Privacy-Preserving Explainable Federated Learning for Intrusion Detection in Private Universities: A Design Science Study with Controlled Benchmark Evaluation

Masanja Cosmas[1], Kenneth Longo Mlelwa[2]*

[1]Institute of Accountancy Arusha, Tanzania

[2]The Mwalimu Nyerere Memorial Academy, Dar es Salaam, Tanzania

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11090145

Received: 04 October 2026; Accepted: 09 October 2026; Published: 22 October 2026

ABSTRACT

Private universities increasingly rely on networked information systems that store sensitive student and staff personally identifiable information (PII). This creates a practical security problem: institutions need to detect attacks effectively without introducing new privacy risks or relying on automated decisions that cannot be explained. This study developed and preliminarily evaluated a privacy-preserving explainable artificial intelligence (PPXAI) framework that integrates federated learning, differential privacy, neural intrusion detection, SHAP-based explanation, and audit-oriented governance. The work followed a Design Science Research approach and began with a mixed-methods diagnosis in two private universities in Iringa Region, Tanzania. Of the 381 targeted questionnaire respondents, 363 returned usable responses. Existing cybersecurity practices were rated moderate (M = 3.34), PII protection challenges were high (M = 3.88), and requirements for privacy and explainability were rated very high (M = 4.23).

These findings informed the framework architecture. Technical validation used a stratified sample of 504,159 CIC-IDS2017 network flows distributed across two non-IID simulated institutional clients. A multilayer perceptron was trained through FedAvg with DP-SGD at a privacy budget of epsilon = 2.99 per client and compared with local-only models, Isolation Forest, and local differential-privacy input perturbation. The tuned PPXAI model achieved 96.96% accuracy, 89.95% precision, 92.32% recall, an F1-score of 91.12%, and an AUC-ROC of 0.993. Recall improved by 15.29 percentage points over local-only training while raw institutional data remained local. SHAP explanations were also tested for faithfulness: removing the five highest-ranked features reduced attack probability by 0.59 on average, compared with 0.04 for randomly selected features.

The model nevertheless missed SSH brute-force and web attacks and generated more false positives than the local-only models. The evidence therefore supports PPXAI as a complementary security layer rather than a stand-alone control, and shows the value of treating privacy, detection, explainability, and governance as linked design requirements in higher-education cybersecurity.

Keywords: federated learning; differential privacy; explainable AI; intrusion detection; higher education cybersecurity

INTRODUCTION

Universities now operate as data-intensive organisations. Admissions, learning management, finance, human resources, libraries, identity services, and research platforms all depend on digital infrastructure that carries personal as well as operational information. This dependence expands the attack surface. Recent research has

documented persistent exposure in the university sector to phishing, credential compromise, vulnerable applications, legacy systems, and fragmented security arrangements [9].

Tanzanian higher-learning institutions face similar concerns, including weaknesses in cybersecurity posture and in the protection of student information systems [8], [13]. Machine-learning intrusion detection can improve the speed and consistency of identifying anomalous traffic, but conventional model training often assumes that data can be centralised. For universities, that assumption creates an additional risk because sensitive records may need to be transferred, pooled, or exposed simply to build a better detector. Federated learning offers a different arrangement: participating institutions keep raw data locally and share model updates instead [11].

That approach reduces unnecessary data movement, but it does not by itself guarantee privacy because model updates may still leak information under some threat conditions [7]. Differential privacy addresses part of this problem by limiting what can be inferred about individual records through gradient clipping and calibrated noise during training [1]. A separate concern is whether security decisions can be understood and defended. An alert that cannot be interrogated is difficult to justify during incident response, audit, or administrative review.

This is why SHAP and related explanation methods have become increasingly relevant in intrusion-detection research: they make the contribution of individual features more visible to analysts [2], [10], [12]. The issue is especially important in educational institutions, where technical teams must respond quickly while still meeting accountability and data-protection obligations. Recent work on federated learning in education also shows the practical value of keeping data local while allowing collaborative model training [4]. What remains less well integrated is the relationship among privacy-preserving learning, explainable intrusion detection, and institutional governance.

These areas are often examined separately, even though a deployable university security system has to satisfy all three at once. Private universities may operate heterogeneous systems, face staffing and resource constraints, and still be expected to produce auditable security decisions. This study therefore treats detection effectiveness, privacy protection, explainability, and auditability as simultaneous design requirements. The PPXAI framework is grounded in requirements identified in two Tanzanian private universities and its technical core is evaluated through a controlled and reproducible benchmark experiment.

The paper makes three contributions. First, it translates institution-level cybersecurity and privacy requirements into architectural choices, rather than starting with a predictive model and adding privacy or explanation mechanisms afterwards. Second, it evaluates a federated neural detector under a formal differential-privacy budget and compares it with other privacy-preserving alternatives on the same held-out test data. Third, it evaluates explanation faithfulness and operational usefulness alongside conventional detection measures.

The aim is not to claim that the framework is ready for production deployment. Instead, the study asks whether an integrated architecture provides a credible technical basis for later institutional pilots.

RELATED WORK

Federated learning was developed to allow a shared model to be trained across decentralised datasets without requiring participating clients to transfer their raw data to a central repository [11]. This is attractive in cybersecurity because institutions can learn from a broader set of attack patterns while retaining local control of network and user records. Its privacy benefits, however, depend on the surrounding threat model and safeguards. Reviews of the field continue to identify risks such as inference, poisoning, reconstruction, and attacks on model updates, which means that federated learning should not be treated as a complete privacy solution on its own [7].

Differential privacy provides one of the most established mechanisms for limiting record-level disclosure during model training. In DP-SGD, individual gradient contributions are clipped before calibrated noise is

added, allowing the privacy guarantee to be expressed through epsilon and delta [1]. The practical trade-off is unavoidable: stronger privacy may reduce predictive utility, whereas a weak privacy setting may offer little meaningful protection. For this reason, a credible evaluation should report model performance together with the privacy parameters under which that performance was obtained.

Explainability addresses a different weakness. Intrusion-detection models can classify traffic accurately while offering little insight into why a particular flow was labelled malicious. Lundberg and Lee [10] introduced SHAP as a unified feature-attribution approach, and subsequent cybersecurity studies have shown its value for interpreting intrusion and malware predictions [2], [12]. Still, an explanation that looks plausible is not necessarily faithful to the underlying model.

Explanation quality should therefore be tested by checking whether the features identified as important actually influence the model's output. Higher-education cybersecurity studies have documented both sector-specific exposure and institutional constraints [9]. In Tanzania, existing research also points to the need for stronger cybersecurity arrangements and better protection of information-management systems [8], [13]. Chen et al. [4] further demonstrate the relevance of federated learning in education-oriented privacy settings.

What is still less developed is an integrated artefact that begins with institutional requirements and combines federated learning, formal privacy protection, explainable detection, and governance-oriented logging. That gap provides the basis for the Design Science approach adopted in this study.

METHODOLOGY

Research design and institutional requirement elicitation

The study used Design Science Research (DSR) because the central objective was to build and evaluate an artefact that responds to a defined organisational problem [5], [6], [14]. A pragmatic mixed-methods strategy supported the problem-diagnosis stage. Fieldwork was conducted at Ruaha Catholic University and the University of Iringa in Tanzania. The target population comprised 7,828 students and staff.

Using Yamane's formula at a 5% precision level, the calculated quantitative sample was 381 [16]. Of these, 363 respondents returned usable questionnaires, equivalent to a 95.3% usable response rate. The questionnaire covered four areas on five-point scales: current cybersecurity practices, PII protection challenges, privacy and explainability requirements, and policy/data-protection alignment. Semi-structured interviews were conducted with information-rich participants drawn from ICT, systems, data-governance, and management roles.

Their accounts helped explain the survey patterns and identify practical implementation constraints. Interview data were coded thematically and compared with the quantitative findings [3]. The purpose of this stage was to elicit design requirements rather than to test causal hypotheses.

PPXAI architecture

The framework that emerged from the field evidence was organised around six connected functions: local data retention and preprocessing, local neural intrusion detection, federated model aggregation, differential-privacy protection during client training, SHAP-based explanation of alerts, and governance-oriented logging with human review. A deliberate boundary was maintained between model assistance and institutional decision authority. Raw traffic and PII remain at the local site, while model updates contribute to collaborative learning. Alerts are accompanied by feature-level explanations, and analyst actions are retained so that decisions can later be reviewed or audited.

Table 1. Institutional design requirements and PPXAI responses

Design requirement	Evidence from field stage	PPXAI response
Uneven cybersecurity maturity	Current practices: M = 3.34	Layered architecture that can complement

		existing controls
High PII protection difficulty	PII challenges: M = 3.88	Local data processing and privacy-preserving training
Strong privacy and explainability demand	Privacy/explainability: M = 4.23	Federated learning, DP-SGD, SHAP explanations
Strong governance and policy demand	Policy/alignment: M = 4.18	Audit logging, explanation retention, human review

Controlled technical validation

Technical evaluation used the CIC-IDS2017 Machine Learning CVE benchmark [15]. After removing missing or infinite values and exact duplicates, 2,520,798 flows remained. A 20% stratified sample of 504,159 flows was then used for the experiment. The data were separated into training, validation, and held-out test partitions, with the test set excluded from model tuning.

Two non-IID clients were created to represent the participating institutional sites. Seventy numeric flow features were retained after constant columns were removed, and standardisation was performed using training-set statistics only.

Table 2. Controlled benchmark configuration

Parameter	Configuration
Dataset	CIC-IDS2017 Machine Learning CVE
Flows used	504,159 stratified flows
Training / validation / test	299,974 / 52,937 / 151,248
Test composition	25,545 attack; 125,703 benign
Features	70 numeric flow features
Detection model	MLP 64-32-1, ReLU, binary classification
Federated setting	FedAvg, 2 non-IID clients, 30 rounds × 1 local epoch
Differential privacy	DP-SGD; clipping norm 1.0; noise multiplier 0.839
Privacy budget	epsilon = 2.99 per client; delta = 3.3 × 10^-6
Selected PPXAI tuning	Class weight 1; learning rate 0.4; threshold 0.16
Explainability	SHAP GradientExplainer on 300 true-positive alerts
Comparison models	Local-only MLP; Isolation Forest; local DP input perturbation

Model selection was based on validation F1-score. The PPXAI configuration was evaluated against three alternatives: local-only multilayer perceptrons, an Isolation Forest trained on each client's benign traffic, and local differential privacy implemented through Laplace input perturbation at epsilon = 3. The held-out test set was used only for final evaluation. Reported measures included accuracy, precision, recall, F1-score, AUC-ROC, false-positive rate (FPR), and false-negative rate (FNR).

Explanation faithfulness was tested by removing SHAP-ranked features and comparing the resulting reduction in predicted attack probability with the reduction produced by removing randomly selected features.

Privacy threat model, reproducibility scope, and efficiency assumptions

The controlled experiment assumes that participating institutions retain raw traffic locally and that the federated coordinator receives model updates rather than source records. The DP-SGD configuration limits the influence of individual training records under the reported epsilon and delta budget, but it should not be interpreted as protection against every federated-learning threat. The study did not implement secure aggregation, Byzantine-robust aggregation, or explicit defences against malicious clients, poisoning,

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

membership-inference, or gradient-reconstruction attacks. Accordingly, the privacy claim is limited to the formal DP training mechanism evaluated here and does not extend to untested adversarial conditions.

Reproducibility is supported by reporting the benchmark dataset, preprocessing logic, partition sizes, model architecture, federated rounds, differential-privacy parameters, model-selection criterion, and final held-out evaluation protocol. However, the study did not independently benchmark end-to-end runtime, memory use, or communication volume across participating sites. It therefore makes no claim that the present configuration is computationally or communication-optimal for production deployment. These resource costs should be measured directly during multi-site pilots before operational adoption.

Ethical and data-protection safeguards

Names were not required on questionnaires, interview records were coded, and access to the research data was restricted. The technical evaluation used a public benchmark rather than live university traffic, and identifiable participant information was not used to train the intrusion detector. Participation in the field component was voluntary, and the work was conducted under the institutional research procedures applicable to the dissertation.

The technical validation did not use live traffic from Ruaha Catholic University or the University of Iringa, and no participant PII was used to train the intrusion-detection model. The field component was used for requirement elicitation, while the model evaluation relied on the public CIC-IDS2017 benchmark.

RESULTS

Institutional requirements

The requirement-elicitation stage revealed a clear gap between the maturity of current controls and the level of protection expected by users and institutional personnel. Existing cybersecurity practices were rated moderate ($M = 3.34$, $SD = 0.71$), while PII protection challenges were rated high ($M = 3.88$, $SD = 0.66$). Privacy and explainability requirements received the strongest rating ($M = 4.23$, $SD = 0.55$), followed closely by policy and data-protection alignment ($M = 4.18$, $SD = 0.58$). The qualitative findings helped explain these scores.

Basic identity and access controls were more established than continuous monitoring and integrated response, while credential compromise, heterogeneous systems, resource limitations, and the need for auditable security decisions remained recurring concerns.

Overall detection performance

Among the privacy-preserving approaches, PPXAI produced the strongest overall balance of detection measures. Recall reached 92.32%, compared with 77.03% for local-only training and 58.15% for Isolation Forest. The 15.29 percentage-point advantage over local-only training is operationally meaningful because the federated model was able to learn from attack patterns distributed across both clients while the clients retained their raw data locally. That gain came with a cost, however: the local-only models achieved higher precision and a substantially lower false-positive rate.

Table 3. Overall detection performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC-ROC	FPR (%)	FNR (%)
PPXAI + DP (epsilon=2.99)	96.96	89.95	92.32	91.12	0.993	2.1	7.68
Local-only models	95.59	96.2	77.03	85.43	0.97	0.64	22.97
Isolation Forest	89.21	72.54	58.15	64.55	0.838	4.48	41.85
Local DP input perturbation	16.89	16.89	100	28.9	0.516	100	0

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

Attack-specific performance

Attack-level results show why aggregate accuracy should not be interpreted on its own. PPXAI detected 97.8% of port scans and 96.8% of DDoS flows, and it also outperformed the comparison models on DoS GoldenEye, slow denial-of-service, FTP brute-force, and bot traffic. Performance was weaker on DoS Hulk than under local-only training, and the model failed to detect SSH brute-force and web attacks in the held-out test set. These misses matter operationally because credential misuse had already emerged as an important PII risk during the institutional field stage.

Table 4. Attack-specific detection rates

Attack type	PPXAI (%)	Local-only (%)	Isolation Forest (%)
PortScan	97.8	45.6	0.2
DDoS	96.8	79.9	62.3
DoS Hulk	92.1	96	88.8
DoS GoldenEye	87.7	70.9	67.7
DoS slowloris / Slowhttptest	73.6	57.9	67
FTP-Patator brute force	66	49.4	0
SSH-Patator brute force	0	47.7	0
Bot	25.6	14.5	0.8
Web attacks	0	3.5	2.3

Explainability and faithfulness

Expert assessment of the explanation layer was favourable across all measured dimensions. Support for auditability received the highest score (M = 4.42, SD = 0.51), followed by the relevance of highlighted features (M = 4.36, SD = 0.54), overall explainability (M = 4.32, SD = 0.49), usefulness for investigation (M = 4.31, SD = 0.58), and clarity (M = 4.18, SD = 0.61). The deletion test added model-based evidence that the explanations reflected genuine model behaviour rather than visual plausibility alone. Across 300 correctly detected attacks, removing the five highest-ranked SHAP features reduced predicted attack probability by 0.59 on average, compared with 0.04 after removing five random features.

Removing the top ten SHAP-ranked features produced a mean reduction of 0.78, versus 0.08 under random deletion. Backward packet-length standard deviation, packet-length variance, and packet-length standard deviation were among the most influential features.

DISCUSSION

The results suggest that a privacy-preserving detector can improve cross-site attack coverage without requiring institutions to pool raw data. The clearest evidence is the difference in recall between PPXAI and local-only learning. Each local model could learn only from the attack patterns present in its own partition, whereas federated learning allowed the shared model to benefit from patterns distributed across both clients. This result is consistent with the original rationale for federated learning [11] and with more recent privacy-oriented applications in education [4].

The privacy mechanism also exposed a realistic trade-off rather than an idealised one. At epsilon = 2.99, PPXAI retained high AUC-ROC and recall, but its performance did not match a no-privacy pooled model exactly. That loss is consistent with the expected cost of formal privacy protection and underlines why privacy parameters should be reported alongside conventional accuracy measures [1], [7]. The local-DP comparison is particularly revealing.

The present study evaluates one principal DP-SGD operating point (epsilon = 2.99 per client) and therefore does not establish a complete privacy-utility curve. A stronger evaluation should compare several privacy budgets under the same data partitions and model-selection procedure so that institutions can quantify how progressively stronger privacy affects recall, false positives, and minority attack classes. The reported result should therefore be read as evidence for one controlled configuration rather than as an optimal privacy setting.

Adding noise directly to input records at a comparable privacy budget destroyed much of the class information and caused the model to classify all traffic as malicious. In this experiment, protecting the training process through DP-SGD was therefore far more useful than perturbing every input record. Explainability also changed the operational value of the detector. A false-positive rate of 2.10% corresponds to roughly 210 alerts for every 10,000 benign flows, which is considerably higher than the rate produced by the local-only models.

The false-positive burden is operationally important because a university security team may process very large volumes of benign traffic. At scale, even a modest percentage can translate into a substantial analyst workload, so threshold calibration, alert prioritisation, and human-review workflows are necessary parts of deployment. Future pilots should measure not only model accuracy but also alert volume per analyst, time-to-triage, escalation rates, and the proportion of explanations judged actionable by security staff.

That burden is only manageable if analysts can triage alerts efficiently. SHAP therefore had a functional role rather than a cosmetic one. The deletion tests showed that the features presented to analysts materially affected the model's predictions, addressing a common weakness in XAI evaluations that rely only on visually convincing explanations [2], [10], [12]. The attack-specific findings also set clear limits on what the framework can claim.

A detector that misses SSH brute-force and web attacks cannot reasonably be presented as a complete university security solution. This limitation is especially important because the field study identified credential misuse as a major risk to PII. PPXAI is therefore better understood as one layer within a broader control environment that also includes identity and access management, multi-factor authentication, account-lockout policies, authentication-log monitoring, patch management, and incident response. Improved intrusion visibility does not remove the need for these controls.

The Design Science perspective helps explain why the contribution extends beyond tuning a classifier. The artefact was derived from a documented institutional problem, translated into design objectives, implemented as an integrated architecture, and evaluated against both technical and organisational criteria [5], [6], [14]. Privacy, explainability, and auditability were treated as first-order design requirements rather than later additions. This is particularly important in resource-constrained higher-education settings, where even a technically accurate model may be unsuitable if it depends on centralised PII, produces uninterpretable alerts, or cannot support institutional accountability.

The study nevertheless has important limitations. The technical evaluation relied on CIC-IDS2017 rather than live traffic from the participating universities, and the two clients were simulated from a single benchmark dataset. The results therefore demonstrate controlled feasibility, not deployment performance. Real institutional networks may contain different traffic distributions, encrypted services, concept drift, policy constraints, and adversarial behaviour.

Privacy was evaluated through the DP accountant rather than through attack-based leakage tests such as membership inference or gradient reconstruction. In addition, the usefulness of the explanation layer was assessed in a preliminary setting rather than through a longitudinal study of analyst behaviour. These limitations define a clear next step: a pilot using locally collected and ethically governed network traffic, explicit privacy attacks, drift monitoring, and measures of analyst workload.

The attack-level weaknesses also suggest that class imbalance and attack rarity require targeted treatment. Future experiments should test class balancing, targeted augmentation, alternative loss functions, and architectures designed for sparse attack categories, while preserving a strictly held-out evaluation set. Additional benchmark datasets and more heterogeneous federated clients are also needed to determine whether the observed performance generalises beyond the present two-client CIC-IDS2017 configuration.

A further limitation concerns governance and explainability validation. Although expert assessment and deletion-based faithfulness testing provide useful evidence, they do not substitute for longitudinal use by operational analysts. Controlled institutional pilots should therefore examine analyst trust, explanation usefulness, policy compliance, escalation decisions, and the practical consequences of false-positive alerts over time.

CONCLUSION

This study developed and preliminarily evaluated a PPXAI framework for higher-education cybersecurity that integrates federated learning, DP-SGD, neural intrusion detection, SHAP explanations, and governance-oriented auditability. The design was grounded in field evidence from two Tanzanian private universities and then tested under a controlled two-client benchmark configuration. At epsilon = 2.99, PPXAI achieved 96.96% accuracy, 92.32% recall, an F1-score of 91.12%, and an AUC-ROC of 0.993, outperforming the other privacy-preserving approaches on overall recall and several major attack categories. The explanation layer was rated positively by experts and was also supported by deletion-based faithfulness testing.

The results are not uniformly positive: PPXAI produced more false positives than local-only training and failed to detect SSH brute-force and web attacks. A cautious conclusion is therefore warranted. Privacy-preserving, explainable federated detection is technically promising as one component of a broader university security architecture, but live institutional pilots, identity-focused controls, and continuing model governance are required before operational adoption.

Before production use, the framework should be evaluated with live, ethically governed institutional traffic, additional datasets, more heterogeneous clients, multiple differential-privacy budgets, explicit privacy-attack testing, and direct measurement of computational and communication overhead. Such evaluation would allow the present controlled benchmark evidence to be translated into a more robust operational assessment.

ACKNOWLEDGEMENTS

The authors acknowledge the cooperation of the participants and institutional personnel who contributed to the underlying dissertation study, as well as the Institute of Accountancy Arusha for providing the academic environment in which the research was conducted.

DATA AVAILABILITY

The technical evaluation used the publicly available CIC-IDS2017 MachineLearningCVE benchmark. Institutional questionnaire and interview data are not included with the manuscript because they contain context-specific research information collected under confidentiality arrangements. The aggregated results needed to understand the study are reported in the paper.

CONFLICT OF INTEREST

No conflict of interest was reported in the dissertation materials on which this manuscript is based. The authors should ensure that this statement remains accurate at the point of final submission.

FUNDING

No external funding was reported in the dissertation materials on which this manuscript is based.

REFERENCES

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308- 318. https://doi.org/10.1145/2976749.2978318
1. Arreche, O., Guntur, T. R., Roberts, J. W., & Abdallah, M. (2024). E-XAI: Evaluating black-box explainable AI frameworks for network intrusion detection. IEEE Access, 12, 23954- 23988.

https://doi.org/10.1109/ACCESS.2024.3365140
2. Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. Qualitative Psychology, 9(1), 3- 26. https://doi.org/10.1037/qup0000196
3. Chen, S., Qi, X. Z., Han, X. H., Fan, Z. C., & Wang, L. L. (2026). Federated learning for teacher data privacy protection: A study in the context of the PIPL. Frontiers in Big Data, 9, 1681382. https://doi.org/10.3389/fdata.2026.1681382
4. Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS Quarterly, 37(2), 337- 355. https://doi.org/10.25300/MISQ/2013/37.2.01
5. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75- 105. https://doi.org/10.2307/25148625
6. Hu, K., Gong, S., Zhang, Q., Seng, C., Xia, M., & Jiang, S. (2024). An overview of implementing security and privacy in federated learning. Artificial Intelligence Review, 57, 204. https://doi.org/10.1007/s10462-024-10846-8
7. Kifaru, F., Kavuta, K., & Semlambo, A. (2023). Assessment of the impacts of cyber security on student information management systems: A case of Ruaha Catholic University. The Journal of Informatics, 3(1), 51- 67. https://doi.org/10.59645/tji.v3i1.127
8. Lallie, H. S., Thompson, A., Titis, E., & Stephens, P. (2025). Analysing cyber attacks and cyber security vulnerabilities in the university sector. Computers, 14(2), 49. https://doi.org/10.3390/computers14020049
9. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-
10. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Aguera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273-
11. Mohale, V. Z., & Obagbuwa, I. C. (2025). Evaluating machine learning-based intrusion detection systems with explainable AI: Enhancing transparency and interpretability. Frontiers in Computer Science, 7, 1520741. https://doi.org/10.3389/fcomp.2025.1520741
12. Mtakati, B., & Sengati, F. (2021). Cybersecurity posture of higher learning institutions in Tanzania. The Journal of Informatics, 1(1). https://doi.org/10.59645/tji.v1i1.1
13. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45- 77. https://doi.org/10.2753/MIS0742-1222240302
14. Sharafaldin, I., Habibi Lashkari, A., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy.
15. Yamane, T. (1967). Statistics: An introductory analysis (2nd ed.). Harper & Row.