.*, vol. 1, no. 1, 2018, doi: 10.1038/s41746-018-0075-8.