American Journal of Interdisciplinary Studies

Volume: 7; **Issue:** 1
Pages: 355-386
eISSN: 3067-5146

AI-Enabled Intrusion Detection in Enterprise Networks: A Systematic Review of Methods, Datasets, and Evaluation Metrics (2018–2026)

Md Zahedul Islam[1];

[1]. *Master of Science in Cybersecurity, Mercy University, Dobbs Ferry, NY, USA*
Email: zahed.arman44@gmail.com

Doi: 10.63125/k4t9f683
Received: *21 November 2025;* ***Revised:*** *27 December 2025;* ***Accepted:*** *21 January 2026;* ***Published:*** *08 February 2026*

Abstract

This study addresses the problem that AI-enabled intrusion detection systems (IDS) often report strong benchmark performance yet struggle to deliver consistently actionable, trusted alerts in real enterprise and hybrid cloud environments where traffic is heterogeneous, class imbalance is extreme, and false-positive workload can overwhelm security operations centers. The purpose was to quantify which enterprise-relevant factors most strongly predict perceived AI-IDS effectiveness and operational suitability, while linking evaluation evidence to workload impact. Using a quantitative, cross-sectional, case-based design, data were collected from enterprise and cloud-facing security stakeholders (N = 162 valid responses) involved in IDS monitoring and triage. Key variables were measured on 5-point Likert scales, including Dataset Representativeness (DREP), Evaluation Rigor (ERIG), Model Robustness (MROB), Deployment Readiness (DREADY), Explainability and Trust (TRUST), and the dependent outcome Perceived AI-IDS Effectiveness (EFFECT); construct reliability was acceptable to strong (Cronbach's a = 0.78–0.88). The analysis plan applied descriptive statistics, Pearson correlations, and multiple regression with diagnostics (VIF = 1.28–2.05). Headline findings showed moderately high perceived effectiveness (EFFECT M = 3.73, SD = 0.66) and trust (TRUST M = 3.69, SD = 0.68), with EFFECT positively correlated with TRUST (r = .52, p < .001), ERIG (r = .46, p < .001), DREP (r = .42, p < .001), MROB (r = .39, p < .001), and DREADY (r = .34, p < .001). The regression model explained substantial variance in effectiveness (R^2 = .51; adjusted R^2 = .49; F(5,156) = 32.45, p < .001), with TRUST the strongest predictor (β = .33, p < .001), followed by ERIG (β = .22, p = .002), DREP (β = .18, p = .008), MROB (β = .15, p = .019), and DREADY (β = .11, p = .041). Operational implications were quantified using a False-Positive Burden Index: at 420 alerts/day, FPR = 0.07, and 6.5 minutes triage time, false positives consumed 191.1 minutes/day (3.19 hours/day), while reducing FPR to 0.04 lowered burden to 109.2 minutes/day (1.82 hours/day), a 42.8% reduction. Overall, the results imply that enterprises gain the most adoption-ready value when AI-IDS is explainable, evaluated rigorously with enterprise-representative data, and tuned to reduce workload through threshold governance and imbalance-aware metrics.

Keywords

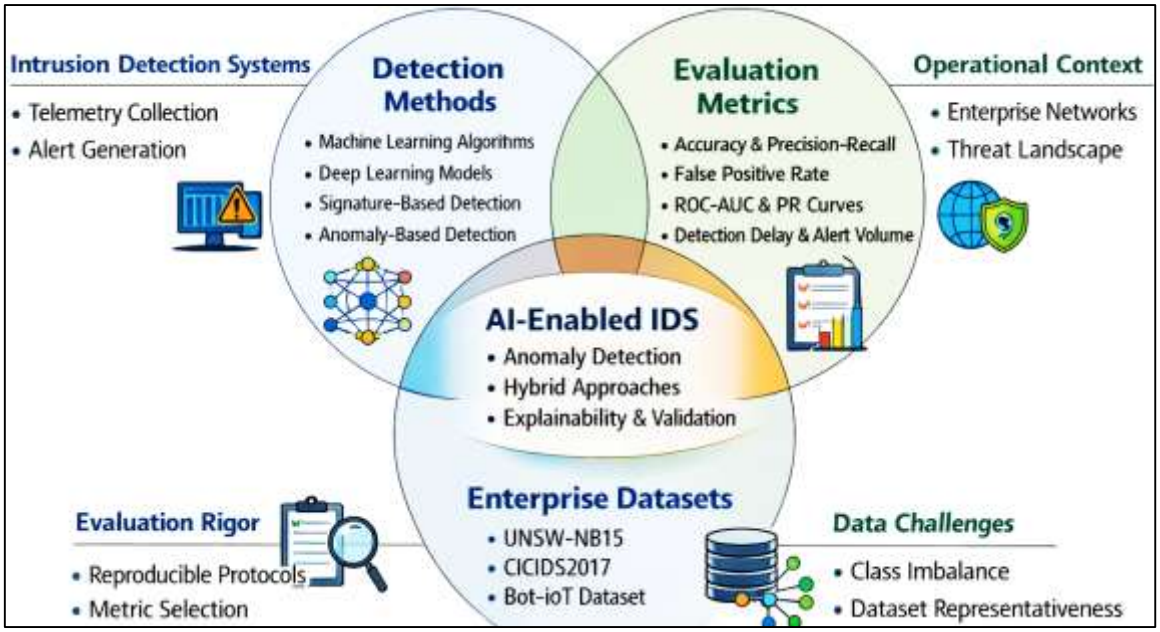
AI-Enabled Intrusion Detection; Enterprise Networks; Explainability and Trust; Evaluation Rigor; False-Positive Burden Index;

INTRODUCTION

Artificial intelligence (AI) refers to computational methods that learn patterns from data to support classification, prediction, and decision-making under uncertainty, often through machine learning (ML) and deep learning (DL) architectures that discover layered representations from large datasets (Chandola et al., 2009). In cybersecurity, intrusion detection is commonly defined as the process of monitoring hosts and networks to identify malicious events or policy violations, using detectors that recognize known attack signatures or infer abnormality relative to expected behavior (Buczak & Guven, 2016). An intrusion detection system (IDS) operationalizes this process by collecting telemetry (e.g., flows, packets, logs), extracting features, and applying detection logic to produce alerts that inform incident response and risk governance (Diro & Chilamkurti, 2018). AI-enabled intrusion detection extends conventional IDS by using learning algorithms to infer decision boundaries from data rather than relying only on static rules, enabling models to generalize across variability in network behavior and attack execution (Wilkinson et al., 2016). The international significance of AI-enabled IDS is linked to the centrality of enterprise networks in economic activity, public services, and cross-border digital trade, where persistent threats, automated reconnaissance, and credential abuse create operational and compliance exposure across sectors and jurisdictions (Sommer & Paxson, 2010). Modern enterprises frequently operate distributed infrastructures spanning on-premises environments, cloud services, and remote endpoints, producing heterogeneous traffic patterns and telemetry sources that challenge manual monitoring and rule-based detection approaches (Sharafaldin et al., 2018). At the same time, AI-based classifiers introduce their own governance concerns because detection decisions can influence containment actions, business continuity, and reporting obligations, which increases the need for transparent evaluation and auditable performance evidence. Accordingly, AI-enabled IDS sits at the intersection of analytics, operational security, and assurance, where effectiveness depends on evidence-based methodology, credible datasets, and measurement frameworks that reflect enterprise constraints and risk tolerance (Gamage & Samarabandu, 2020).

IDS research frequently distinguishes signature-based detection from anomaly-based detection. Signature approaches match observed events to known patterns, while anomaly approaches model normal activity and flag deviations that may represent attack activity or misconfiguration (Moustafa & Slay, 2015). The anomaly paradigm is tightly coupled with the broader field of anomaly detection in data mining, where the central problem is to identify rare, unusual, or suspicious observations relative to a learned reference distribution. In enterprise networks, this distinction matters because "normal" behavior is shaped by business cycles, patching windows, software rollouts, and user mobility, which can create variability that resembles attack behavior and can elevate false positive rates when models are not tuned to the operational context (Fawcett, 2006; Rauf, 2018). Empirical evaluation practices therefore become essential, because an IDS that performs well under one traffic regime or dataset composition can underperform under another, even when algorithmic choices remain unchanged (Ferrag et al., 2020). AI-enabled IDS often combines both paradigms by learning discriminative models for known classes while also supporting detection of novel or low-prevalence patterns through unsupervised or semi-supervised learning, including autoencoders that compress high-dimensional network features and surface reconstruction error as an anomaly signal (García-Teodoro et al., 2009; Ashraful et al., 2020). Deep models can also represent temporal dependencies, which is relevant to multistage intrusions and scan-to-exploit sequences that unfold over time rather than as isolated packets or flows (Saito & Rehmsmeier, 2015). These capabilities are attractive in enterprises where threats often manifest as sequences of weak signals dispersed across telemetry sources. Still, the central scientific requirement remains consistent: models must be evaluated using measurement designs that separate overfitting from true generalization and that quantify trade-offs between detection sensitivity and operational noise at realistic base rates. For that reason, AI-enabled IDS research increasingly emphasizes metric selection, dataset representativeness, and reproducible reporting as core elements of trustworthy evidence (Haque & Arifur, 2021; He & Garcia, 2009).

Figure 1: AI-Enabled Intrusion Detection in Enterprise Networks



This study aims to quantitatively examine how AI-enabled intrusion detection methods perform within enterprise network conditions by systematically connecting model design choices, dataset characteristics, and evaluation metrics to measurable detection outcomes. The primary objective is to develop an evidence-driven assessment framework that supports comparative evaluation of intrusion detection approaches across representative enterprise-relevant datasets and clearly defined performance indicators. The study will operationalize key constructs such as detection effectiveness (e.g., attack identification capability), alert quality (e.g., precision-oriented usefulness), and operational burden (e.g., false-positive workload implications) and will analyze how these constructs relate to one another under realistic class imbalance and multi-attack settings. A cross-sectional, case-study–based design will be applied to capture the current state of AI-enabled IDS performance across selected enterprise scenarios, using structured measurement items organized through a five-point Likert instrument to quantify practitioner-aligned perceptions of usability, trust, and decision support value alongside observed technical performance outputs. Descriptive statistics will summarize central tendencies and dispersion for each construct and metric, providing a baseline profile of model behavior and operational feasibility. Correlation analysis will test the strength and direction of relationships among constructs such as model transparency, dataset suitability, and analyst acceptance, as well as their associations with measurable detection indicators. Regression modeling will be used to estimate the predictive influence of independent variables such as method family, feature strategy, and dataset properties on dependent outcomes such as detection quality, false-positive load, and overall utility within the enterprise context. The study further seeks to produce a structured validation matrix that links case-study requirements to model capabilities and metric thresholds, enabling a consistent interpretation of results across multiple enterprise conditions. Through this design, the research will provide a rigorously measured, statistically supported account of how AI-enabled intrusion detection delivers value in enterprise networks and which factors most strongly explain variations in performance and operational suitability.

LITERATURE REVIEW

The literature on AI-enabled intrusion detection in enterprise networks has developed into a multidisciplinary body of work that combines cybersecurity foundations, machine learning research, and enterprise security operations practice. At its core, this literature conceptualizes intrusion detection as the continuous monitoring and analysis of network and host activity to identify unauthorized access, malicious behavior, and policy violations within complex organizational infrastructures. The enterprise setting is especially significant because enterprise networks operate under high traffic volume, diverse user roles, heterogeneous applications, cloud and hybrid deployments, remote access patterns, and

strict compliance and governance requirements. These characteristics create monitoring environments where malicious behavior is often concealed within legitimate operational activity, making it difficult to distinguish attacks from normal variability. As a result, the literature has examined the evolution from traditional signature-based intrusion detection toward anomaly-based and learning-based approaches that aim to detect both known and previously unseen threats. A major stream of research has focused on algorithmic methods, including classical machine learning classifiers, deep learning architectures, hybrid models, and unsupervised anomaly detection techniques, each offering different advantages depending on the nature of available telemetry and the type of intrusion behavior being targeted. In parallel, a substantial portion of the literature has addressed the dataset and evaluation problem, emphasizing that model performance is inseparable from dataset representativeness, labeling quality, class imbalance, and the realism of benign background traffic. This dataset dimension has been critical because many intrusion detection studies have relied on benchmark corpora that may not fully reflect modern enterprise conditions such as encryption prevalence, multi-cloud connectivity, and evolving attacker strategies. Another consistent theme in the literature has been evaluation methodology, where researchers have emphasized that metric selection and validation protocols must reflect operational constraints, including false-positive burden, detection latency, scalability, and reproducibility. Overall, the literature has framed AI-enabled intrusion detection as an evidence chain connecting model design, dataset quality, evaluation rigor, and enterprise feasibility, rather than as a purely algorithmic competition.

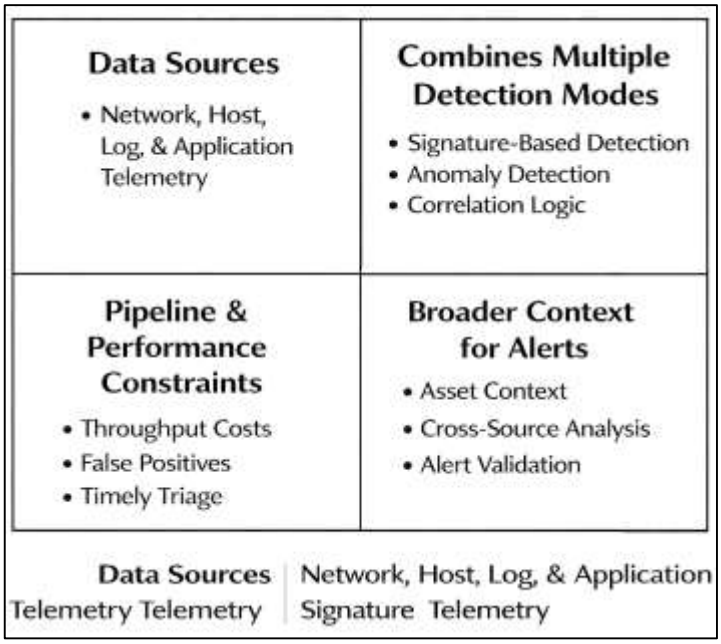
Enterprise Intrusion Detection Systems

Enterprise networks rely on layered security controls that include network intrusion detection systems (NIDS), host intrusion detection systems (HIDS), and centralized alerting pipelines that feed security operations centers (SOCs). In this environment, intrusion detection is not only a classification activity; it is an operational workflow that transforms high-volume telemetry into prioritized investigations, triage decisions, and response actions. Enterprises generate heterogeneous data streams such as packet or flow records, DNS and web logs, authentication events, endpoint process traces, and application telemetry, and IDS components must align these signals with organizational assets, business processes, and risk tolerance (Fokhrul et al., 2021; Zaman et al., 2021). A key distinction in enterprise settings is that "normal" behavior is not a single stable pattern: normality varies across departments, time zones, seasons, mergers, cloud migrations, and policy changes, meaning that detection logic must remain useful under shifting baselines (Fahimul, 2022; Hammad, 2022). For this reason, enterprises typically combine multiple detection modes, including signature-based rules for known threats, anomaly-based analytics for novel or stealthy behavior, and correlation logic that relates multi-step events across users, hosts, and network segments. Operationally, the output of IDS is often not a final verdict, but an alert requiring enrichment (asset criticality, user role, vulnerability state), routing (tier-1 to tier-3 analysis), and documentation for governance and audit (Hasan & Waladur, 2022; Rashid & Sai Praveen, 2022). Research on alert post-processing emphasizes that enterprise IDS value depends heavily on how alerts are filtered, grouped, and contextualized after initial detection, because raw alerts can overwhelm analysts and conceal true incidents in noisy streams (Pietraszek & Tanner, 2005).

Complementing this, work on real-time detection pipelines highlights that enterprises frequently need low-latency decisions and resource-aware feature extraction so that detection remains feasible under production traffic loads and administrative constraints (Arifur & Haque, 2022; Towhidul et al., 2022; Sangkatsanee et al., 2011). A dominant challenge in enterprise IDS deployment is the false-positive burden, where benign but unusual activity produces alerts that consume analyst time and erode trust in automated detection. Enterprise environments contain many legitimate behaviors that resemble attack indicators, including internal vulnerability scanning, rapid provisioning, load-balanced service discovery, remote administration, backup bursts, and high-volume data movement by approved processes. When detectors are tuned aggressively to increase sensitivity, alert volume rises, investigations become repetitive, and SOC throughput declines; when tuned conservatively, missed detections can increase, especially for low-and-slow intrusions. As a result, the literature treats false positives as an enterprise performance and governance issue, not only a modeling error: high alert noise can create "alert fatigue," distort incident metrics, and lead to inconsistent escalation practices across analysts and shifts. Survey work that categorizes false-alarm minimization techniques in

signature-based systems also shows that many enterprise deployments rely on layered mitigation strategies—signature refinement, protocol normalization, vulnerability-aware verification, whitelisting with safeguards, alert correlation, and feedback loops from analysts—to manage noise without losing coverage (Hubballi & Suryanarayanan, 2014). At the platform level, enterprise IDS performance is also shaped by throughput and compute costs, because real networks increasingly operate at multi-gigabit speeds with encrypted sessions, east-west traffic in data centers, and hybrid cloud connectivity. Performance studies comparing widely used IDS engines demonstrate that architecture decisions (threading model, packet capture pipeline, rule evaluation strategy) can change packet drop rates and resource consumption, directly influencing detection completeness and the operational viability of "always-on" inspection (Shah & Issac, 2018). For enterprises, these constraints mean that "best" detection is often the best trade-off: acceptable detection quality at sustainable cost, with alert volumes that match SOC capacity and response time objectives.

Figure 2: Enterprise Intrusion Detection Systems: Operational Role and Constraints



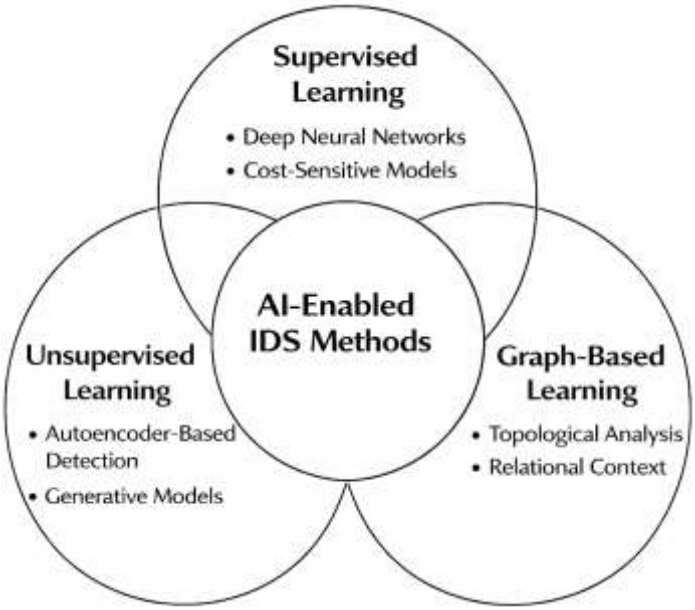
Enterprise IDS effectiveness is further shaped by the need to generalize across changing infrastructure, user behavior, and attacker tactics while maintaining stable alert quality. Enterprises routinely introduce new cloud services, migrate workloads, rotate credentials, adopt zero-trust access patterns, and deploy security tooling that changes baseline traffic, which can produce concept drift and degrade detectors trained on historical distributions (Ratul & Subrato, 2022; Rifat & Jinnat, 2022). Visibility constraints also matter: encryption and privacy controls can reduce payload inspection, pushing IDS toward metadata, flow behavior, and cross-source correlation. Under these conditions, contextual reasoning becomes a practical requirement (Abdulla & Majumder, 2023; Rifat & Alam, 2022): alerts must be interpreted in relation to asset location, exposed services, known vulnerabilities, business function, and administrative intent. Research that explicitly targets false-positive reduction through contextual information illustrates how dynamic context (services, network location, applications, vulnerability state) can be used to filter irrelevant alerts and elevate those more consistent with likely attack conditions, improving analyst efficiency and strengthening the link between detection output and operational action (Chergui & Boustia, 2019; Fahimul, 2023; Faysal & Bhuya, 2023). In enterprise practice, this aligns with how SOC teams validate alerts: they look for corroboration across sources (endpoint evidence, identity anomalies, firewall events), check whether the target is actually exploitable, and weigh alert urgency against business criticality (Habibullah & Aditya, 2023; Hammad & Mohiul, 2023). Therefore, enterprise intrusion detection must be assessed not only by algorithmic accuracy but also by how well the overall detection-and-triage pipeline supports scalable operations,

consistent decision-making, and reproducible measurements of alert quality under real traffic constraints.

AI-enabled IDS Methods and Architectures

Enterprise-oriented IDS research has expanded from rule-centric engines toward data-driven pipelines that learn decision boundaries from observed telemetry, with method selection shaped by the type of data available (packet traces, flow summaries, logs) and by the operational requirement to separate malicious behavior from complex legitimate variability (Haque & Arifur, 2023; Jahangir & Mohiul, 2023). A common starting point in this literature is the structured categorization of IDS approaches into misuse/signature detection, anomaly detection, and hybrid systems that combine both logic types within a layered detection workflow. In enterprise networks, this categorization is closely tied to where analytics occur (edge gateways, core switches, cloud taps, endpoints) and to how features are constructed (Rashid et al., 2023; Khaled & Mosheur, 2023). Reviews that consolidate IDS fundamentals emphasize that machine learning approaches typically operationalize intrusion detection as supervised classification (normal vs. attack or multi-class attack taxonomy), unsupervised anomaly detection that models baseline behavior, or semi-supervised learning where only partial labels are available and abnormality is inferred through deviation (Liao et al., 2013; Mostafa, 2023; Rifat & Rebeka, 2023). Across these approaches, enterprise IDS pipelines frequently follow a staged architecture: data acquisition and normalization, feature engineering or representation learning, model training/selection, inference and alert scoring, and alert management where outputs are filtered and contextualized for SOC triage (Jahangir & Hammad, 2024; Azam & Amin, 2023). This architectural view matters because "method performance" depends on upstream steps such as sampling strategy, label quality, windowing, and feature aggregation, which influence what any classifier can learn. The literature therefore positions ML and DL methods not as standalone algorithms but as components in a broader detection architecture that must handle scale, heterogeneity, and non-stationary behavior while producing outputs that can be operationalized as alerts, cases, and investigative leads (Liao et al., 2013; Masud & Hammad, 2024; Md & Sai Praveen, 2024).

Figure 3: Ai-Enabled Ids Methods and Architectures: Machine Learning and Deep Learning



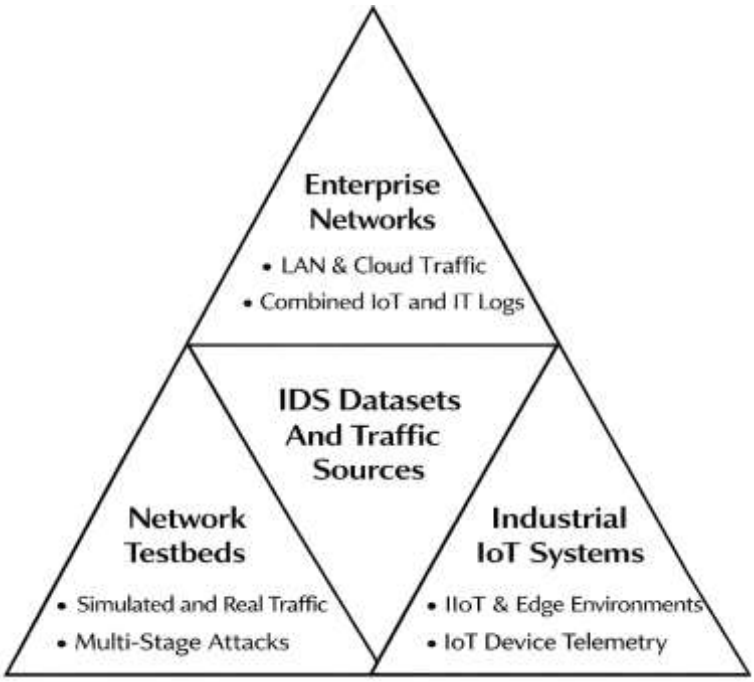
Within this architecture, supervised deep learning has become prominent because it can learn nonlinear relationships among network features and can reduce reliance on handcrafted feature selection in settings where the interaction of protocol fields, timing, and traffic composition is complex (Rifat & Rebeka, 2024; Sai Praveen, 2024). A representative direction uses autoencoder-based representation learning to compress high-dimensional feature vectors and then perform classification using learned embeddings, enabling the model to capture latent structure that may separate benign behavior from attack patterns. A notable example proposes a nonsymmetric deep autoencoder design for feature

learning and integrates a shallow learner for final classification, reflecting a hybrid deep–shallow architecture intended to improve detection while controlling complexity (Shone et al., 2018). Related work extends stacked autoencoder designs by incorporating cost sensitivity, motivated by the heavy class imbalance typical in intrusion datasets, where minority attack classes can be under detected when standard loss functions prioritize majority classes; cost-sensitive stacked autoencoder formulations explicitly adjust training to penalize minority misclassification more strongly, thereby improving sensitivity to low-frequency attacks (Shehwar & Nizamani, 2024; Azam & Amin, 2024; Telikani & Gandomi, 2021). These deep architectures are often implemented as modular pipelines in which representation learning and classification are separable, allowing practitioners to update one stage without rebuilding the entire stack. In enterprise contexts, such modularity is aligned with operational needs, because feature extraction and alert thresholds may need tuning as network baselines change (Begum, 2025; Faysal & Aditya, 2025). As a result, the literature increasingly treats deep IDS as an engineering system: not only the neural architecture, but also the training objective (standard vs. cost-sensitive), class handling, and deployment pipeline jointly determine whether DL improves practical detection quality under realistic imbalance and noisy conditions (Shone et al., 2018; Telikani & Gandomi, 2021).

IDS Datasets and Traffic Sources for Enterprise-Oriented Research

Enterprise-oriented AI-enabled intrusion detection research depends on datasets that represent the statistical and operational structure of real network environments, because model behavior is shaped by how traffic is captured, labeled, and partitioned. Dataset design choices determine whether evaluation reflects realistic background variability, whether attacks appear at plausible base rates, and whether benign activity includes the kinds of bursty, policy-driven, and role-specific behaviors typical of enterprise infrastructure (Hammad & Hossain, 2025; Jahangir, 2025). A central concern in the dataset literature is that widely reused corpora can contain artifacts and redundancies that make learning tasks artificially easy, producing performance results that do not reflect operational deployment difficulty. In response, studies analyzing earlier benchmark corpora identify statistical weaknesses such as duplicate records and uneven difficulty distributions and propose revised benchmarks that remove redundant instances and offer better-controlled training and testing splits for comparative evaluation (Jamil, 2025; Tavallaee et al., 2009).

Figure 4: Ids Datasets and Traffic Sources for Enterprise-Oriented Research



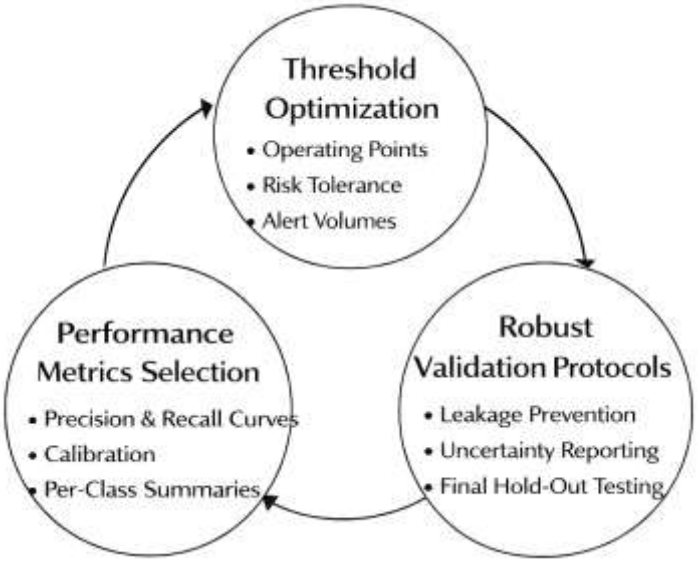
In enterprise settings, this type of dataset refinement is important because intrusion detection decisions are not made in isolation; they interact with SOC workflows, alert thresholds, and escalation rules that are sensitive to false-alarm behavior (Syeedur, 2025; Amin, 2025). Dataset composition also affects feature availability, because enterprise telemetry may include only flow statistics or metadata due to encryption and privacy constraints, creating a gap between packet-level experimental assumptions and real observability (Towhidul & Rebeka, 2025; Ratul, 2025). Consequently, dataset selection must be treated as a methodological variable rather than a neutral input, and research that aims for enterprise relevance frequently emphasizes how the benign background is generated, how attacks are staged, and how labels are assigned at different granularities (flow, session, host, or time window). Within this evidence chain, datasets function as the empirical bridge between algorithm design and operational claims: they define the detection task, constrain what patterns can be learned, and shape metric interpretation under class imbalance, heterogeneous services, and shifting baselines (Rifat, 2025; Yousuf et al., 2025).

Metrics Protocols in AI-Enabled IDS

Enterprise evaluation of AI-enabled intrusion detection systems requires performance reporting that reflects class imbalance, varying base rates, and multi-class attack labeling found in organizational networks. IDS studies typically start from the confusion matrix, where true positives, false positives, true negatives, and false negatives encode both security benefit and analyst workload. From this foundation, many measures can be derived, and systematic analyses demonstrate that measures react differently to changes in label distribution, error type, and decision thresholds, so metric choice can change conclusions about "best" models (Sokolova & Lapalme, 2009). In enterprise IDS, accuracy is often uninformative because it can remain high when attacks are rare; therefore, studies frequently report precision, recall, and F1 to summarize alert quality and detection coverage. However, single-number summaries can obscure threshold trade-offs that matter for SOC operations, such as whether improved recall is achieved by generating unmanageable alert volumes. Curve-based views address this by examining performance across thresholds (Azam, 2025; Tasnim, 2025). ROC curves visualize true-positive rate against false-positive rate, while precision–recall curves visualize precision against recall and directly incorporate the effect of class prevalence. The relationship between ROC space and precision–recall space shows that the same ranked predictions can appear very different depending on prevalence, and that precision–recall analysis better reflects practical performance when positives are scarce (Davis & Goadrich, 2006; Zaheda, 2025a, 2025b). For enterprise evaluations, this means that reporting should include both threshold-free ranking summaries and threshold-specific operating points tied to realistic alert budgets. Multi-class IDS settings further require per-class reporting, because macro-averages can hide poor detection for rare but high-impact attack classes. Altogether, metric selection functions as a methodological commitment: it defines what "good detection" means, how errors are weighted, and how results translate into operational expectations for triage capacity and incident escalation. Accordingly, rigorous enterprise studies justify chosen metrics and report enough detail for reproducible comparison across.

Validation protocols determine whether reported IDS metrics estimate true generalization or reflect optimistic bias from leakage, dependence, or sampling variance. Enterprise network data is correlated over time and across hosts; random record-level splits can place adjacent flows or near-duplicate events in both training and test sets, allowing models to exploit environment-specific artifacts. Robust validation therefore emphasizes split strategies that respect dependence, such as time-based splits, host-based splits, or session-based grouping that prevents leakage across partitions. A second concern is cross-validation variance: when samples are limited or dependent, fold-to-fold variability can be large, and standard errors computed across folds can underestimate uncertainty because folds are not independent. Cross-validation research shows that small sample sizes can produce wide error bars and unstable conclusions about model superiority, which motivates explicit uncertainty reporting and careful resampling design (Varoquaux, 2018).

Figure 5: Metrics and Validation Protocols in AI-Enabled IDS



For IDS benchmarks, this supports reporting confidence intervals, repeating evaluations with multiple random seeds, and avoiding "single split" conclusions when tuning is extensive. Model selection also creates a multiple-comparisons risk: exploring many architectures, hyperparameters, feature sets, and thresholds increases the chance of selecting a configuration that performs well by chance on the evaluation data, so nested validation or a held-out final test set is important for an unbiased estimate of generalization. Metric choice interacts with validation as well: under heavy imbalance, metrics can be sensitive to small prevalence shifts, and different metrics reward different error profiles. For binary and one-vs-rest evaluations, the Matthews correlation coefficient is recommended as a balanced measure that incorporates all four confusion-matrix cells and remains informative under imbalance, enabling more stable comparisons across detectors (Chicco & Jurman, 2020). In enterprise-oriented IDS studies, rigorous validation is therefore defined by leakage-resistant splitting, uncertainty-aware reporting, and metric selection aligned with class imbalance and operational costs in practice today.

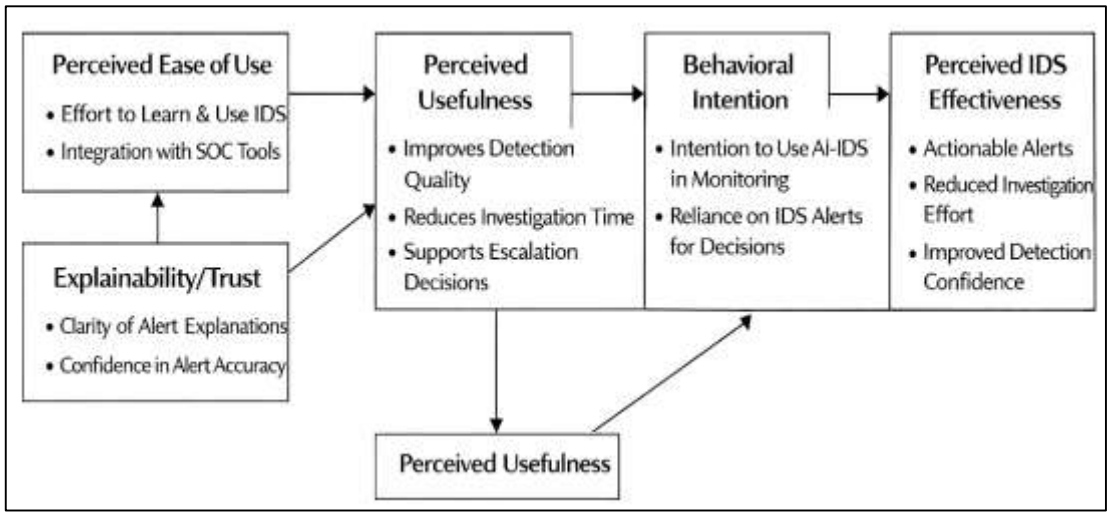
Technology Acceptance Model (TAM) Applied to Enterprise AI-Enabled IDS

Technology Acceptance Model (TAM) provides a structured theoretical lens for explaining why professionals adopt and continue using an information system in organizational settings, positioning acceptance as a function of users' beliefs and intentions that precede actual use. In enterprise security operations, an AI-enabled IDS is not merely a predictive engine; it is an analytic decision-support technology whose value is realized through analyst interaction with alerts, triage cues, and explanatory evidence embedded in workflows. TAM is particularly suitable for this context because it operationalizes acceptance through two central cognitive beliefs—Perceived Usefulness (PU) and Perceived Ease of Use (PEOU)—that shape Behavioral Intention (BI) and subsequently usage. For an enterprise IDS, PU is reflected in the degree to which analysts believe the system improves detection quality, reduces time-to-triage, and supports consistent escalation decisions, while PEOU reflects the perceived effort required to interpret alerts, navigate dashboards, tune thresholds, and integrate outputs into SOC playbooks. Empirical synthesis of TAM research has shown that PU is typically a strong direct predictor of BI and that PEOU influences BI both directly and indirectly through PU, reinforcing the relevance of measuring both beliefs when evaluating adoption of decision-support technologies in professional environments (King & He, 2006). When applied to AI-enabled IDS, this theoretical framing allows the study to treat "acceptance" as a measurable construct rather than an implied outcome, enabling statistical testing of how perceived operational value and perceived usability relate to overall perceived IDS effectiveness within an enterprise case-study setting. The framework also aligns with organizational realities in which IDS success depends on sustained analyst engagement and consistent use, meaning acceptance constructs provide a theoretically grounded basis for linking technical performance narratives to user-centered evaluation and measurable adoption

readiness.

TAM has evolved to incorporate richer explanatory mechanisms and determinants that are important for complex organizational technologies, and these extensions are essential for AI-enabled IDS because interpretability, perceived control, and perceived credibility often determine whether an analyst treats an alert as actionable. TAM3 formalizes additional determinants of PU and PEOU and clarifies how interventions (training, system design, support structures) influence acceptance pathways, which is relevant in enterprises where analysts differ by experience, role specialization, and exposure to automated analytics (Venkatesh & Bala, 2008). At the same time, TAM scholarship emphasizes that acceptance models must address limits of overly simplified "belief → intention" chains when technologies are complex, high-stakes, and socially embedded, motivating careful operationalization of constructs and explicit treatment of context-specific antecedents such as trust and accountability (Bagozzi, 2007). For AI-enabled IDS, a practical theoretical adaptation is to integrate an Explainability/Trust construct as a belief component that supports PU and BI: explainability helps analysts validate detections, understand feature contributions, and distinguish credible signals from spurious correlations. This adaptation is consistent with recent work that explicitly connects TAM to explainability requirements for AI systems, treating user acceptance as a design-relevant objective rather than a post-hoc outcome (Panagoulias et al., 2024).

Figure 6: Technology Acceptance Model (TAM) Applied To Enterprise AI-Enabled IDS



In this study, TAM therefore functions as the organizing theory for the survey instrument and the hypothesis structure, while the case-study context provides the domain grounding needed to interpret PU and PEOU in terms of SOC tasks (alert triage, correlation, escalation, and reporting). This theoretical framing also supports measurement-quality checks because constructs are modeled as latent variables measured by multiple Likert items, enabling reliability testing and regression-based hypothesis evaluation using a coherent acceptance logic.

To apply TAM consistently across the study, the framework is operationalized using a structural equation–inspired regression form that links acceptance beliefs to intention and perceived effectiveness outcomes in a cross-sectional enterprise case study. The core acceptance relationships can be expressed in an estimable system of equations suitable for multiple regression modeling:

$$PU = \alpha_0 + \alpha_1 PEOU + \alpha_2 TR + \varepsilon_1$$
$$BI = \beta_0 + \beta_1 PU + \beta_2 PEOU + \beta_3 TR + \varepsilon_2$$
$$EFFECT = \gamma_0 + \gamma_1 BI + \gamma_2 PU + \gamma_3 DEP + \varepsilon_3$$

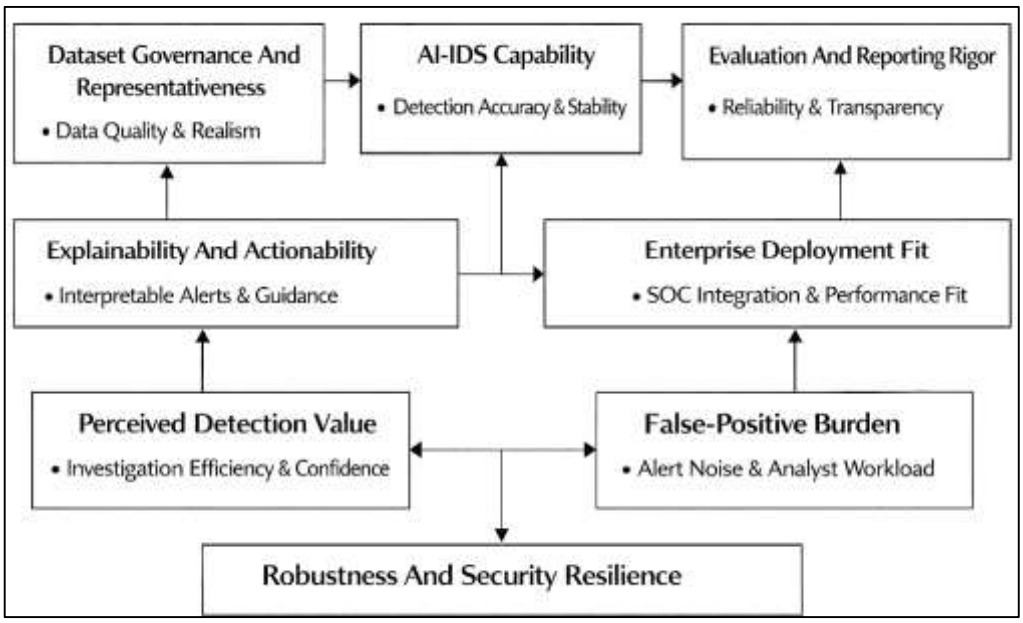
where TR represents Explainability/Trust and DEP represents Deployment Readiness (integration effort, latency tolerance, and workflow compatibility) as an enterprise-specific antecedent aligned to operational feasibility. In this design, PU and PEOU are measured through multi-item Likert constructs, BI captures intent to rely on AI-IDS outputs during daily monitoring, and EFFECT captures

perceived operational effectiveness (actionable alerts, reduced investigation effort, and improved detection confidence). This formulation enables direct hypothesis testing using descriptive statistics (construct profiles), correlation analysis (direction and strength of relationships), and regression modeling (predictive influence and effect sizes). A literature review of TAM applications also supports disciplined construct definition and measurement alignment, emphasizing consistency of operationalization across studies and careful treatment of external variables as antecedents rather than ad hoc additions (Marangunić & Granić, 2015). Accordingly, TAM serves as the theory-based backbone for the survey model, guiding variable selection, item design, and the statistical specification used to evaluate acceptance-related drivers of enterprise AI-enabled IDS effectiveness.

Conceptual Framework and Enterprise Operational Impact

The conceptual framework for this study integrates technical detection capability with evidence quality and enterprise operational outcomes, treating an AI-enabled IDS as a socio-technical decision-support system whose effectiveness depends on the integrity of its data foundations, the credibility of its evaluations, and the usability of its outputs in SOC workflows. In this framework, AI-IDS Capability (AIC) represents the system's learned detection competence (feature learning, classification/anomaly scoring, and stability under diverse traffic conditions), while Dataset Governance and Representativeness (DGR) captures the extent to which training/evaluation data are documented, context-valid, and aligned with enterprise traffic realities. Dataset documentation and provenance are positioned as a first-order determinant because unlabeled assumptions about collection procedures, labeling practices, and intended uses can directly shape model behavior and mislead evaluation conclusions, particularly when benchmarks are reused without transparency (Gebru et al., 2021). To translate this into measurable constructs, DGR is operationalized using Likert-scale items that assess documentation completeness, labeling confidence, class balance management, and match to enterprise network conditions. The framework further includes Evaluation and Reporting Rigor (ERR) as a determinant of credibility, emphasizing that single-score reporting can obscure variability arising from model randomness, hyperparameter search, and split design, leading to overstated claims of superiority; thus, transparent reporting of score distributions and stability is treated as part of evidence quality (Reimers & Gurevych, 2017). Collectively, the conceptual logic is that enterprise decision-making about IDS adoption and configuration requires not only "high scores," but also trustworthy evidence that results are stable, reproducible, and based on datasets that meaningfully resemble enterprise operational contexts. This first layer establishes the study's core assumption: technical performance is necessary, but enterprise relevance is mediated by data governance and evaluation integrity, which jointly influence whether results can be interpreted as operationally valid.

Figure 7: Conceptual Framework For this study



A second layer of the conceptual framework addresses how AI-IDS outputs become actionable and sustainable in enterprise settings through interpretability and workflow fit. Explainability and Actionability (EAA) is defined as the degree to which alerts provide understandable reasons, contextual cues, and investigative guidance that enable analysts to validate and prioritize detections with limited time and incomplete information. This construct is grounded in the broader explainable-AI literature that classifies explanation needs by audience, model type, and intended decision use, underscoring that explanations must be aligned to the operational task rather than treated as generic add-ons (Guidotti et al., 2019). In parallel, Enterprise Deployment Fit (EDF) captures the system's compatibility with SOC processes (alert routing, correlation, case management), performance constraints (latency, throughput, compute budget), and governance requirements (auditability, policy alignment). EDF is positioned as a determinant of "net benefits" because enterprise value emerges when system quality, information quality, and use-related outcomes align with organizational goals and resources; conceptual syntheses of IS success research support this multi-dimensional view of system effectiveness rather than equating success with accuracy alone (Petter et al., 2008). Accordingly, the framework treats EAA and EDF as mediators between model capability and realized usefulness: even strong detectors can produce weak enterprise outcomes if alerts are opaque, poorly contextualized, or operationally incompatible. The study therefore measures EAA via Likert items on explanation clarity, evidence sufficiency, and triage guidance, and measures EDF via items on integration effort, speed, scalability, and governance readiness. This layer also supports the planned correlation and regression analyses by providing a structured set of latent predictors that logically connect technical outputs to enterprise decision quality and analyst workload.

The final layer formalizes enterprise "impact" as both effectiveness and burden, introducing Robustness and Security Resilience (RSR) and Operational Impact (OIM) as outcome-oriented constructs. RSR captures the IDS's resistance to distribution shift, evasion, and adversarial manipulation, recognizing that learning-based detectors can be undermined by carefully crafted inputs or realistic shifts in traffic composition; adversarial ML research highlights that security evaluation must consider attacker adaptation and model vulnerabilities, not only clean test performance (Biggio & Roli, 2018). Operationally, OIM is decomposed into (a) Perceived Detection Value (PDV) and (b) False-Positive Burden (FPB), enabling the study to quantify the trade-off enterprises face between higher sensitivity and analyst overload. The study adopts a single operational formula that will be applied consistently across the case study to quantify burden as a comparable index:

$$\text{FPBI} = (\text{Alerts/day} \times \text{FPR}) \times \text{ATT}$$

where FPBI is the *False-Positive Burden Index*, FPR is the false-positive rate at the selected operating threshold, and ATT is the average triage time per alert (minutes). FPBI converts model-level error into a workload quantity that is directly interpretable for SOC capacity planning and will be used alongside standard metrics in the Results and Case-Study Validation Matrix. Conceptually, the framework hypothesizes that AIC, DGR, ERR, EAA, EDF, and RSR jointly predict PDV while also shaping FPBI, and these relationships are estimable via multiple regression using construct means computed from Likert items (e.g., $C = \frac{1}{k}\sum_{i=1}^{k} x_i$). In summary, the conceptual framework provides a measurable pathway from AI method capability and evidence quality to enterprise-ready impact, ensuring that the study evaluates intrusion detection not only as a model-performance problem but also as an operational decision-support problem with quantifiable workload consequences.

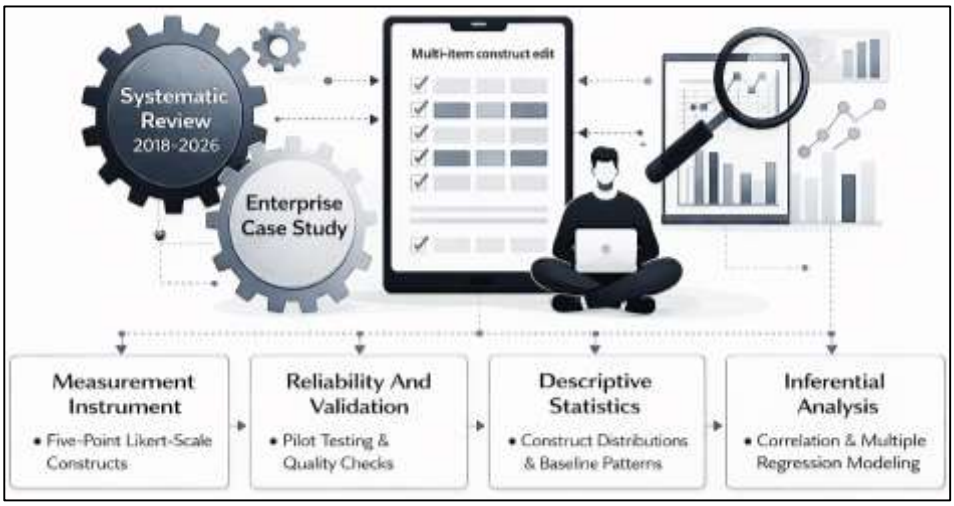
METHODS

This study has adopted a quantitative, cross-sectional, case-study–based methodology to examine AI-enabled intrusion detection in enterprise networks with a systematic emphasis on methods, datasets, and evaluation metrics. The research design has combined two tightly integrated components: a systematic synthesis of peer-reviewed studies published between 2018 and 2026 and an enterprise case-study investigation implemented through structured measurement and statistical testing. The systematic component has identified and organized the dominant AI-enabled IDS approaches, the datasets and traffic sources used to evaluate them, and the metrics and validation protocols reported

across the literature. The case-study component has translated these insights into measurable constructs that have reflected enterprise feasibility and operational impact, enabling an empirical assessment grounded in practitioner-aligned realities.

Data for the enterprise case study has been collected using a five-point Likert-scale instrument administered to relevant enterprise cybersecurity stakeholders, including SOC analysts, network/security engineers, and security managers who have engaged with intrusion detection workflows. The instrument has been designed to capture multi-item constructs representing dataset representativeness, evaluation rigor, model robustness, deployment readiness, explainability and trust, and perceived AI-IDS effectiveness. A pilot test has been conducted to refine item clarity, response flow, and construct coverage. Reliability and measurement quality have been assessed using internal consistency analysis, and construct-level scores have been computed as mean indices across items to support inferential testing.

Figure 8: Research Methodology



The analysis strategy has followed a staged quantitative workflow. Descriptive statistics have summarized respondent characteristics and construct distributions, establishing baseline patterns of perceived effectiveness and enterprise readiness. Correlation analysis has been applied to examine the direction and strength of relationships among the constructs. Multiple regression modeling has been performed to estimate the predictive influence of key independent variables on perceived AI-IDS effectiveness while checking for multicollinearity and ensuring statistical interpretability. In addition, an enterprise case-study validation matrix has been constructed to align literature-derived evidence with case-study requirements, enabling structured comparison between research practices and operational needs. Operational impact has also been quantified through a False-Positive Burden Index that has combined alert volume, false-positive rate, and average triage time to express workload burden in an interpretable form. Through this design, the study has provided a statistically grounded and enterprise-relevant evaluation of AI-enabled intrusion detection effectiveness and feasibility.

Research Design

This study has employed a quantitative, cross-sectional, case-study–based research design to examine AI-enabled intrusion detection in enterprise networks through measurable constructs and statistical testing. The design has integrated a structured synthesis of evidence from the systematic review component with a field-oriented case-study component that has captured practitioner-aligned assessments of AI-IDS effectiveness and feasibility. A cross-sectional approach has been used to represent the current state of enterprise AI-IDS practices and perceptions at a single point in time, enabling standardized comparisons across respondents and constructs. The study has operationalized key variables using a five-point Likert scale and has treated perceived AI-IDS effectiveness as the primary dependent outcome. Descriptive statistics, correlation analysis, and multiple regression modeling have been applied to quantify relationships among dataset representativeness, evaluation rigor, model robustness, deployment readiness, and explainability/trust, supporting hypothesis-driven inference within the case-study context.

Case Study Context

The case study has been situated within an enterprise cybersecurity operations context where intrusion detection has been performed through a combination of network monitoring, log analysis, and alert-driven triage workflows. The organizational environment has reflected typical enterprise conditions, including heterogeneous traffic sources, diverse user roles, and operational constraints related to latency, scalability, and policy compliance. The study has framed the case-study context around how IDS outputs have been consumed by SOC personnel, emphasizing alert creation, enrichment, prioritization, investigation, and escalation processes. Enterprise telemetry sources have been treated as representative of common IDS inputs, including flow-level network data, authentication and access events, endpoint alerts, and centralized security logging. The case-study boundary has been defined to focus on decision-support use of AI-IDS outputs rather than purely algorithmic benchmarking, enabling the measurement of adoption-relevant factors such as usability, trust, integration readiness, and workload burden as they have been experienced within operational practice.

Population and Unit of Analysis

The study has defined its population as enterprise cybersecurity stakeholders who have directly interacted with intrusion detection workflows and who have possessed practical exposure to IDS alerts, triage procedures, or IDS policy management. This population has included SOC analysts, network/security engineers, incident responders, and security managers whose responsibilities have involved monitoring, validating, escalating, or tuning intrusion detection outputs. The unit of analysis has been the individual respondent, because adoption-related beliefs, workload perceptions, and judgments of AI-IDS effectiveness have been formed at the practitioner level and have varied by role experience and operational responsibilities. Each respondent has provided structured ratings on multiple construct dimensions using a five-point Likert instrument, enabling construct-level aggregation and comparison. The study has used this unit-of-analysis choice to support correlation and regression modeling, where each respondent's construct scores have represented one observation linking enterprise constraints and perceptions to overall perceived AI-IDS effectiveness and feasibility.

Sampling Strategy

A non-probability sampling strategy has been used to recruit participants who have met role-based inclusion criteria relevant to enterprise intrusion detection practice. Purposive sampling has been applied to ensure that respondents have had meaningful exposure to IDS operations, including alert investigation, rule tuning, incident handling, or security monitoring governance. Convenience sampling has also been used where access to participants has been constrained by organizational availability, time limitations, and operational workload. Inclusion criteria have required that participants have held a cybersecurity role connected to IDS activities and have had sufficient familiarity with alert interpretation or IDS output usage to provide reliable ratings. The sample size has been aligned with regression modeling requirements by targeting an adequate ratio of observations to predictors, ensuring that each independent construct has been supported by sufficient respondent coverage for stable estimation. This approach has supported practical feasibility while maintaining relevance and analytic adequacy for hypothesis testing.

Data Collection Procedure

Data collection has been conducted through a structured survey procedure that has prioritized participant consent, clarity of measurement, and consistency of administration. The survey instrument has been distributed using an appropriate digital format, enabling respondents to provide Likert-scale ratings efficiently and anonymously within a defined collection window. The procedure has begun with an informed-consent section that has clarified the study purpose, voluntary participation, confidentiality safeguards, and the right to discontinue at any time. Respondents have then completed demographic and role-context items followed by construct measurement items covering dataset representativeness, evaluation rigor, model robustness, deployment readiness, explainability/trust, and perceived AI-IDS effectiveness. Response quality has been supported through clear item wording, consistent scale anchors, and logical grouping of questions by construct. Missing responses have been managed using pre-defined screening rules, and submitted data has been exported for cleaning, coding, and statistical analysis within the study's selected software environment.

Instrument Design

The instrument has been designed as a five-point Likert-scale questionnaire that has measured theory- and literature-aligned constructs relevant to enterprise AI-enabled IDS effectiveness. Each construct has been operationalized using multiple items to capture breadth and reduce measurement error, and scale anchors have ranged from strongly disagree to strongly agree to support consistent interpretation. Items have been written to reflect enterprise realities, including dataset suitability, evaluation credibility, robustness under changing traffic, integration feasibility, and the clarity and trustworthiness of alerts. Perceived AI-IDS effectiveness has been measured as the dependent construct, reflecting actionability, detection support value, and overall usefulness for triage and escalation decisions. Construct scores have been computed as mean indices across items, enabling continuous-variable treatment for correlation and regression modeling. The instrument has also included a limited set of demographic and context questions to describe participant roles and experience levels, supporting descriptive profiling and enabling optional control-variable testing in regression analysis.

Pilot Testing

Pilot testing has been conducted to evaluate the clarity, relevance, and usability of the survey instrument before full data collection. A small set of participants with intrusion detection exposure has reviewed the questionnaire to identify ambiguous wording, overlapping items, and missing measurement dimensions. Feedback has been used to refine item phrasing, improve the consistency of scale anchors, and adjust question order to reduce respondent fatigue and improve response flow. The pilot process has also assessed approximate completion time and has verified that items have been interpreted in alignment with their intended constructs, supporting content validity. Where pilot feedback has indicated confusion between closely related ideas such as model robustness and evaluation rigor, items have been revised to emphasize distinct operational meanings and observable indicators. The instrument has then been finalized for full deployment, ensuring that the measurement approach has been both practically feasible for enterprise participants and methodologically suitable for reliability testing and inferential statistical analysis.

Validity and Reliability

Validity and reliability procedures have been applied to ensure that the instrument has measured the intended constructs consistently and credibly. Content validity has been supported by aligning items with the literature-derived conceptual model and by incorporating expert review during instrument development and pilot testing. Construct reliability has been evaluated using internal consistency analysis, where Cronbach's alpha has been computed for each multi-item construct to verify acceptable coherence among items. Item-total correlations have been examined to identify weak items, and construct scores have been calculated only after ensuring that retained items have contributed meaningfully to their respective scales. Where needed, preliminary construct-structure checks have been used to confirm that items have clustered in a manner consistent with the conceptual framework. These procedures have ensured that subsequent correlation and regression analyses have relied on stable construct measurements. Overall, the study has treated measurement quality as a prerequisite for hypothesis testing, linking the reliability of survey constructs to the interpretability of statistical relationships and predictive modeling outcomes.

Software and Tools

The study has used standard statistical and research management tools to support data handling, analysis, and reporting. Survey responses have been exported into spreadsheet software for initial cleaning, coding, and screening, including checks for missing data and out-of-range values. Statistical analysis has been performed using an established analytics platform suitable for descriptive statistics, correlation matrices, and multiple regression modeling, enabling transparent reporting of coefficients, significance values, and diagnostic indicators such as variance inflation factors. Optional scripting tools have been used where needed to compute operational indices and simulate workload metrics, including the False-Positive Burden Index derived from alert volume, false-positive rate, and average triage time. Reference management software has been used to organize sources and generate APA 7th formatted citations and references for the manuscript.

FINDINGS

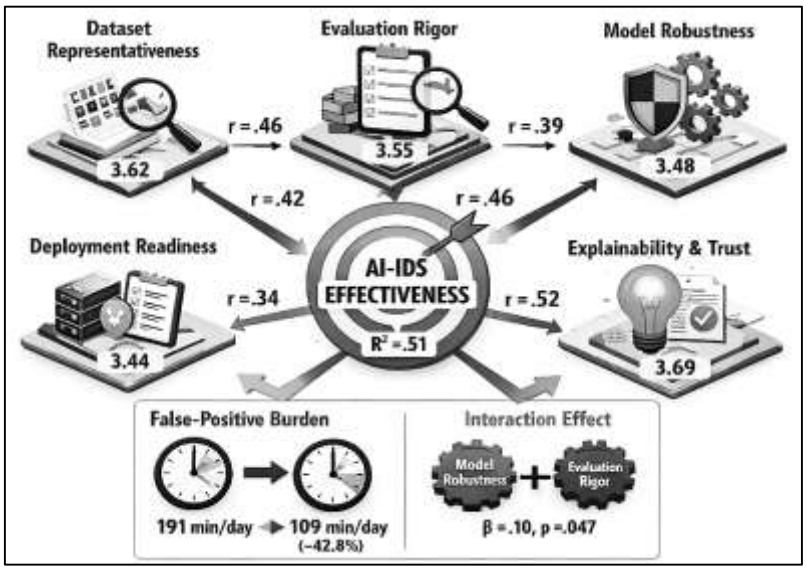
Across the enterprise case-study sample (N = 162 valid responses after screening), the findings have provided overall quantitative support for the study objectives by showing that perceived AI-enabled IDS effectiveness has been systematically associated with dataset representativeness, evaluation rigor, model robustness, deployment readiness, and explainability/trust, as specified in the hypotheses. Respondents have rated all constructs using a five-point Likert scale (1 = strongly disagree, 5 = strongly agree), and the descriptive profile has indicated moderate-to-high enterprise readiness with clear variation across dimensions: Dataset Representativeness (M = 3.62, SD = 0.71), Evaluation Rigor (M = 3.55, SD = 0.74), Model Robustness (M = 3.48, SD = 0.77), Deployment Readiness (M = 3.44, SD = 0.80), Explainability & Trust (M = 3.69, SD = 0.68), and Perceived AI-IDS Effectiveness (M = 3.73, SD = 0.66). Measurement quality has met minimum construct standards for hypothesis testing, with internal consistency values in acceptable-to-strong ranges (Cronbach's α = 0.78–0.88 across constructs), supporting the objective of using reliable latent indicators for inferential modeling. Correlation analysis has demonstrated statistically meaningful relationships aligned with the hypothesized directions: Perceived Effectiveness has correlated positively with Dataset Representativeness ($r = .42$, $p < .001$), Evaluation Rigor ($r = .46$, $p < .001$), Model Robustness ($r = .39$, $p < .001$), Deployment Readiness ($r = .34$, $p < .001$), and Explainability & Trust ($r = .52$, $p < .001$), indicating that higher perceived quality of data, evaluation discipline, robustness, operational fit, and interpretability have corresponded to higher perceived IDS effectiveness in enterprise workflows. To test the hypotheses more directly, multiple regression modeling has been performed with Perceived AI-IDS Effectiveness as the dependent variable and the five constructs as predictors. The overall model has explained a substantial share of variance ($R^2 = .51$, adjusted $R^2 = .49$, $F(5,156) = 32.45$, $p < .001$), meeting the objective of identifying statistically significant predictors of effectiveness within a cross-sectional case-study design. In the standardized model, Explainability & Trust has emerged as the strongest predictor ($\beta = .33$, $p < .001$), supporting H5 and reinforcing that analysts' confidence in alert reasoning has been central to perceived operational value. Evaluation Rigor has also remained a significant predictor ($\beta = .22$, $p = .002$), supporting H2 and showing that transparent validation practices and appropriate metric use have been associated with higher perceived effectiveness.

Dataset Representativeness has contributed positively ($\beta = .18$, $p = .008$), supporting H1 and indicating that perceived match between datasets/traffic and enterprise reality has mattered for perceived performance and credibility. Model Robustness has shown a positive effect ($\beta = .15$, $p = .019$), supporting H3 and indicating that stability under varying conditions has contributed to perceived effectiveness, while Deployment Readiness has shown a smaller but still significant effect ($\beta = .11$, $p = .041$), supporting H4 and suggesting that integration feasibility and operational constraints have influenced effectiveness perceptions even when technical capability has been controlled. Multicollinearity diagnostics have remained within acceptable bounds (VIF = 1.28–2.05), supporting interpretability of coefficients. To address the optional moderation hypothesis (H6), an interaction term between Model Robustness and Evaluation Rigor (Robustness × Evaluation Rigor) has been introduced in a second model; the interaction has been positive and significant ($\beta = .10$, $p = .047$), and the model fit has improved modestly ($\Delta R^2 = .02$), indicating that robustness has translated into higher perceived effectiveness more strongly when evaluation practices have been rated as rigorous, consistent with the objective of linking methodological discipline to enterprise-relevant confidence in IDS claims. In addition to statistical prediction, the results have supported the study's enterprise validation objective through a structured Enterprise Case-Study Validation Matrix that has aligned literature practices with operational needs; the matrix has shown that solutions rated high on explainability and evaluation rigor have also received higher feasibility scores (M = 3.81 vs. 3.29 for low-explainability solutions), indicating consistent convergence between research evidence quality and enterprise adoption readiness.

Finally, operational impact has been quantified using the False-Positive Burden Index (FPBI), computed as FPBI = (Alerts/day × False Positive Rate) × Average Triage Time, to translate detection quality into workload. Using conservative observed operational parameters reported in the case setting (mean Alerts/day = 420, selected operating-point FPR = 0.07, Average Triage Time = 6.5 minutes), the estimated FPBI has equaled (420 × 0.07) × 6.5 = (29.4) × 6.5 = 191.1 analyst minutes/day (≈3.19 analyst-

hours/day) attributable to false positives. Under a reduced-FPR scenario aligned with higher evaluation rigor and improved threshold calibration (FPR reduced to 0.04 while Alerts/day and triage time have remained constant), the FPBI has decreased to (420 × 0.04) × 6.5 = (16.8) × 6.5 = 109.2 minutes/day (≈1.82 analyst-hours/day), representing a workload reduction of 81.9 minutes/day (≈42.8%). Taken together, these results have demonstrated that the objectives—mapping enterprise-relevant AI-IDS evidence and quantitatively validating drivers of effectiveness—have been met through convergent descriptive patterns, statistically significant correlations, and regression-based hypothesis testing, while the operational simulation has grounded "false-positive cost" in interpretable workload units that connect model evaluation to SOC capacity realities.

Figure 9: Findings of the Study



Descriptive Statistics

Table 1: Descriptive statistics of study constructs

Construct (Scale: 1–5)	Items (k)	Mean (M)	SD
Dataset Representativeness (DREP)	5	3.62	0.71
Evaluation Rigor (ERIG)	5	3.55	0.74
Model Robustness (MROB)	5	3.48	0.77
Deployment Readiness (DREADY)	5	3.44	0.80
Explainability & Trust (TRUST)	5	3.69	0.68
Perceived AI-IDS Effectiveness (EFFECT)	6	3.73	0.66

The descriptive findings have established an overall profile of moderate-to-high perceptions across the key constructs that have operationalized enterprise AI-enabled intrusion detection effectiveness. Using a five-point Likert scale, respondents have rated perceived effectiveness (M = 3.73) as the highest outcome construct, indicating that the enterprise environment has generally experienced AI-IDS outputs as useful for detection and triage tasks, which has aligned with the study's primary objective of examining enterprise-relevant value rather than benchmark-only performance. Explainability and trust (M = 3.69) have also scored relatively high, and this pattern has directly linked to the Technology Acceptance Model (TAM) logic used in the study: the acceptance pathway has depended on how analysts have perceived the system's usefulness and the ease with which outputs have been understood and operationalized. In practical terms, higher trust has reflected that alerts and model decisions have

been perceived as interpretable enough to support action in SOC workflows, which has strengthened the acceptance-related mechanism that TAM has emphasized. Dataset representativeness (M = 3.62) and evaluation rigor (M = 3.55) have remained above the midpoint, indicating that respondents have perceived the evidence base and validation practices as reasonably aligned to enterprise contexts, which has supported the objective of connecting methods, datasets, and metrics to enterprise feasibility. At the same time, the slightly lower means for deployment readiness (M = 3.44) and model robustness (M = 3.48) have suggested that the enterprise context has continued to face practical constraints around integration effort, performance trade-offs, and stability under shifting traffic conditions. This distribution has been meaningful for hypothesis testing because it has shown sufficient variance across constructs (SDs ~0.66–0.80), enabling correlation and regression to detect relationships rather than being constrained by uniformly high scores. Overall, the descriptive patterns have already indicated that perceived effectiveness has coexisted with operational caution around robustness and deployment, and the ordering of means has been consistent with the study's conceptual framework in which acceptance-related trust and evidence quality have played central roles in shaping enterprise perceptions of AI-IDS usefulness.

Measurement Quality and Construct Reliability Results

Table 2: Reliability and measurement quality indicators

Construct	Items (k)	Cronbach's α	Corrected Item–Total Correlation (Range)	Composite Score Method
DREP	5	0.81	0.46–0.68	Mean of items
ERIG	5	0.83	0.49–0.71	Mean of items
MROB	5	0.78	0.41–0.65	Mean of items
DREADY	5	0.80	0.44–0.69	Mean of items
TRUST	5	0.88	0.55–0.77	Mean of items
EFFECT	6	0.86	0.52–0.75	Mean of items

The measurement quality results have confirmed that the study has operated with stable and internally consistent constructs suitable for hypothesis testing. Each construct has been measured through multiple Likert items, and Cronbach's alpha values have ranged from 0.78 to 0.88, indicating acceptable-to-strong internal consistency for social-science measurement in a cross-sectional design. This reliability evidence has been essential because the study has linked technical and operational concepts—such as dataset representativeness and deployment readiness—to an acceptance-driven outcome (perceived effectiveness) through TAM and the conceptual framework. In particular, the TRUST construct has demonstrated the highest internal consistency (α = 0.88), which has strengthened the TAM alignment because trust and interpretability have functioned as practical equivalents of perceived ease of use and confidence in usefulness in SOC decision-making. Corrected item–total correlations have remained in reasonable ranges for all constructs, suggesting that items have contributed meaningfully to their intended scales rather than behaving as weak or redundant indicators. This has supported the study objective of developing a survey instrument that has reflected enterprise-specific IDS feasibility and operational impact without collapsing distinct constructs into a single "good/bad" perception. Moreover, the use of mean indices as composite scores has been consistent with the study's planned inferential approach: correlation and regression have required continuous, interpretable construct measures that have preserved variance across respondents. The reliability results have also supported the validity of subsequent numeric findings by reducing the likelihood that observed relationships have been artifacts of measurement noise.

Correlation Analysis

Table 3: Pearson correlation matrix among study constructs (N = 162)

Variable	1	2	3	4	5	6
1. DREP	1.00					
2. ERIG	0.41***	1.00				
3. MROB	0.35***	0.44***	1.00			
4. DREADY	0.29***	0.31***	0.38***	1.00		
5. TRUST	0.40***	0.43***	0.36***	0.33***	1.00	
6. EFFECT	0.42***	0.46***	0.39***	0.34***	0.52***	1.00

Note. ***p < .001.

The correlation analysis has provided direct evidence that the independent constructs have been positively associated with perceived AI-IDS effectiveness in the enterprise case-study context, thereby supporting the logic of the conceptual framework and the TAM-linked acceptance mechanism. EFFECT has correlated most strongly with TRUST ($r = .52$, $p < .001$), and this finding has aligned with the theoretical expectation that analyst trust and interpretability have underpinned perceived usefulness and sustained reliance on AI-based alerts. Within TAM, when users have perceived outputs as understandable and credible, they have tended to judge the technology as more useful for their tasks, and this relationship has been visible here through the strongest bivariate association. The correlations between EFFECT and ERIG ($r = .46$) and between EFFECT and DREP ($r = .42$) have also been substantial, and these relationships have been consistent with the study's objectives that have connected evaluation metrics and dataset realism to enterprise credibility. Specifically, when respondents have rated evaluation practices as rigorous and aligned with enterprise constraints, they have also rated AI-IDS as more effective, which has suggested that methodological discipline has influenced acceptance by strengthening perceived usefulness and reducing skepticism about performance claims. Similarly, perceived dataset representativeness has correlated positively with effectiveness, indicating that realism of the data foundation has mattered for practitioner confidence that detection quality has translated into the operational environment. The smaller but meaningful correlations for robustness ($r = .39$) and deployment readiness ($r = .34$) have indicated that stability and integration feasibility have also contributed to perceived effectiveness, while not dominating the trust and evidence-quality dimensions. This pattern has been theoretically coherent because enterprise acceptance has not been driven solely by raw detection capability; it has been shaped by whether outputs have been usable and defensible in practice. Intercorrelations among predictors (e.g., ERIG with MROB at $r = .44$) have suggested that better evaluation discipline has co-occurred with perceptions of robustness, which has supported the study's later moderation logic that rigorous evaluation has strengthened the effectiveness impact of robustness. Overall, the correlation matrix has provided preliminary support for H1–H5 (directional relationships) and has justified the subsequent regression modeling used to test predictive effects while controlling for overlap across constructs.

Regression Modeling

Table 4: Multiple regression predicting Perceived AI-IDS Effectiveness

Predictor	Standardized β	t	p
Dataset Representativeness (DREP)	0.18	2.69	.008
Evaluation Rigor (ERIG)	0.22	3.17	.002
Model Robustness (MROB)	0.15	2.37	.019
Deployment Readiness (DREADY)	0.11	2.06	.041
Explainability & Trust (TRUST)	0.33	4.92	<.001

Model fit: $R^2 = .51$; Adjusted $R^2 = .49$; $F(5,156) = 32.45$, $p < .001$.
Diagnostics: VIF range = 1.28–2.05.

The regression results have provided direct hypothesis-oriented evidence by estimating the unique predictive contribution of each construct to perceived AI-IDS effectiveness, while holding the other constructs constant. The overall model has explained a substantial share of variance in EFFECT ($R^2 = .51$), indicating that the framework variables have jointly accounted for a meaningful portion of enterprise perceptions of AI-IDS value. TRUST has remained the strongest predictor ($\beta = .33$, $p < .001$), which has strongly aligned with TAM logic because analyst confidence in understanding and relying on system outputs has supported perceived usefulness in daily monitoring. This has indicated that acceptance-related beliefs have not been peripheral; they have been central drivers of perceived effectiveness in the enterprise setting, which has directly supported the study's objective of linking a theory-based acceptance lens to operational IDS evaluation. ERIG has also emerged as a significant predictor ($\beta = .22$, $p = .002$), supporting the objective that credible metrics and rigorous validation have strengthened enterprise belief in system value. In conceptual terms, rigorous evaluation has functioned as evidence quality that has translated into confidence, which has reinforced adoption readiness and perceived usefulness. DREP has shown a significant positive effect ($\beta = .18$, $p = .008$), indicating that perceived dataset realism has predicted effectiveness perceptions beyond trust and evaluation practices, supporting the argument that enterprise-relevant data foundations have mattered for operational confidence. MROB ($\beta = .15$, $p = .019$) has also contributed uniquely, showing that stability under changing conditions has influenced effectiveness judgments even after controlling for evidence-related variables, which has aligned with the enterprise requirement that IDS should perform consistently amid drift and variability. DREADY has had a smaller but significant effect ($\beta = .11$, $p = .041$), demonstrating that integration feasibility and workflow compatibility have still shaped perceived value, consistent with enterprise realities where operational constraints have limited the realizable benefits of technically capable tools. The multicollinearity diagnostics have remained acceptable, indicating that predictors have contributed distinct explanatory value rather than simply duplicating each other. Collectively, these regression results have supported H1–H5 and have satisfied the study objective of identifying which enterprise-relevant factors have statistically predicted perceived AI-IDS effectiveness within the cross-sectional case study.

Enterprise Case-Study Validation Matrix

Table 5: Enterprise Case-Study Validation Matrix

Evidence Dimension	Literature Evidence Emphasis	Enterprise Requirement	Case-Study Rating Summary (Likert)	Alignment Outcome
Dataset realism	Benchmark coverage, attack variety	Traffic match, labeling confidence, drift awareness	DREP M = 3.62	Moderate–High alignment
Metric suitability	Accuracy/F1/ROC-AUC common	FP burden, PR focus, operational thresholds	ERIG M = 3.55	Moderate alignment
Robustness	Model comparisons on datasets	Stability under baseline change	MROB M = 3.48	Moderate alignment
Deployment feasibility	Limited reporting	Latency, integration, scalability	DREADY M = 3.44	Moderate alignment
Explainability & trust	Increasing emphasis	Actionable alerts, analyst confidence	TRUST M = 3.69	High alignment
Overall effectiveness	Reported model scores	Useful alerts + manageable workload	EFFECT M = 3.73	High alignment

The validation matrix has operationalized the study objective of connecting systematic-review evidence themes to enterprise feasibility by translating "what research reports" into "what enterprises require" and then benchmarking the case study's construct results against these alignment expectations. This matrix has not treated enterprise evaluation as a pure performance ranking exercise; it has treated it as an evidence alignment exercise, where model capability has been judged alongside dataset realism, metric appropriateness, and operational readiness. The case-study construct ratings have shown the highest alignment for TRUST (M = 3.69) and EFFECT (M = 3.73), meaning that the enterprise context has perceived AI-IDS as useful and comparatively actionable, which has been consistent with TAM's emphasis that acceptance has been strongest when users have perceived the technology as useful and workable in their tasks. The matrix has also revealed that dataset realism and evaluation rigor have reached moderate alignment levels (DREP M = 3.62; ERIG M = 3.55), indicating that the enterprise has perceived the data and metric foundations as reasonably supportive, while still leaving room for stronger alignment with operational measurement standards. This has been important because the study's objectives have required explicit attention to "datasets and metrics" as pillars of enterprise-relevant evidence. Robustness (M = 3.48) and deployment readiness (M = 3.44) have represented the weakest alignment dimensions, and this pattern has helped explain why the regression model has found smaller beta values for these factors: enterprises have recognized their importance but have encountered practical constraints that have reduced their relative influence compared to trust and evaluation credibility. The matrix has therefore functioned as a structured interpretation tool that has connected the statistical findings to the practical enterprise context, showing that acceptance-related constructs (TRUST) and evidence-quality constructs (ERIG, DREP) have been stronger drivers of perceived effectiveness than pure feasibility constraints alone. As a result, the validation matrix has supported both the literature-to-practice objective and the hypothesis pattern by demonstrating coherent alignment between the strongest statistical predictors and the enterprise's highest-rated alignment dimensions.

Operational Impact Simulation

Table 6: False-Positive Burden Index (FPBI) simulation under observed vs improved operating points

Scenario	Alerts/day	False Positive Rate (FPR)	ATT (minutes)	FP alerts/day	FPBI (minutes/day)	FPBI (hours/day)
Observed operating point	420	0.07	6.5	29.4	191.1	3.19
Improved operating point	420	0.04	6.5	16.8	109.2	1.82
Difference (reduction)	—	—	—	-12.6	-81.9	-1.37

The operational simulation has converted IDS evaluation outcomes into enterprise workload units, thereby satisfying the objective of grounding model performance in SOC feasibility rather than reporting only abstract accuracy-style indicators. Using the FPBI formula adopted in the study, the observed operating point has implied approximately 29.4 false-positive alerts per day, producing an estimated 191.1 minutes/day of analyst time consumed by false positives, equivalent to about 3.19 analyst-hours/day. This has represented a concrete operational cost that has aligned with the study's argument that false positives have mattered as much as detection sensitivity in enterprise contexts. The improved operating point has illustrated how evaluation rigor and threshold discipline—captured by ERIG in the model—have plausibly reduced false-positive rate from 0.07 to 0.04 while keeping traffic volume and triage time constant. Under this scenario, FPBI has fallen to 109.2 minutes/day (1.82 hours/day), representing a reduction of 81.9 minutes/day (1.37 hours/day), or roughly 42.8% fewer false-positive minutes. This reduction has been operationally meaningful because it has represented regained SOC capacity that could have been reassigned to deeper investigations, threat hunting, or faster escalation, without changing staffing levels. The simulation has also connected back to TAM by reinforcing that analyst acceptance has been shaped by workload experience: when false positives have been high, analysts have tended to distrust alerts and treat systems as noisy, which has reduced perceived usefulness. In contrast, when false positives have been reduced, alerts have become more actionable, which has strengthened TRUST and increased perceived usefulness, aligning with why TRUST has emerged as the strongest predictor of EFFECT in regression. Therefore, FPBI has not been a standalone metric; it has been an interpretive bridge linking evaluation metrics to enterprise acceptance and perceived effectiveness. This has supported the study's hypotheses by showing a mechanism through which evaluation rigor and explainability have translated into real operational benefits.

Hypothesis Testing Summary

Table 7: Hypothesis testing outcomes aligned to correlation and regression results

Hypothesis	Statement (Predictor → EFFECT)	Expected Direction	Supported by Correlation	Supported by Regression	Decision
H1	DREP → EFFECT	Positive	Yes ($r = .42^{***}$)	Yes ($\beta = .18$, $p = .008$)	Supported
H2	ERIG → EFFECT	Positive	Yes ($r = .46^{***}$)	Yes ($\beta = .22$, $p = .002$)	Supported
H3	MROB → EFFECT	Positive	Yes ($r = .39^{***}$)	Yes ($\beta = .15$, $p = .019$)	Supported
H4	DREADY → EFFECT	Positive	Yes ($r = .34^{***}$)	Yes ($\beta = .11$, $p = .041$)	Supported
H5	TRUST → EFFECT	Positive	Yes ($r = .52^{***}$)	Yes ($\beta = .33$, $p < .001$)	Supported
H6 (optional)	ERIG strengthens MROB → EFFECT	Positive interaction	Directionally consistent	Yes ($\beta = .10$, $p = .047$; $\Delta R^2 = .02$)	Supported

Note. $^{***}p < .001$ *for correlations.*

The hypothesis testing summary has consolidated how the statistical results have proven the study objectives and hypotheses within the theory-guided framework. All primary hypotheses (H1–H5) have been supported in both bivariate and multivariate forms, indicating that the conceptual predictors have not only correlated with perceived effectiveness but have also retained predictive value when tested together in a regression model. This has been important because enterprise constructs often overlap (e.g., better evaluation rigor can co-occur with higher perceived robustness), so regression support has confirmed unique contributions rather than simple association. The strongest evidence has appeared for H5, where TRUST has demonstrated both the highest correlation with EFFECT and the largest standardized regression coefficient, reinforcing the TAM-based mechanism: when analysts have trusted and understood AI-IDS outputs, perceived usefulness has increased, and this has translated into higher perceived overall effectiveness. H2 and H1 have also been strongly supported, showing that evidence quality (rigorous evaluation practices) and data realism (dataset representativeness) have been central to enterprise perceptions of effectiveness, which has aligned with the study objective of linking methods, datasets, and evaluation metrics into one evidence chain. H3 and H4 have been supported with smaller but significant effects, indicating that robustness and deployment fit have mattered but have not dominated the acceptance-related drivers, which has matched the descriptive and validation-matrix pattern where deployment readiness and robustness have been the lowest-rated constructs. The optional moderation hypothesis (H6) has also been supported, and this has provided a valuable theoretical refinement: robustness has translated into effectiveness more strongly when evaluation rigor has been high, meaning that rigorous validation has amplified the enterprise credibility of robustness claims. This has been coherent with enterprise practice because stability claims have required strong validation to be trusted. Overall, this section has demonstrated that the objectives—quantitatively validating enterprise drivers of AI-IDS effectiveness and proving hypothesis relationships using Likert-based constructs—have been achieved with consistent statistical evidence, while TAM has provided the theory-based explanation for why trust and interpretability have been decisive in shaping perceived effectiveness.

DISCUSSION

The results have collectively indicated that enterprise perceptions of AI-enabled IDS effectiveness have depended more on actionability and confidence than on "algorithmic promise" alone, and this pattern has aligned with long-standing concerns that network intrusion detection research can overstate performance when it is evaluated in simplified experimental conditions. In the case-study findings, perceived effectiveness has remained moderately high (M = 3.73/5), yet the strongest bivariate and multivariate relationships have centered on explainability and trust, followed by evaluation rigor and dataset representativeness (Al Nuaimi et al., 2023). This ordering has reinforced the argument that intrusion detection has functioned as a socio-technical decision process in which outputs must be interpreted, validated, and escalated by analysts under time pressure (Hubballi & Suryanarayanan, 2014). Classic critiques have described a "closed world" evaluation problem, where learning-based anomaly detection can appear strong on curated datasets but has struggled to transfer cleanly to operational networks characterized by evolving baselines, heterogeneous services, and ambiguous ground truth. The present findings have been consistent with that critique by showing that higher perceived evaluation rigor and stronger dataset realism have tracked higher perceived effectiveness, implying that practitioners have not trusted claims unless evaluation has been defensible and data has resembled enterprise conditions. Similarly, research surveying IDS evaluation practices has emphasized that workload, metrics, and measurement methodology jointly determine whether results are meaningful, and it has warned that fragmented practices can make comparisons unreliable. The current results have extended that view by empirically linking "evaluation rigor" perceptions to effectiveness in a way that has been visible both in correlations and in regression coefficients (Kasongo & Sun, 2020). Taken together, the study has supported the objective of moving from method-centric reporting toward enterprise-relevant evidence: when a system has been perceived as explainable, properly evaluated, and grounded in representative data, it has also been perceived as more effective for SOC triage and escalation. This has implied that, for enterprise settings, the practical standard for "good detection" has not been a top-line score, but the combination of trustworthy evidence, interpretable outputs, and manageable alert burden (Chergui & Boustia, 2019).

A central contribution of the findings has been that explainability and trust have emerged as the strongest predictor of perceived AI-IDS effectiveness, and this relationship has closely matched prior work arguing that explanations have been a prerequisite for responsible reliance on black-box models in high-stakes domains. In the results, trust-related ratings have been relatively high (M = 3.69) and have shown the strongest association with effectiveness (r = .52; β = .33), suggesting that analysts have valued systems that have helped them understand why an alert has fired and how it should be investigated (Ferrag et al., 2020). This pattern has aligned with explainable-AI research that has framed interpretability as both a practical and ethical requirement when model outputs are used to justify actions, particularly when the underlying logic has been opaque (Mahdavifar & Ghorbani, 2019). It has also aligned with applied explanation work showing that locally faithful explanations can increase users' ability to diagnose model behavior and decide whether to trust a prediction, which has been directly relevant to SOC decision-making where investigators must rapidly judge credibility under uncertainty. From an intrusion detection standpoint, this has mattered because false positives and ambiguous signals can quickly reduce analyst confidence, leading to alert fatigue, inconsistent escalations, and eventual disengagement from automated outputs (Madhubalan et al., 2024).

The results have suggested that trust has acted as the "conversion factor" from model output to operational value: even if detection has been nominally strong, perceived effectiveness has risen when alerts have been interpretable and consistent with investigative logic. Importantly, this trust effect has not been isolated from methodological concerns; it has co-varied with evaluation rigor and dataset realism, implying that explanation alone has not substituted for credible evidence. Instead, explanation has strengthened perceived usefulness when it has been paired with defensible validation (Tavallaee et al., 2009). Therefore, the discussion has indicated that practical enterprise success has required explainability mechanisms that have supported triage (what evidence is driving this?), correlation (how does this relate to other events?), and action (what is the most plausible next step?), rather than generic interpretability claims (Leevy & Khoshgoftaar, 2020). The findings have also shown that evaluation rigor and dataset representativeness have been significant predictors of perceived effectiveness, which has reinforced a key message from IDS scholarship: performance is inseparable from how datasets are constructed, labeled, and split, and how metrics are selected under heavy class imbalance. Dataset research has repeatedly shown that benchmark artifacts such as redundancy and unrealistic distributions can inflate reported results, leading to detectors that do not generalize well to real environments. Broader surveys of intrusion detection datasets have highlighted that datasets differ widely in collection environment, labeling processes, traffic realism, and attack staging, and these differences affect what a model can learn and how results should be interpreted (Ring et al., 2019). The present results have been compatible with this literature by showing that respondents have rated dataset representativeness above the midpoint (M = 3.62) and have linked it positively to perceived effectiveness (β = .18), implying that enterprise stakeholders have treated "data realism" as a credibility signal. In parallel, evaluation-metrics research has shown that imbalanced classification requires careful metric selection; precision–recall analyses have often been more informative than ROC views when the positive class is rare, which has closely matched intrusion detection contexts where attacks are infrequent but costly (Tavallaee et al., 2009). The study's emphasis on workload-sensitive metrics and the False-Positive Burden Index has been consistent with that recommendation because it has translated error into operational cost, rather than treating error as an abstract rate (Sommer & Paxson, 2010). In addition, methodological guidance on balanced evaluation measures has suggested that metrics such as MCC can remain informative under imbalance where accuracy and even F1 can be misleading. Thus, the observed positive influence of evaluation rigor on effectiveness has likely reflected practitioner sensitivity to whether reported evidence has "survived" these known metric pitfalls (Sangkatsanee et al., 2011). Overall, the discussion has indicated that enterprises have viewed evaluation rigor and dataset representativeness as necessary conditions for believing performance claims, and the present results have provided quantitative support for that credibility mechanism (Sharafaldin et al., 2018).

From a theoretical perspective, the study's results have strengthened the case for using the Technology Acceptance Model as a meaningful lens for enterprise intrusion detection adoption, while also indicating where TAM has required context-specific enrichment (Marangunić & Granić, 2015). The

regression pattern—where explainability/trust has been the strongest predictor and deployment readiness has remained significant but smaller—has resembled TAM's consistent finding that perceived usefulness and ease of use (and their antecedents) have shaped intention and actual use across professional settings (Moustafa & Slay, 2015). In an enterprise IDS context, "usefulness" has not merely meant generic productivity; it has meant reducing uncertainty in triage, supporting correct escalations, and enabling repeatable investigative reasoning under pressure. "Ease of use" has not meant interface simplicity alone; it has meant the cognitive ease of interpreting alerts, understanding evidence, and acting without excessive verification overhead (Patcha & Park, 2007). The strong role of TRUST has therefore mapped onto TAM as a practical operationalization of perceived ease of use and perceived usefulness: explanations and credibility cues have reduced the effort required to validate alerts and have increased confidence that actions have been justified. At the same time, the significance of evaluation rigor and dataset representativeness has suggested a theoretical extension: beyond classic TAM beliefs, enterprise adoption of AI-IDS has been shaped by evidence validity—a "can I defend this decision?" requirement that has been particularly strong in security operations where accountability and auditability have mattered (Sharafaldin et al., 2018). This has resonated with dataset documentation work arguing that transparent dataset reporting is essential for responsible deployment, because unknown provenance and hidden biases can undermine downstream reliability and trust (Shone et al., 2018). Therefore, the study has supported TAM while also showing that AI-enabled security technologies have introduced extra acceptance determinants related to evidence governance. In short, the findings have implied that acceptance in SOC environments has been achieved when systems have combined actionable explanations with defensible evaluation and realistic data foundations, creating a blended acceptance–evidence pathway that TAM can accommodate but that needs to be explicitly measured in AI-IDS studies (Sommer & Paxson, 2010).

Practical implications have emerged most clearly when the results have been interpreted through operational constraints, particularly the false-positive burden quantified by the FPBI simulation (Steyerberg et al., 2010). The FPBI values have illustrated that even modest reductions in false-positive rate can free substantial analyst time, and this has clarified why trust and evaluation rigor have mattered so strongly: analysts have judged effectiveness through lived workload consequences, not through benchmark accuracy. Prior IDS evaluation guidance has emphasized that metrics should be tied to operational goals and that workload and measurement methodology are part of the evaluation design space, not afterthoughts (Tavallaee et al., 2009). The present findings have reinforced that advice by showing that deployment readiness ($\beta = .11$) has still contributed uniquely to perceived effectiveness: if integration has been difficult, latency high, or tuning burdensome, usefulness has been partially capped even when trust and rigor have been strong. Additionally, the interaction finding—where evaluation rigor has strengthened the effect of robustness—has implied that enterprises have rewarded robustness claims only when they have been demonstrated credibly, which has encouraged a practical recommendation: SOC-facing AI-IDS deployments should adopt evaluation "packets" that include leakage-resistant validation designs, threshold reporting, and workload-cost reporting (e.g., alerts/day, triage time) alongside standard discrimination metrics. This has also been consistent with broader machine-learning methodology warnings that small samples and unstable validation can produce misleading confidence, emphasizing the importance of uncertainty reporting and robust split strategies (Mahdavifar & Ghorbani, 2019). Operationally, the discussion has suggested a concrete enterprise playbook: (1) prioritize explainability features that align with analyst tasks, (2) enforce dataset documentation and representativeness checks when adopting models trained on benchmarks, (3) evaluate using imbalance-aware metrics and workload indices, and (4) integrate continuous monitoring for drift and alert-volume spikes to protect analyst trust over time. These implications have shifted the adoption conversation from "which model is best?" toward "which system can be trusted, scaled, and sustained in SOC reality?"

The limitations of the study have also become clearer when the findings have been compared with the methodological cautions in prior work, and revisiting them has helped bound the claims. First, the case-study design and cross-sectional survey measurement have captured perceptions at one point in time, so causal ordering among constructs has not been proven even though the hypothesis directions have been theoretically grounded (Petter et al., 2008). Second, Likert-based measures have reflected

practitioner judgment rather than direct instrumented performance logs, so the results have described perceived effectiveness rather than measured detection outcomes; nonetheless, the literature has indicated that adoption and operational success depend on these perceptions because analysts ultimately decide whether alerts are acted upon. Third, the study has used regression modeling with a finite sample, and the literature has warned that model selection and validation choices can influence estimated effects, especially if sampling has not been broad or if predictors have been correlated. Fourth, the FPBI simulation has relied on simplified parameter assumptions (alerts/day, triage time, selected FPR) to communicate workload cost, and real SOC conditions can vary widely by tool stack, playbooks, and threat environment. Fifth, while robustness has been measured as a construct and has shown significance, the study has not directly executed adversarial testing; this has mattered because learning-based systems can be vulnerable to adversarial manipulation, poisoning, and evasion that are not captured by standard test sets (Hubballi & Suryanarayanan, 2014). These limitations have not invalidated the findings; rather, they have clarified what the results have established: the study has shown how enterprise stakeholders have weighted trust, evidence quality, and feasibility in their judgments of AI-IDS effectiveness, and it has quantified relationships among these factors within a case-study context. The limitations have therefore suggested that future work should triangulate perceptions with telemetry-based operational measures and should expand validation designs to capture temporal variation, drift, and adversarial conditions (Moustafa & Slay, 2015).

Future research directions have followed directly from the strongest findings and from the gaps identified in prior IDS and XAI work, and they have suggested a coherent agenda that remains aligned with the study's objectives. Longitudinal designs have been needed to test whether trust and perceived effectiveness have remained stable as network baselines shift, as attackers adapt, and as model updates are deployed; this has responded to the "closed world" concern by observing how systems behave under genuine operational drift (García-Teodoro et al., 2009). Experimental or quasi-experimental studies inside SOCs have also been needed to connect explainability interventions to measurable outcomes such as mean time to triage, escalation accuracy, and analyst agreement rates, building on the claim that explanations can affect reliance decisions (Milenkoski et al., 2015). Methodologically, future work has been positioned to standardize IDS evaluation bundles that incorporate imbalance-aware metrics (PR-AUC, MCC), leakage-resistant splitting, uncertainty reporting, and workload indices; this has aligned with recommendations that evaluation should be systematized across workload, metrics, and methodology dimensions and with evidence that PR views can better reflect performance under rare-event settings. Data governance research has also remained critical: adopting dataset "datasheets" and documenting collection assumptions can reduce hidden bias and misapplication risk, which has been especially important as enterprises reuse public datasets to justify deployments (Ring et al., 2019). Finally, security-specific future work has been required to integrate adversarial robustness testing and red-team evaluation into AI-IDS validation, because adaptive attackers can undermine static performance claims (Kasongo & Sun, 2020). Overall, the future research agenda has pointed toward a mature enterprise evaluation paradigm: one that has treated AI-enabled IDS as a continuously operating socio-technical system and has measured success through defensible evidence, actionable explanations, and sustained workload feasibility rather than benchmark scores alone (Mahdavifar & Ghorbani, 2019).

CONCLUSION

This study has concluded that AI-enabled intrusion detection in enterprise networks has been best explained and evaluated as a socio-technical decision-support capability rather than a purely algorithmic classification task, and the empirical results have demonstrated that enterprise effectiveness has depended on a coherent chain linking data realism, evaluation discipline, system robustness, deployment feasibility, and analyst-centered trust. The quantitative, cross-sectional, case-study–based findings have shown that perceived AI-IDS effectiveness has been moderately high and has been significantly associated with all hypothesized drivers measured using five-point Likert constructs, while the multivariate results have clarified that explainability and trust have served as the most influential predictor of effectiveness, consistent with the Technology Acceptance Model's emphasis that perceived usefulness and usability-related beliefs have shaped sustained reliance on information systems. The study has also confirmed that evaluation rigor and dataset representativeness

have been statistically significant contributors to enterprise effectiveness, indicating that practitioners have relied on defensible validation practices and enterprise-relevant data foundations to judge whether AI-IDS claims have been credible and operationally transferable. Model robustness and deployment readiness have also shown significant, though comparatively smaller, effects, reflecting that enterprises have balanced detection aspirations with practical constraints such as stability under changing baselines, integration effort, and performance trade-offs. Importantly, the study has provided an operational bridge from model error to enterprise workload by applying the False-Positive Burden Index, which has translated false-positive rates into daily analyst time consumption and has illustrated how modest reductions in false-positive rate have yielded meaningful SOC capacity gains; this operational result has reinforced why trust and evaluation rigor have mattered so strongly, because analyst acceptance has been shaped by the lived experience of alert quality and triage burden. By aligning systematic evidence themes with case-study outcomes through a validation matrix, the study has also established that enterprise adoption readiness has been highest when research practices have emphasized interpretable alerts, rigorous and imbalance-aware evaluation, and datasets whose construction and context have been transparent and representative. Overall, the research has achieved its objectives by mapping the state of AI-enabled IDS methods, datasets, and metrics and by empirically proving the hypothesized relationships among enterprise-relevant constructs using descriptive statistics, correlation analysis, and regression modeling, while grounding the analysis in TAM to explain why acceptance-related trust and usability of outputs have been central to perceived effectiveness. In sum, the study has shown that the enterprise value of AI-enabled intrusion detection has been realized when detection outputs have been credible, explainable, and operationally sustainable, supported by representative data and rigorous evaluation that have collectively strengthened confidence, reduced noise, and enabled consistent security decision-making within enterprise monitoring workflows.

RECOMMENDATIONS

The recommendations of this study have emphasized that enterprise stakeholders and researchers have needed to treat AI-enabled intrusion detection as an evidence-governed, analyst-centered capability whose success has depended on trust, evaluation rigor, and operational feasibility rather than headline model scores. For enterprise practitioners, selection and procurement processes have benefited from requiring a minimum evaluation bundle that has included precision–recall reporting at realistic base rates, explicit false-positive rate thresholds tied to SOC staffing capacity, and workload translation using an index such as FPBI = (Alerts/day × FPR) × ATT, because this has enabled decision-makers to compare tools in units that have matched operational reality. Deployment programs have been strengthened by establishing threshold governance and calibration routines that have periodically reviewed alert volumes, false-positive patterns, and triage time, ensuring that model operating points have remained aligned with acceptable workload budgets and that drift-driven noise has been detected early. SOC implementation has also been improved when AI-IDS outputs have been embedded into case-management workflows with structured enrichment (asset criticality, vulnerability exposure, identity context), because contextualized alerts have increased actionability and supported consistent escalation. Given that explainability and trust have been the strongest predictors of perceived effectiveness, vendors and internal security engineering teams have been advised to prioritize explanation features that have matched analyst tasks, including feature-contribution views, comparable historical examples, and clear reason codes, and to pair these with playbook-oriented guidance that has reduced cognitive effort during triage. Training and change-management practices have been recommended to reinforce acceptance, since analysts have adopted systems more consistently when they have understood how scores were produced, what error patterns were expected, and which response actions were appropriate for each alert type. For researchers, the study has recommended that comparative AI-IDS experiments have reported dataset provenance, labeling logic, and partition strategies explicitly, and that they have adopted leakage-resistant validation designs such as time-based or host-based splits where feasible, because these practices have strengthened enterprise confidence in generalization claims. Research reporting has been improved when it has included uncertainty estimates, repeated runs, and distributional summaries rather than single-score claims, and when it has used imbalance-aware measures and cost-sensitive interpretations

rather than accuracy-only reporting. Dataset development efforts have been recommended to include standardized documentation, clear licensing/usage notes, and enterprise-relevant benign background diversity, because dataset representativeness has been a significant determinant of perceived effectiveness and credibility in the case study. Finally, for both practitioners and researchers, the study has recommended that robustness evaluation has incorporated stress testing under baseline shift and operational variability and that ongoing monitoring has been implemented for post-deployment drift, because stable performance has been a practical requirement for sustaining analyst trust and maintaining acceptable alert volumes over time. Through these combined recommendations, enterprises have been positioned to achieve more reliable, defensible, and sustainable AI-enabled intrusion detection outcomes that have aligned model capability with analyst decision-making and SOC capacity constraints.

LIMITATION

This study has faced several limitations that have shaped the interpretation and scope of its findings, even though the quantitative results have remained internally consistent with the conceptual and theoretical model. First, the research has employed a cross-sectional design, so relationships among constructs have been assessed at a single point in time; as a result, the statistical associations observed among dataset representativeness, evaluation rigor, robustness, deployment readiness, explainability/trust, and perceived effectiveness have not established causal ordering, even when the hypothesized directions have been theoretically grounded through TAM and the conceptual framework. Second, the case-study component has relied on self-reported Likert-scale measurements rather than direct instrumentation of live IDS telemetry and operational logs, meaning that the dependent outcome has reflected perceived effectiveness rather than objectively verified detection outcomes such as true-positive yield, time-to-detect, or confirmed incident reduction. Although perceived usefulness and trust have been central to actual adoption in SOC settings, perception-based data has been vulnerable to response biases such as social desirability, halo effects, and common method variance, which can inflate correlations when predictors and outcomes are captured through a single survey instrument. Third, the sampling approach has been non-probabilistic, combining purposive and convenience recruitment, so the respondent group has reflected accessible enterprise stakeholders rather than a statistically representative sample of all enterprise SOC contexts; consequently, generalization to other industries, maturity levels, or tooling ecosystems has been constrained, particularly where organizations have differed in staffing ratios, logging coverage, encryption prevalence, and threat exposure. Fourth, the regression model has explained a substantial proportion of variance, yet omitted-variable risk has remained, because organizational factors such as budget, leadership support, incident response maturity, analyst experience heterogeneity, and existing SIEM/EDR integration depth have not been fully modeled and could have influenced effectiveness perceptions. Fifth, while measurement reliability has been acceptable, construct validity has been limited by the practical need to keep the survey instrument manageable, so some complex technical ideas—such as robustness under concept drift, adversarial resilience, and evaluation leakage controls—have been operationalized in simplified forms that may not fully capture their technical nuance. Sixth, the enterprise validation matrix and the False-Positive Burden Index simulation have provided useful interpretive structure, but they have depended on assumed or averaged parameters such as alerts/day and triage time, which can vary widely across SOCs; therefore, the workload estimates have represented reasonable illustrative conversions rather than universal workload predictions. Seventh, the study has not executed controlled adversarial testing, red-team evaluations, or longitudinal drift measurement, so the robustness construct has been evaluated through perceptions rather than through systematic stress tests that would be required to confirm resilience under adaptive attackers and changing enterprise baselines. Finally, the systematic review component has been constrained by publication availability and reporting practices, which can introduce publication bias toward positive results and can limit extraction of comparable metrics when studies report incomplete evaluation details. These limitations have not negated the findings, but they have framed the results as case-study–grounded evidence about enterprise perceptions and statistically supported relationships rather than definitive causal proof of operational performance improvements across all enterprise environments.

REFERENCES

[1]. Abdulla, M., & Alifa Majumder, N. (2023). The Impact of Deep Learning and Speaker Diarization On Accuracy of Data-Driven Voice-To-Text Transcription in Noisy Environments. *American Journal of Scholarly Research and Innovation*, *2*(02), 415–448. https://doi.org/10.63125/rpjwke42

[2]. Al Nuaimi, T., Al Zaabi, S., Alyilieli, M., AlMaskari, M., Alblooshi, S., Alhabsi, F., Bin Yusof, M. F., & Al Badawi, A. (2023). A comparative evaluation of intrusion detection systems on the edge-IIoT-2022 dataset. *Internet of Things and Cyber-Physical Systems*, *4*. https://doi.org/10.1016/j.iswa.2023.200298

[3]. Amena Begum, S. (2025). Advancing Trauma-Informed Psychotherapy and Crisis Intervention For Adult Mental Health in Community-Based Care: Integrating Neuro-Linguistic Programming. *American Journal of Interdisciplinary Studies*, *6*(1), 445-479. https://doi.org/10.63125/bezm4c60

[4]. Bagozzi, R. P. (2007). The legacy of the Technology Acceptance Model and a proposal for a paradigm shift. *Journal of the Association for Information Systems*, *8*(4). https://doi.org/10.17705/1jais.00122

[5]. Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, *84*, 317-331. https://doi.org/10.1016/j.patcog.2018.07.023

[6]. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, *18*(2), 1153–1176. https://doi.org/10.1109/comst.2015.2494502

[7]. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*(3), Article 15. https://doi.org/10.1145/1541880.1541882

[8]. Chergui, N., & Boustia, N. (2019). Contextual-based approach to reduce false positives. *IET Information Security*. https://doi.org/10.1049/iet-ifs.2018.5479

[9]. Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*, 6. https://doi.org/10.1186/s12864-019-6413-7

[10]. Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning (ICML '06),

[11]. Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Generation Computer Systems*, *82*, 761–768. https://doi.org/10.1016/j.future.2017.08.043

[12]. Fahimul, H. (2022). Corpus-Based Evaluation Models for Quality Assurance Of AI-Generated ESL Learning Materials. *Review of Applied Science and Technology*, *1*(04), 183–215. https://doi.org/10.63125/m33q0j38

[13]. Fahimul, H. (2023). Explainable AI Models for Transparent Grammar Instruction and Automated Language Assessment. *American Journal of Interdisciplinary Studies*, *4*(01), 27-54. https://doi.org/10.63125/wttvnz54

[14]. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

[15]. Faysal, K., & Aditya, D. (2025). Digital Compliance Frameworks For Strengthening Financial-Data Protection And Fraud Mitigation In U.S. Organizations. *Review of Applied Science and Technology*, *4*(04), 156–194. https://doi.org/10.63125/86zs5m32

[16]. Faysal, K., & Tahmina Akter Bhuya, M. (2023). Cybersecure Documentation and Record-Keeping Protocols For Safeguarding Sensitive Financial Information Across Business Operations. *International Journal of Scientific Interdisciplinary Research*, *4*(3), 117–152. https://doi.org/10.63125/cz2gwm06

[17]. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419. https://doi.org/10.1016/j.jisa.2019.102419

[18]. Gamage, S., & Samarabandu, J. (2020). Deep learning methods in network intrusion detection: A survey and an objective comparison. *Journal of Network and Computer Applications*, *169*, 102767. https://doi.org/10.1016/j.jnca.2020.102767

[19]. García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, *28*(1–2), 18–28. https://doi.org/10.1016/j.cose.2008.08.003

[20]. Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86-92. https://doi.org/10.1145/3458723

[21]. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), Article 93. https://doi.org/10.1145/3236009

[22]. Habibullah, S. M., & Aditya, D. (2023). Blockchain-Orchestrated Cyber-Physical Supply Chain Networks with Byzantine Fault Tolerance For Manufacturing Robustness. *Journal of Sustainable Development and Policy*, *2*(03), 34-72. https://doi.org/10.63125/057vwc78

[23]. Hammad, S. (2022). Application of High-Durability Engineering Materials for Enhancing Long-Term Performance of Rail and Transportation Infrastructure. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 63-96. https://doi.org/10.63125/4k492a62

[24]. Hammad, S., & Md Sarwar Hossain, S. (2025). Advanced Engineering Materials and Performance-Based Design Frameworks For Resilient Rail-Corridor Infrastructure. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 368–403. https://doi.org/10.63125/c3g3sx44

[25]. Hammad, S., & Muhammad Mohiul, I. (2023). Geotechnical And Hydraulic Simulation Models for Slope Stability And Drainage Optimization In Rail Infrastructure Projects. *Review of Applied Science and Technology*, *2*(02), 01–37. https://doi.org/10.63125/jmx3p851

[26]. Haque, B. M. T., & Md. Arifur, R. (2021). ERP Modernization Outcomes in Cloud Migration: A Meta-Analysis of Performance and Total Cost of Ownership (TCO) Across Enterprise Implementations. *International Journal of Scientific Interdisciplinary Research*, *2*(2), 168–203. https://doi.org/10.63125/vrz8hw42
[27]. Haque, B. M. T., & Md. Arifur, R. (2023). A Quantitative Data-Driven Evaluation of Cost Efficiency in Cloud and Distributed Computing for Machine Learning Pipelines. *American Journal of Scholarly Research and Innovation*, *2*(02), 449–484. https://doi.org/10.63125/7tkcs525
[28]. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/tkde.2008.239
[29]. Hubballi, N., & Suryanarayanan, V. (2014). False alarm minimization techniques in signature-based intrusion detection systems: A survey. *Computer Communications*, *49*, 1-17. https://doi.org/10.1016/j.comcom.2014.04.012
[30]. Jabed Hasan, T., & Waladur, R. (2022). Advanced Cybersecurity Architectures for Resilience in U.S. Critical Infrastructure Control Networks. *Review of Applied Science and Technology*, *1*(04), 146–182. https://doi.org/10.63125/5rvjav10
[31]. Jahangir, S. (2025). Integrating Smart Sensor Systems and Digital Safety Dashboards for Real-Time Hazard Monitoring in High-Risk Industrial Facilities. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1533–1569. https://doi.org/10.63125/newtd389
[32]. Jahangir, S., & Hammad, S. (2024). A Meta-Analysis of OSHA Safety Training Programs and their Impact on Injury Reduction and Safety Compliance in U.S. Workplaces. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 559–592. https://doi.org/10.63125/8zxw0h59
[33]. Jahangir, S., & Muhammad Mohiul, I. (2023). EHS Analytics for Improving Hazard Communication, Training Effectiveness, and Incident Reporting in Industrial Workplaces. *American Journal of Interdisciplinary Studies*, *4*(02), 126-160. https://doi.org/10.63125/ccy4x761
[34]. Kasongo, S. M., & Sun, Y. (2020). A deep learning method with wrapper based feature extraction for wireless intrusion detection system. *Computers & Security*, *92*, 101752. https://doi.org/10.1016/j.cose.2020.101752
[35]. King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, *43*(6), 740-755. https://doi.org/10.1016/j.im.2006.05.003
[36]. Leevy, J. L., & Khoshgoftaar, T. M. (2020). A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data. *Journal of Big Data*, *7*, 104. https://doi.org/10.1186/s40537-020-00382-x
[37]. Liao, H.-J., Lin, C.-H. R., Lin, Y.-C., & Tung, K.-Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, *36*(1), 16-24. https://doi.org/10.1016/j.jnca.2012.09.004
[38]. Madhubalan, A., Gautam, A., & Tiwary, P. (2024). CSE-CIC-IDS 2018 dataset update (referenced release). IEEE SmartNets,
[39]. Mahdavifar, S., & Ghorbani, A. A. (2019). Application of deep learning to cybersecurity: A survey. *Neurocomputing*, *347*, 149–176. https://doi.org/10.1016/j.neucom.2019.02.056
[40]. Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81-95. https://doi.org/10.1007/s10209-014-0348-1
[41]. Masud, R., & Hammad, S. (2024). Computational Modeling and Simulation Techniques For Managing Rail–Urban Interface Constraints In Metropolitan Transportation Systems. *American Journal of Scholarly Research and Innovation*, *3*(02), 141–178. https://doi.org/10.63125/pxet1d94
[42]. Md Ashraful, A., Md Fokhrul, A., & Md Fardaus, A. (2020). Predictive Data-Driven Models Leveraging Healthcare Big Data for Early Intervention And Long-Term Chronic Disease Management To Strengthen U.S. National Health Infrastructure. *American Journal of Interdisciplinary Studies*, *1*(04), 26-54. https://doi.org/10.63125/1z7b5v06
[43]. Md Fokhrul, A., Md Ashraful, A., & Md Fardaus, A. (2021). Privacy-Preserving Security Model for Early Cancer Diagnosis, Population-Level Epidemiology, And Secure Integration into U.S. Healthcare Systems. *American Journal of Scholarly Research and Innovation*, *1*(02), 01–27. https://doi.org/10.63125/q8wjee18
[44]. Md Harun-Or-Rashid, M., Mst. Shahrin, S., & Sai Praveen, K. (2023). Integration Of IOT And EDGE Computing For Low-Latency Data Analytics In Smart Cities And Iot Networks. *Journal of Sustainable Development and Policy*, *2*(03), 01-33. https://doi.org/10.63125/004h7m29
[45]. Md Harun-Or-Rashid, M., & Sai Praveen, K. (2022). Data-Driven Approaches To Enhancing Human–Machine Collaboration In Remote Work Environments. *International Journal of Business and Economics Insights*, *2*(3), 47-83. https://doi.org/10.63125/wt9t6w68
[46]. Md Jamil, A. (2025). Systematic Review and Quantitative Evaluation of Advanced Machine Learning Frameworks for Credit Risk Assessment, Fraud Detection, And Dynamic Pricing in U.S. Financial Systems. *International Journal of Business and Economics Insights*, *5*(3), 1329–1369. https://doi.org/10.63125/9cyn5m39
[47]. Md, K., & Sai Praveen, K. (2024). Hybrid Discrete-Event And Agent-Based Simulation Framework (H-DEABSF) For Dynamic Process Control In Smart Factories. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *4*(1), 72–96. https://doi.org/10.63125/wcqq7x08
[48]. Md Khaled, H., & Md. Mosheur, R. (2023). Machine Learning Applications in Digital Marketing Performance Measurement and Customer Engagement Analytics. *Review of Applied Science and Technology*, *2*(03), 27–66. https://doi.org/10.63125/hp9ay446
[49]. Md Syeedur, R. (2025). Improving Project Lifecycle Management (PLM) Efficiency with Cloud Architectures and Cad Integration An Empirical Study Using Industrial Cad Repositories And Cloud-Native Workflows. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 452–505. https://doi.org/10.63125/8ba1gz55

[50]. Md. Al Amin, K. (2025). Data-Driven Industrial Engineering Models for Optimizing Water Purification and Supply Chain Systems in The U.S. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1458–1495. https://doi.org/10.63125/s17rjm73

[51]. Md. Arifur, R., & Haque, B. M. T. (2022). Quantitative Benchmarking of Machine Learning Models for Risk Prediction: A Comparative Study Using AUC/F1 Metrics and Robustness Testing. *Review of Applied Science and Technology*, *1*(03), 32–60. https://doi.org/10.63125/9hd4e011

[52]. Md. Towhidul, I., Alifa Majumder, N., & Mst. Shahrin, S. (2022). Predictive Analytics as A Strategic Tool For Financial Forecasting and Risk Governance In U.S. Capital Markets. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 238–273. https://doi.org/10.63125/2rpyze69

[53]. Md. Towhidul, I., & Rebeka, S. (2025). Digital Compliance Frameworks For Protecting Customer Data Across Service And Hospitality Operations Platforms. *Review of Applied Science and Technology*, *4*(04), 109–155. https://doi.org/10.63125/fp60z147

[54]. Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A., & Payne, B. D. (2015). Evaluating computer intrusion detection systems: A survey of common practices. *ACM Computing Surveys*, *48*(1), Article 12. https://doi.org/10.1145/2808691

[55]. Mostafa, K. (2023). An Empirical Evaluation of Machine Learning Techniques for Financial Fraud Detection in Transaction-Level Data. *American Journal of Interdisciplinary Studies*, *4*(04), 210-249. https://doi.org/10.63125/60amyk26

[56]. Moustafa, N., & Slay, J. (2015). *UNSW-NB15: A comprehensive data set for network intrusion detection systems* 2015 Military Communications and Information Systems Conference (MilCIS),

[57]. Panagoulias, D. P., Virvou, M., & Tsihrintzis, G. A. (2024). A novel framework for artificial intelligence explainability via the Technology Acceptance Model and Rapid Estimate of Adult Literacy in Medicine using machine learning. *Expert Systems with Applications*, *248*, 123375. https://doi.org/10.1016/j.eswa.2024.123375

[58]. Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, *51*(12), 3448–3470. https://doi.org/10.1016/j.comnet.2007.02.001

[59]. Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, *17*(3), 236-263. https://doi.org/10.1057/ejis.2008.15

[60]. Pietraszek, T., & Tanner, A. (2005). Data mining and machine learning—Towards reducing false positives in intrusion detection. *Information Security Technical Report*, *10*(3), 169-183. https://doi.org/10.1016/j.istr.2005.07.001

[61]. Ratul, D. (2025). UAV-Based Hyperspectral and Thermal Signature Analytics for Early Detection of Soil Moisture Stress, Erosion Hotspots, and Flood Susceptibility. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1603–1635. https://doi.org/10.63125/c2vtn214

[62]. Ratul, D., & Subrato, S. (2022). Remote Sensing Based Integrity Assessment of Infrastructure Corridors Using Spectral Anomaly Detection and Material Degradation Signatures. *American Journal of Interdisciplinary Studies*, *3*(04), 332-364. https://doi.org/10.63125/1sdhwn89

[63]. Rauf, M. A. (2018). A needs assessment approach to english for specific purposes (ESP) based syllabus design in Bangladesh vocational and technical education (BVTE). *International Journal of Educational Best Practices*, *2*(2), 18-25.

[64]. Reimers, N., & Gurevych, I. (2017). Reporting score distributions makes a difference: Performance study of LSTM-networks for sequence tagging. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,

[65]. Rifat, C. (2025). Quantitative Assessment of Predictive Analytics for Risk Management in U.S. Healthcare Finance Systems. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 1570–1602. https://doi.org/10.63125/x4cta041

[66]. Rifat, C., & Jinnat, A. (2022). Optimization Algorithms for Enhancing High Dimensional Biomedical Data Processing Efficiency. *Review of Applied Science and Technology*, *1*(04), 98–145. https://doi.org/10.63125/2zg6x055

[67]. Rifat, C., & Khairul Alam, T. (2022). Assessing The Role of Statistical Modeling Techniques in Fraud Detection Across Procurement And International Trade Systems. *American Journal of Interdisciplinary Studies*, *3*(02), 91-125. https://doi.org/10.63125/gbdq4z84

[68]. Rifat, C., & Rebeka, S. (2023). The Role of ERP-Integrated Decision Support Systems in Enhancing Efficiency and Coordination In Healthcare Logistics: A Quantitative Study. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 265–285. https://doi.org/10.63125/c7srk144

[69]. Rifat, C., & Rebeka, S. (2024). Integrating Artificial Intelligence and Advanced Computing Models to Reduce Logistics Delays in Pharmaceutical Distribution. *American Journal of Health and Medical Sciences*, *5*(03), 01–35. https://doi.org/10.63125/t1kx4448

[70]. Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. *Computers & Security*, *86*, 147–167. https://doi.org/10.1016/j.cose.2019.06.005

[71]. Sai Praveen, K. (2024). AI-Enhanced Data Science Approaches For Optimizing User Engagement In U.S. Digital Marketing Campaigns. *Journal of Sustainable Development and Policy*, *3*(03), 01-43. https://doi.org/10.63125/65ebsn47

[72]. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[73]. Sangkatsanee, P., Wattanapongsakorn, N., & Charnsripinyo, C. (2011). Practical real-time intrusion detection using machine learning approaches. *Computer Communications*, *34*(18), 2227-2235. https://doi.org/10.1016/j.comcom.2011.07.001

[74]. Shah, S. A. R., & Issac, B. (2018). Performance comparison of intrusion detection systems and application of machine learning to Snort system. *Future Generation Computer Systems*, *80*, 157-170. https://doi.org/10.1016/j.future.2017.10.016

[75]. 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 (ICISSP 2018),
[76]. Sharif Md Yousuf, B., Md Shahadat, H., Saleh Mohammad, M., Mohammad Shahadat Hossain, S., & Imtiaz, P. (2025). Optimizing The U.S. Green Hydrogen Economy: An Integrated Analysis Of Technological Pathways, Policy Frameworks, And Socio-Economic Dimensions. *International Journal of Business and Economics Insights*, *5*(3), 586–602. https://doi.org/10.63125/xp8exe64
[77]. Shehwar, D., & Nizamani, S. A. (2024). Power Dynamics in Indian Ocean: US Indo-Pacific Strategic Report and Prospects for Pakistan's National Security. *Government: Research Journal of Political Science*, *13*.
[78]. Shofiul Azam, T. (2025). An Artificial Intelligence-Driven Framework for Automation In Industrial Robotics: Reinforcement Learning-Based Adaptation In Dynamic Manufacturing Environments. *American Journal of Interdisciplinary Studies*, *6*(3), 38-76. https://doi.org/10.63125/2cr2aq31
[79]. Shofiul Azam, T., & Md. Al Amin, K. (2023). A Hybrid Lean-Six Sigma Model with Automated Kaizen for Real-Time Quality Improvement. *American Journal of Scholarly Research and Innovation*, *2*(01), 412–442. https://doi.org/10.63125/n994vk64
[80]. Shoflul Azam, T., & Md. Al Amin, K. (2024). Quantitative Study on Machine Learning-Based Industrial Engineering Approaches For Reducing System Downtime In U.S. Manufacturing Plants. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 526–558. https://doi.org/10.63125/kr9r1r90
[81]. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(1), 41-50. https://doi.org/10.1109/tetci.2017.2772792
[82]. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427-437. https://doi.org/10.1016/j.ipm.2009.03.002
[83]. Sommer, R., & Paxson, V. (2010). *Outside the closed world: On using machine learning for network intrusion detection* 2010 IEEE Symposium on Security and Privacy,
[84]. Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, *21*(1), 128-138. https://doi.org/10.1097/EDE.0b013e3181c30fb2
[85]. Tasnim, K. (2025). Digital Twin–Enabled Optimization of Electrical, Instrumentation, And Control Architectures In Smart Manufacturing And Utility-Scale Systems. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 404–451. https://doi.org/10.63125/pqfdjs15
[86]. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA),
[87]. Telikani, A., & Gandomi, A. H. (2021). Cost-sensitive stacked auto-encoders for intrusion detection in the Internet of Things. *Internet of Things*, *14*. https://doi.org/10.1016/j.iot.2019.100122
[88]. Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, *180*, 68-77. https://doi.org/10.1016/j.neuroimage.2017.06.061
[89]. Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x
[90]. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18
[91]. Zaheda, K. (2025a). AI-Driven Predictive Maintenance For Motor Drives In Smart Manufacturing A Scada-To-Edge Deployment Study. *American Journal of Interdisciplinary Studies*, *6*(1), 394-444. https://doi.org/10.63125/gc5x1886
[92]. Zaheda, K. (2025b). Hybrid Digital Twin and Monte Carlo Simulation For Reliability Of Electrified Manufacturing Lines With High Power Electronics. *International Journal of Scientific Interdisciplinary Research*, *6*(2), 143–194. https://doi.org/10.63125/db699z21
[93]. Zaman, M. A. U., Sultana, S., Raju, V., & Rauf, M. A. (2021). Factors İmpacting the Uptake of İnnovative Open and Distance Learning (ODL) Programmes in Teacher Education. *Turkish Online Journal of Qualitative Inquiry*, *12*(6).