Intelligent Network Monitoring and Proactive Cyber Threat Detection for Improving Enterprise IT Infrastructure Resilience in Saudi Arabia.

Abstract

Enterprise cyber resilience depends on detecting hostile activity early enough to preserve essential services, yet intrusion detection research frequently evaluates classifiers without examining their operational consequences. This review examines how intelligent network monitoring, contextual threat analysis, and controlled response can be combined within Saudi enterprise information technology environments. A structured integrative synthesis of 30 DOIverified studies published between 2020 and 2025 compares telemetry design, learning approaches, generalisation, adversarial robustness, analyst workload, and recoveryoriented assessment. Evidence indicates that high benchmark accuracy alone does not establish readiness for deployment: dataset leakage, shifting traffic distributions, rare attacks, incomplete asset context, and excessive alerts can each undermine otherwise promising models. The article develops a reference architecture linking network flows, endpoint and identity events, threat intelligence, riskbased correlation, human review, and reversible containment. It also proposes an evaluation protocol using crossenvironment testing, alert investigation effort, detection latency, and service continuity indicators. Saudi application requires explicit attention to asset criticality, data governance, nationally relevant cybersecurity controls, and the differing needs of industrial and office networks. The main contribution is a decision framework that treats detection as one component of measurable operational resilience. Its value lies in defining what organisations should validate before automating response, where further empirical work is needed, and how security analytics can support continuity without implying untested performance gains.

Keywords

Cyber resilience; enterprise network monitoring; intrusion detection; cyber threat intelligence; security operations; Saudi Arabia; operational continuity

UNDER PEER REVIEW IN IJAR

1 Introduction

Enterprise networks are assembled from switches, servers, wireless access points, identity services, cloud connections, remote access gateways, and operational systems. Their traffic reflects legitimate business activity alongside configuration errors, software changes, and malicious attempts to obtain access. Conventional monitoring identifies equipment failures and capacity constraints, whereas security monitoring asks whether observed behaviour indicates exploitation, lateral movement, data misuse, or impending disruption. Combining these perspectives is difficult because a suspicious event has different consequences on an ordinary workstation, a payment system, and a control network. A recent survey of enterprise asset behaviour and distributed attacks demonstrates why static device inspection offers only a partial view of threats that travel through connected assets [1].

Intrusion detection scholarship has advanced rapidly through machine learning, deep learning, graph analysis, and collaborative inference. Reviews of anomalybased methods establish a broad methodological base, while also identifying persistent variation in dataset choices and validation practices [2,3]. Artificial intelligence surveys explain the appeal of automating pattern recognition over volumes of heterogeneous network records [4]. Nevertheless, an experimental classifier is not equivalent to an enterprise defence capability. Its features may depend on traffic that cannot legally or practically be collected; alerts may arrive without ownership information; and an accurate classification may still reach an analyst after the affected service has failed.

The distinction is particularly important for organisations modernising digital infrastructure in Saudi Arabia. Expansion of interconnected services increases dependence on reliable authentication, segmented networks, managed endpoints, and recoverable systems. The Saudi National Cybersecurity Authority's Essential Cybersecurity Controls provide a relevant governance context, but the existence of controls does not establish the effectiveness of a particular analytical model. Research must connect cybersecurity monitoring to consequences that infrastructure owners can measure: time to notice harmful activity, time to contain it, continuity of essential functions, and evidence available for restoration decisions.

This paper therefore reviews detection technologies as components of a wider operational system. It examines the quality of their inputs, the assumptions behind reported outcomes, the handling of uncertainty, and the safeguards surrounding automated action. Its central argument is that resilience requires a credible sequence from observation to a proportionate response, supported by accountable human decisions and recoverable configurations. The article offers a framework and an assessment protocol rather than claiming to have implemented or experimentally tested a Saudi enterprise platform.

2 Aim Objectives and Review Questions

The aim is to determine how intelligent network monitoring and proactive cyber threat detection can be configured, evaluated, and governed to improve the ability of Saudi enterprises to maintain and restore essential IT services during cyber incidents. The emphasis is on translating research findings into defensible infrastructure design decisions rather than ranking algorithms by an isolated accuracy score.

Five objectives organise the review. First, identify the network, endpoint, identity, and asset data needed for reliable enterprise threat interpretation. Second, compare statistical, learningbased, graphbased, and intelligenceassisted detection according to deployment constraints and failure modes. Third, assess how enrichment, alert prioritisation, analyst investigation, and response automation affect containment opportunities. Fourth, construct an architecture that connects detection outputs with service criticality, governance, and recovery. Fifth, specify an evaluation method that tests transferability and operational consequences in Saudi enterprise settings.

UNDER PEER REVIEW IN IJAR

The corresponding questions are whether monitoring data are sufficiently representative; which analytical approaches remain dependable when traffic and attackers change; how uncertain detections should influence action; and which metrics show improvement in service resilience rather than merely in classification. These questions guide the evidence synthesis and the proposed research agenda.

3 Review Methodology

This study uses a structured integrative review because its subject crosses empirical intrusion detection, network engineering, organisational response, and cybersecurity governance. A metaanalysis would be inappropriate: the source studies differ in traffic distributions, label definitions, sampling procedures, baselines, metrics, and operating environments. Combining their reported accuracy values into a single estimate would suggest a comparability that the underlying evaluations do not support. A qualitative synthesis can instead identify consistent mechanisms, conditional findings, and unresolved methodological disagreements.

The evidence window was January 2020 through December 2025. Targeted bibliographic discovery used combinations of enterprise network monitoring, network intrusion detection, cyber threat intelligence, anomaly detection, concept drift, explainability, federated learning, adversarial attacks, alert correlation, and cyber resilience. DOI metadata and publication dates were crosschecked against registration records, and relevant survey references informed thematic coverage. Eligible works were scholarly journal articles or research reports addressing at least one of the questions above and providing a traceable DOI. Papers were selected for substantive relevance and complementary methods, not for a favourable reported result. Nonscholarly marketing material, duplicate versions, and material outside the date window were excluded from the cited set.

Thirty studies form the analytical corpus. They include surveys, model studies, dataset investigations, and research on security operations. Their heterogeneity is deliberate: an enterprise architecture cannot be assessed using detection accuracy papers alone. An evidence matrix recorded each study's task, environment, data type, proposed method, evaluation emphasis, and limitations relevant to deployment. Study claims were interpreted in light of their own scope; laboratory results were not recast as verified Saudi operating outcomes. Where available metadata alone could not resolve an empirical detail, the discussion avoids numerical performance comparisons and limits its interpretation to the published research question and methodological implications.

The synthesis proceeded through four analytical lenses: observability, inference reliability, response usability, and service resilience. Evidence about data provenance and class imbalance informed the first; model generalisation and robustness informed the second; explainability and alert operations informed the third; and incident outcomes and recoverability informed the fourth. Points of agreement were distinguished from hypotheses requiring field trials. This procedure is reproducible at the level of DOI selection and conceptual coding, but it is a purposive integrative review rather than a claim that every eligible publication in every database was screened. That distinction constrains the conclusions appropriately.

Quality appraisal considered whether a study identified the network environment, described the provenance of its observations, separated model development from evaluation, and disclosed limitations that affect use outside the experiment. Greater interpretive weight was placed on studies discussing dataset transfer, adversarial conditions, or operational decisions rather than solely reporting a favourable headline metric. Thematic coding was used to compare mechanisms, not to assign numerical scores to studies. Because the evidence set contains both empirical and conceptual papers, no pooled effect size, publicationbias statistic, or unsupported studyselection flow count is presented. All proposed measures below are research recommendations, not results observed in Saudi organisations.

UNDER PEER REVIEW IN IJAR

4 Monitoring Data and the Problem of Visibility

The first engineering decision is which observations should be trusted. Packet payload inspection may reveal attack signatures but creates capacity, encryption, retention, and privacy problems. Flow records provide source and destination attributes, protocols, volumes, and timing with lower storage demands, although they may conceal applicationlevel intent. Domain name, proxy, authentication, endpoint, configuration, and vulnerability records provide further context. No single feed describes the full sequence from entry to persistence and service disruption. The practical objective is sufficient event linkage to support decisions, not maximal collection of every available byte.

Observability also depends on an accurate asset inventory. A connection to an unclassified address is difficult to prioritise, whereas the same event associated with a privileged identity and a businesscritical server may justify urgent investigation. Asset ownership, exposure, maintenance windows, segmentation boundaries, expected communication partners, and recovery dependencies should therefore accompany technical telemetry. Time synchronisation, stable identifiers, and documented collection gaps are prerequisites for meaningful correlation. The enterprise survey by Lyu and colleagues supports investigating asset relationships rather than interpreting each isolated record independently [1].

Training data introduce additional weaknesses. Dataset studies show how methods and benchmarks vary and why class definitions need careful scrutiny [2]. Generating labelled observations is difficult because benign administration can resemble discovery or lateral movement, while rare attacks provide too few representative examples [14,15]. Resampling can change how a learner interprets minority classes, but it does not recreate the structure of an unseen campaign [11,17]. Class weights, thresholds, and sampling procedures must be reported together; otherwise, apparent improvements may arise from a changed evaluation distribution rather than genuine detection capability.

Data access is also uneven across enterprise environments. Industrial devices may use legacy protocols or have narrow maintenance windows. Remote workers traverse encrypted tunnels, and cloud service logs may differ from onpremises packet captures. A monitoring design should document blind spots, define collection ownership, and preserve enough timestamped evidence for incident reconstruction. Visibility gaps should appear explicitly in analyst dashboards rather than being silently treated as absence of malicious activity.

5 Detection Methods and Their Limits

Signature and rulebased detection remain valuable for known malicious indicators, clear policy violations, and transparent control testing. Their weakness is dependence on recognisable patterns: modified payloads, new command sequences, and distributed lowvolume behaviour can evade a narrow rule set. Supervised learning adds flexibility by extracting distinctions from labelled examples. Surveys describe decision trees, ensemble methods, neural networks, and hybrid systems, but they also show that a preferred algorithm depends heavily on preprocessing and dataset characteristics [3,4]. A useful enterprise design retains stable rules while evaluating adaptive models as additional evidence channels.

Deep learning methods can model temporal sequences and nonlinear relations that simple thresholds miss. Work on adaptable detection for previously unseen attacks suggests potential for transferable feature learning [6]. Yet previously unseen does not automatically mean detectable: new attacks may exploit features absent from training, and a model can confuse legitimate software deployments with malicious novelty. Claims of zeroday detection therefore require a precise definition of novelty, an appropriately separated evaluation set, and operational followup. Detection performance on a heldout random sample is a weaker test than performance after equipment changes or against an attack family excluded from training.

UNDER PEER REVIEW IN IJAR

Distribution shift is a central risk. Network traffic changes with business calendars, migrations, new applications, revised access policies, and adversarial behaviour. Research on concept drift and incremental learning motivates monitoring the stability of feature distributions and model decisions over time [10]. Crossdataset generalisation studies further challenge the assumption that strong results on one benchmark transfer to another organisation [16]. Practical controls include timeordered testing, frozen baselines, shadow deployment, change logging, and rollback after unexplained alert escalation. Updating a model solely because its output volume changes can be dangerous: a genuine ongoing attack may be mistaken for benign drift.

Adversarial manipulation adds a separate source of uncertainty. Attackers may alter traffic timing, packet properties, or event sequences to reduce model confidence. Surveys and comparative studies examine how crafted inputs can affect intrusion detectors and why defensive evaluation must include an adaptive opponent [5,18,19]. Training on adversarial examples can be useful, but it is not a universal guarantee; defences may overfit the attack construction used in testing. Enterprises need a control path that remains usable when the model is distrusted, including deterministic rules, analyst escalation, and retained original logs.

Network topology provides information that flat event tables obscure. Graph methods represent hosts, identities, services, and connections as related entities, supporting analysis of propagation and coordinated behaviour [25]. Recent privacypreserving graph research also examines ways of learning from relational information without exposing all underlying traffic [26]. These methods demand careful graph construction: ephemeral addresses, shared service accounts, and missing identity links can create false relationships. Graph explanations should identify the observed edge or sequence behind a warning, not simply present a visually impressive risk map.

Softwaredefined networking offers another point of control through programmable policies and centralised traffic visibility. Survey evidence describes opportunities and security challenges across this setting [20,22]. Flowbased timeseries work illustrates how temporal observations can inform anomaly detection [21]. Centralised control also concentrates operational risk: an erroneous isolation command can interrupt several applications simultaneously. Accordingly, the ability to generate a control action should never be confused with authority to execute it without validation.

6 Enrichment Correlation and Human Decisions

Threat intelligence becomes useful when it adds meaning to local observations. An indicator may identify an infrastructure component used in a known campaign, but its value depends on freshness, confidence, context, and relevance to the monitored environment. Blindly matching a destination against a feed can create false positives, especially where services share hosting infrastructure. Intelligence should be recorded with its source, observation date, confidence, and expiration rule. A locally unusual event supported by relevant intelligence is stronger evidence than either component alone.

Collaborative detection studies propose sharing models or signals across organisations to broaden visibility beyond a single network [7,8]. Other work considers combined distributed learning and trusted exchange mechanisms [9]. The deployment question is whether shared information genuinely improves detection without leaking network structure, personal data, or confidential operations. Federated learning reduces some rawdata movement, but gradients and updates still require privacy assessment, poisoning resistance, participation governance, and mechanisms for excluding compromised contributors. Shared threat intelligence is therefore a controlled input, not an unquestioned source of truth.

Individual detections rarely reveal an entire incident. Alert correlation can assemble authentication failures, network probing, suspicious access, and data movement into a coherent investigation context [27]. Correlation rules must also handle outoforder arrival, duplicate records, missing hosts, and identity reassignment. A practical case should state what occurred, the affected asset, the supporting evidence, the competing benign explanation,

UNDER PEER REVIEW IN IJAR

and the likely service consequence. Analysts can then distinguish genuine uncertainty from a lack of useful information in the interface.

Explainable detection research addresses the need to understand why a model labelled traffic as suspicious [12,13]. Posthoc explanations may reveal influential features, yet they do not prove that the resulting attribution is causally correct. Features can reflect dataset artefacts or shortcuts. Explanation quality should be assessed by whether analysts can verify the cited behaviour, reproduce the relevant evidence, and recognise limitations. An explanation that highlights an impossible feature should prompt model review rather than confidence in the alert.

Security operations research identifies alert fatigue as a distinct organisational problem [28]. Work on collaboration between analysts and automated systems suggests that technology should reduce repetitive investigation while preserving the operator's ability to challenge a conclusion [29]. Effective triage needs deduplication, service context, clear evidence trails, and feedback about false alarms. More detections are not necessarily better if the extra workload delays response to the cases that matter. Queue age, time spent per true incident, and discarded highrisk alerts should accompany conventional precision and recall.

7 A Resilience Oriented Architecture

Figure 1 presents a conceptual architecture that connects detection with operational continuity. It begins with monitored infrastructure and identity events; the collection layer applies timestamp normalisation, provenance checks, and data minimisation. An enrichment service resolves devices, users, applications, change windows, and recovery priorities. Detection then combines fixed signatures, anomaly scoring, graph relationships, and intelligence indicators. Correlation groups signals into cases, while a response controller applies approval rules. A separate resilience layer records the effect of actions on service availability and recovery readiness.

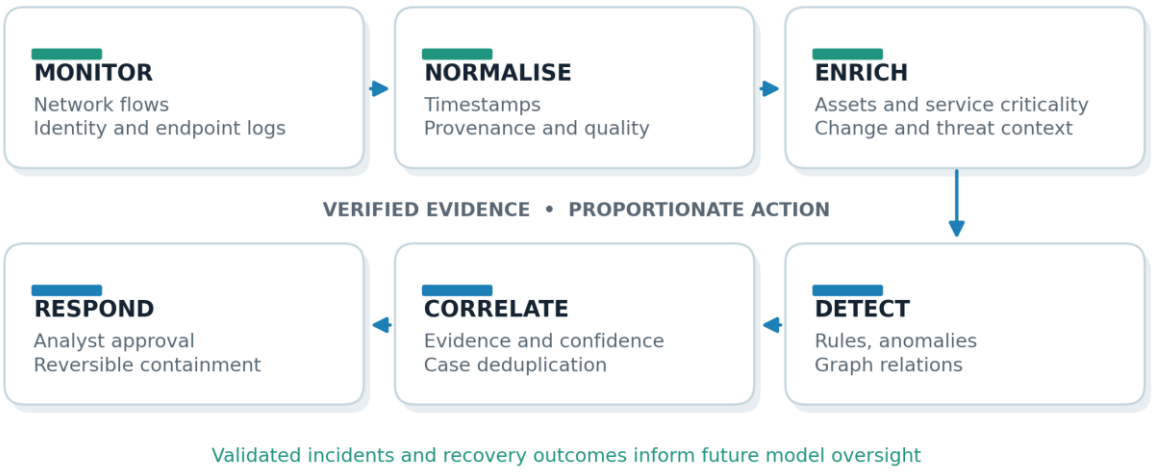

Figure 1. Proposed evidencetoresponse architecture. Arrows indicate information flow; the note identifies governed learning and recovery review.

The design deliberately separates a model's recommendation from a production change. A highconfidence alert on a noncritical test host may permit a reversible network restriction. A similar warning involving a clinical, financial, or industrial dependency may require human confirmation and a staged response. Decision rights should be defined before deployment, with clear owners for approvals, exceptions, and postincident analysis. A feedback loop returns investigation outcomes to model oversight, but retraining is undertaken only after label quality and distribution changes have been checked.

UNDER PEER REVIEW IN IJAR

Resilience is an outcome across time rather than a static property of a detector. Operational response research and industrial case analysis show why cybersecurity must consider organisational dependencies and recovery actions [23,30]. Preparation includes protected backups, knowngood configurations, alternative communication channels, and rehearsed restoration procedures. Detection can increase the time available for these measures, but it cannot substitute for them. A response system that shuts down a critical switch unnecessarily may create a service incident even while suppressing the suspicious traffic.

The framework also supports graduated autonomy. Level one presents corroborated evidence for review. Level two permits humanapproved containment with precomputed impact estimates. Level three automates tightly bounded, reversible actions under tested conditions and escalates ambiguous cases. These are proposed design levels, not a maturity score established by the reviewed literature. Their purpose is to align the strength of evidence with the potential harm of intervention.

8 Designing for Saudi Enterprise Conditions

Saudi enterprises encompass conventional corporate IT, distributed service operations, logistics, industrial environments, and publicfacing digital platforms. Their exposure profiles and recovery constraints cannot be treated as identical. Businesscritical assets should be classified through service owners, and monitoring coverage should be mapped against authentication paths, network boundaries, remote administration, and dependencies on external providers. The National Cybersecurity Authority's regulatory controls provide a national reference for governance decisions; any formal compliance assessment must examine the applicable control text and evidence separately from this review.

Local deployment also raises questions about data handling. Centralising full packet captures from multiple sites may be disproportionate or technically infeasible. Derived flow features, pseudonymised identifiers, scoped retention, rolebased access, and local processing can reduce exposure, provided investigators can still reconstruct a relevant sequence under authorised conditions. Organisations should document the legal and contractual basis for collection and sharing, including transfers between business units or providers. Privacypreserving collaboration remains an engineering research problem rather than an excuse to assume all information exchange is harmless [7,26].

Operational technology deserves special attention. Asset discovery or active probing that is routine in office networks may be unsafe for fragile controllers. Passive observation, vendor consultation, tested maintenance windows, and predefined isolation paths are more suitable starting points. Research on cyber resilience in industrial environments supports studying dependencies between technical incidents and physical operations [23]. Similarly, a cloud migration changes telemetry locations, identities, and network routes; monitoring rules must be reassessed alongside the migration plan rather than copied unchanged from the previous environment.

Skills and institutional ownership matter as much as tools. Network teams understand routes and failure domains; security analysts interpret attack evidence; application owners understand acceptable interruption; recovery teams know restoration sequences. Joint exercises should test whether these groups can assemble an incident narrative quickly and reach a proportionate decision.

Governance also requires agreement about when a warning becomes an incident. The monitoring team may identify anomalous traffic first, but the service owner is often better placed to determine whether a process can be suspended safely. A documented escalation schedule should identify substitutes for absent decision makers and specify the evidence retained after any emergency action. Testing should include conflicting recommendations from detection models, delayed identity logs, and threats during planned maintenance. These cases reveal whether a formally approved response procedure remains usable when its assumptions fail. A sound implementation must preserve accountability without allowing procedural delay to obscure a genuine attack.

UNDER PEER REVIEW IN IJAR

9 Operational Evaluation and Risk Prioritisation

Table 1 summarises the main analytical families and their operational tradeoffs. Its purpose is to prevent the common mistake of selecting the algorithm with the highest reported accuracy irrespective of evidence quality, installation cost, and consequences of failure. An enterprise evaluation should begin with a defined threat scenario, such as internal discovery followed by suspicious remote access, and specify which event sources make the scenario observable.

Table 1. Detection approaches and tradeoffs for enterprise deployment, synthesised from the reviewed literature.

Analytical family	Evidence contribution	Operational limitation	Bestsuited role
Rules and signatures	Known indicators and transparent policy violations	Incomplete coverage of novel or altered behaviours	Deterministic baseline and safety control
Supervised classification	Patterns represented in labelled flow and event data	Class imbalance and poor transfer between sites	Secondary evidence with drift monitoring
Temporal and deep learning	Complex sequences and nonlinear network behaviour	Opaque predictions and costly retraining	Candidate detection with separate validation
Graphbased analysis	Relationships among hosts, users and services	Identity ambiguity and incomplete topology	Lateral movement and multiasset correlation
Collaborative intelligence	Crossorganisation indicators and learned patterns	Data exposure, feed quality and poisoned updates	Timebounded context enrichment
Analystguided correlation	Incident narrative with corroborating events	Workload, duplicate alerts and delayed evidence	Riskranked case management

Evaluation must distinguish offline classification from live decision support. A credible dataset preserves event chronology, host identity boundaries, and the separation of training from testing environments. Maintainable log dataset research highlights the importance of documenting collection and reproduction assumptions [24]. Splitbyhost or splitbysite tests reveal transfer problems concealed by random record splits. Where attacks are rare, precision, recall, false alarms per monitored asset, and expected review workload convey more information than aggregate accuracy. Detection latency should be measured from the earliest observable hostile action, while containment latency should include analyst and approval delays.

Evaluation reports should state exactly what counts as a true incident, a false alert, an inconclusive case, and a successful intervention. Changes in asset numbers or collection coverage can alter raw alert counts even if underlying threat conditions stay constant. Reporting rates per monitored asset or service and supplying confidence intervals can make comparisons more credible. Analysts should also inspect performance separately for rare attack types and critical applications because an average may hide unacceptable failures. Operational testing must record negative effects of containment, including blocked authorised connections, user disruption, restoration effort, and any temporary loss of monitoring visibility.

Figure 2 offers a qualitative decision matrix using evidence confidence and potential service impact. A lowconfidence warning on a critical service should trigger rapid verification and protective preparation, not immediate disruptive isolation. Strong evidence against a lowimpact endpoint may justify a bounded, reversible restriction. High impact combined with strong corroboration merits incidentcommand escalation and preapproved

UNDER PEER REVIEW IN IAR

containment steps. The matrix does not allocate numerical thresholds; those require validation against an organisation's incident costs and risk appetite.

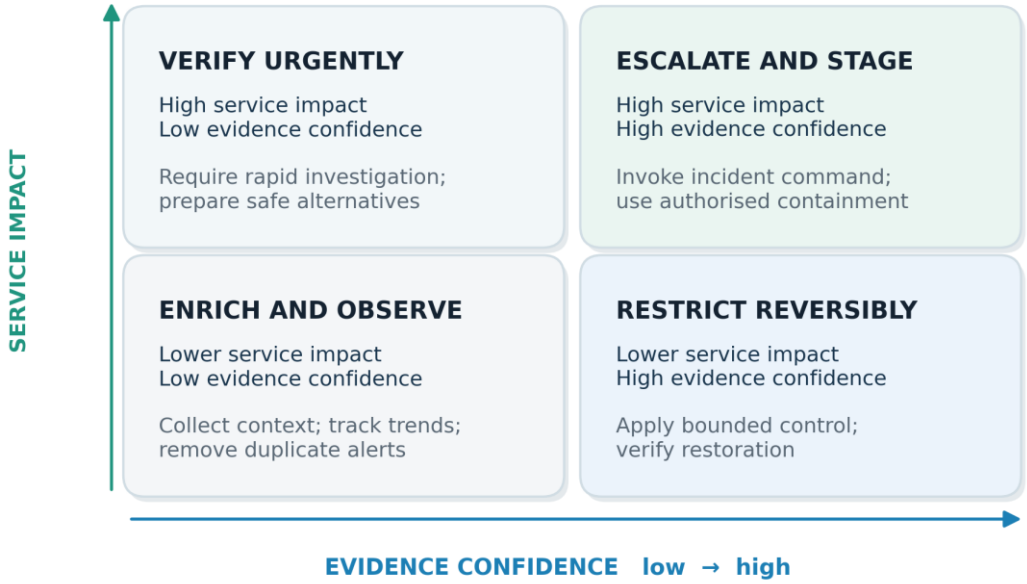

Figure 2. Conceptual incident triage matrix relating evidence confidence to potential service impact. Categories require local validation.

Table 2 translates these principles into proposed acceptance criteria. Rather than asserting universal thresholds, it specifies the evidence each deployment should collect before a model is trusted. Teams should measure alert volume across routine change windows, test response scripts in a staging environment, and inspect performance after new software or network segments are introduced. A measured reduction in mean time to detect is meaningful only if containment does not increase avoidable downtime. Formal service continuity indicators provide the essential counterweight to security model metrics.

Table 2. Proposed evaluation evidence for Saudi enterprise pilots; no numerical acceptance thresholds are assumed.

Assessment domain	Evidence to collect	Operational decision supported
Coverage and data quality	Asset coverage; missing logs; time skew; provenance	Identify blind spots before accepting detection claims
Detection reliability	Precision, recall and false alerts by site, time and threat class	Choose thresholds without masking rare critical attacks
Generalisation and drift	Timeseparated and crosssite tests; model version history	Determine when recalibration or rollback is necessary
Analyst usefulness	Investigation duration; duplicate case rate; explanation checks	Estimate staffing needs and evidence sufficiency
Response safety	Approval delay; containment reversibility; authorised service disruption	Limit automation and prevent avoidable outages

UNDER PEER REVIEW

Assessment domain	Evidence to collect	Operational decision supported
Resilience outcome	Time to detect, contain and restore; business service availability	Test whether threat response preserves continuity

10 Critical Discussion

Three tensions recur across the evidence. The first concerns sensitivity and operational workload. Models tuned to identify every unusual event can overwhelm incident responders, while aggressive filtering may conceal a slow intrusion. Alert correlation and analystcentred designs provide ways to manage that tradeoff, but published studies do not justify a universal threshold for acceptable false alarms [2729]. Sitespecific cost modelling and review capacity should determine the operating point, with periodic reconsideration as the threat environment changes.

The second tension concerns adaptation and assurance. Incremental learning may restore relevance under changed traffic, yet uncontrolled updates can absorb attacker behaviour into a new definition of normal. Generalisation research, dataset scrutiny, and adversarial evaluations jointly favour controlled change management [5,10,16]. Model versions, training provenance, threshold revisions, and rollback decisions should be auditable. Independent rulebased safeguards remain important when learning systems behave unexpectedly.

The third tension concerns security action and business availability. Faster containment can interrupt an attack, but incorrect network isolation can also interrupt the enterprise. The relevant objective is minimising total service harm, including cyber consequences, response side effects, and recovery time. This is why asset criticality and decision accountability belong inside the detection architecture rather than in a separate policy document. Research linking organisational resilience to incident outcomes provides a basis for examining the relationship empirically [30].

The literature nevertheless has limitations. Benchmark studies often share datasets, definitions, and experimental assumptions, so numerous positive results do not necessarily represent independent evidence. Studies of alert management and operational recovery are less directly comparable with model experiments. Saudispecific evidence on integrated detectiontorecovery pipelines is sparse within this selected corpus. The review itself is purposive, not exhaustive; its conclusions are conceptual propositions that require field validation. No numerical improvement in Saudi enterprise resilience can responsibly be attributed to the proposed framework at this stage.

11 Research Agenda and Practical Implications

The strongest next study would be a prospective, multisite evaluation that begins with a baseline security operations process. Participating organisations could compare ruleonly monitoring with an additional enrichment and correlation layer, while preserving incident response procedures and documenting each deployment change. Network events should be collected under approved governance arrangements, with identifiers handled according to access policies. Analysts should label investigations independently where feasible, and disagreements should be resolved through recorded evidence rather than majority voting alone.

The study should test at least four conditions: routine enterprise traffic, planned infrastructure change, a simulated internal attack sequence, and an unexpected service incident. Redteam exercises can supply reproducible attack narratives, but their coverage must not be presented as proof of detection against all real adversaries. Prospective observations should capture falsealert burden, evidence completeness, response latency, escalation quality, containment side effects, and recovery time. Randomised scenario order or matched baseline periods would improve causal interpretation, although such designs may be constrained by operational safety.

UNDER PEER REVIEW IN IJAR

External validity deserves explicit attention. Results from a Riyadh office network may not generalise to distributed industrial sites or enterprises with different identity platforms. The same detection pipeline should be evaluated under changing asset inventories, encrypted traffic, and altered log retention rules. Model transfer should be judged on postdeployment behaviour, not only on a benchmark score. Public reporting should describe anonymisation procedures, missing telemetry, service criticality, and any incidents excluded from independent analysis.

For practitioners, the nearterm recommendation is to establish an observability register, build a conservative correlation prototype, and rehearse reversible response before adopting extensive autonomous action. Procurement should request evidence of crossenvironment validation, transparent alert explanations, update governance, and measurable analyst workload. Security operations, infrastructure engineering, and business continuity owners should jointly sign off on the acceptance criteria. These steps make intelligent monitoring an accountable service capability rather than an isolated software acquisition.

12 Conclusion

Intelligent monitoring can improve enterprise cyber resilience when the organisation understands what is observable, how uncertain a detection is, and which actions preserve essential services. The reviewed work offers promising analytical mechanisms, including adaptive anomaly detection, graph relationships, collaborative intelligence, and alert correlation. It also exposes persistent weaknesses in benchmark transfer, model robustness, explainability, and workload assessment. This review responds by proposing a layered architecture in which asset context, human judgment, bounded automation, and recovery evidence are inseparable from detection. For Saudi enterprises, the next priority is controlled field evaluation against nationally relevant governance requirements and measurable service continuity outcomes. Publication of credible operational results, including failures and false alarms, would provide stronger evidence for adoption than another isolated claim of high classification accuracy.

UNDER PEER REVIEW IN IJAR

References

1. Minzhao Lyu, Hassan Habibi Gharakheili, Vijay Sivaraman (2024). A Survey on Enterprise Network Security: Asset Behavioral Monitoring and Distributed Attack Detection. IEEE Access, 12, 8936389383. https://doi.org/10.1109/access.2024.3419068
2. Zhen Yang, Xiaodong Liu, Tong Li, Di Wu, Jinjiang Wang, Yunwei Zhao, et al. (2022). A systematic literature review of methods and datasets for anomalybased network intrusion detection. Computers & Security, 116, 102675. https://doi.org/10.1016/j.cose.2022.102675
3. Oluwadamilare Harazeem Abdulganiyu, Taha Ait Tchakoucht, Yakub Kayode Saheed (2023). A systematic literature review for network intrusion detection system (IDS). International Journal of Information Security, 22(5), 11251162. https://doi.org/10.1007/s10207023006822
4. Mohammed Sayeeduddin Habeeb, T. Ranga Babu (2022). Network intrusion detection system: A survey on artificial intelligence-based techniques. Expert Systems, 39(9), e13066. https://doi.org/10.1111/exsy.13066
5. Ke He, Dan Dongseong Kim, Muhammad Rizwan Asghar (2023). Adversarial Machine Learning for Network Intrusion Detection Systems: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, 25(1), 538566. https://doi.org/10.1109/comst.2022.3233793
6. Mahdi Soltani, Behzad Ousat, Mahdi Jafari Siavoshani, Amir Hossein Jahangir (2023). An adaptable deep learningbased intrusion detection system to zeroday attacks. Journal of Information Security and Applications, 76, 103516. https://doi.org/10.1016/j.jisa.2023.103516
7. Mohanad Sarhan, Siamak Layeghy, Nour Moustafa, Marius Portmann (2022). Cyber Threat Intelligence Sharing Scheme Based on Federated Learning for Network Intrusion Detection. Journal of Network and Systems Management, 31(1), 3. https://doi.org/10.1007/s10922022096913
8. Massimo Guarascio, Nunziato Cassavia, Francesco Sergio Pisani, Giuseppe Manco (2022). Boosting CyberThreat Intelligence via Collaborative Intrusion Detection. Future Generation Computer Systems, 135, 3043. https://doi.org/10.1016/j.future.2022.04.028
9. Tongtong Jiang, Guowei Shen, Chun Guo, Yunhe Cui, Bo Xie (2023). BFLS: Blockchain and Federated Learning for sharing threat detection models as Cyber Threat Intelligence. Computer Networks, 224, 109604. https://doi.org/10.1016/j.comnet.2023.109604
10. Farah Jemili, Khaled Jouini, Ouajdi Korbaa (2024). Intrusion detection based on concept drift detection and online incremental learning. International Journal of Pervasive Computing and Communications, 21(1), 81115. https://doi.org/10.1108/ijpcc1220230358
11. Lan Liu, Pengcheng Wang, Jun Lin, Langzhou Liu (2021). Intrusion Detection of Imbalanced Network Traffic Based on Machine Learning and Deep Learning. IEEE Access, 9, 75507563. https://doi.org/10.1109/access.2020.3048198
12. Yiwen Wang, Lei Xu, Wanli Liu, Rongzhen Li, Junjie Gu (2023). Network Intrusion Detection Based on Explainable Artificial Intelligence. Wireless Personal Communications, 131(2), 11151130. https://doi.org/10.1007/s11277023104727
13. Pieter Barnard, Nicola Marchetti, Luiz A. DaSilva (2022). Robust Network Intrusion Detection Through Explainable Artificial Intelligence (XAI). IEEE Networking Letters, 4(3), 167171. https://doi.org/10.1109/lnet.2022.3186589
14. Jose Carlos Mondragon, Paula Branco, GuyVincent Jourdan, Andres Eduardo GutierrezRodriguez, Rajesh Roshan Biswal (2025). Advanced IDS: a comparative study of datasets and machine learning algorithms for network flowbased intrusion detection systems. Applied Intelligence, 55(7), 608. https://doi.org/10.1007/s10489025064224
15. Ryosuke Ishibashi, Kohei Miyamoto, Chansu Han, Tao Ban, Takeshi Takahashi, Jun'ichi Takeuchi (2022). Generating Labeled Training Datasets Towards Unified Network Intrusion Detection Systems. IEEE Access, 10, 5397253986. https://doi.org/10.1109/access.2022.3176098
16. Marco Cantone, Claudio Marrocco, Alessandro Bria (2024). Machine Learning in Network Intrusion Detection: A CrossDataset Generalization Study. IEEE Access, 12, 144489144508. https://doi.org/10.1109/access.2024.3472907
17. Sikha Bagui, Kunqi Li (2021). Resampling imbalanced data for network intrusion detection datasets. Journal of Big Data, 8(1), 6. https://doi.org/10.1186/s4053702000390x
18. Elie Alhajjar, Paul Maxwell, Nathaniel Bastian (2021). Adversarial machine learning in Network Intrusion Detection Systems. Expert Systems with Applications, 186, 115782. https://doi.org/10.1016/j.eswa.2021.115782
19. Houda Jmila, Mohamed Ibn Khedher (2022). Adversarial machine learning for network intrusion detection: A comparative study. Computer Networks, 214, 109073. https://doi.org/10.1016/j.comnet.2022.109073
20. Vitor Gabriel da Silva Ruffo, Daniel Matheus Brandão Lent, Mateus Komarchesqui, Vinícius Ferreira Schiavon, Marcos Vinicius Oliveira de Assis, Luiz Fernando Carvalho, et al. (2024). Anomaly and intrusion detection using deep learning for softwaredefined networks: A survey. Expert Systems with Applications, 256, 124982. https://doi.org/10.1016/j.eswa.2024.124982
21. Sultan Zavrak, Murat Iskefiyeli (2023). Flowbased intrusion detection on softwaredefined networks: a multivariate time series anomaly detection approach. Neural Computing and Applications, 35(16), 1217512193. https://doi.org/10.1007/s00521023083765
22. Tohid Jafarian, Mohammad Masdari, Ali Ghaffari, Kambiz Majidzadeh (2020). A survey and classification of the security anomaly detection mechanisms in software defined networks. Cluster Computing, 24(2), 12351253. https://doi.org/10.1007/s10586020031841

UNDER PEER REVIEW IN IJAR

23. Kirsty Perrett, Ian David Wilson (2023). A cyber resilience analysis case study of an industrial operational technology environment. Environment Systems and Decisions, 43(2), 178190. https://doi.org/10.1007/s10669023098951
24. Max Landauer, Florian Skopik, Maximilian Frank, Wolfgang Hotwagner, Markus Wurzenberger, Andreas Rauber (2023). Maintainable Log Datasets for Evaluation of Intrusion Detection Systems. IEEE Transactions on Dependable and Secure Computing, 20(4), 34663482. https://doi.org/10.1109/tdsc.2022.3201582
25. Tristan Bilot, Nour El Madhoun, Khaldoun Al Agha, Anis Zouaoui (2023). Graph Neural Networks for Intrusion Detection: A Survey. IEEE Access, 11, 4911449139. https://doi.org/10.1109/access.2023.3275789
26. Xinjun Pei, Xiaoheng Deng, Shengwei Tian, Ping Jiang, Yunlong Zhao, Kaiping Xue (2025). A PrivacyPreserving Graph Neural Network for Network Intrusion Detection. IEEE Transactions on Dependable and Secure Computing, 22(1), 740756. https://doi.org/10.1109/tdsc.2024.3417853
27. Bronagh Lanigan, Zeinab Rezaeifar, Federico Cruciani, Michael Milliken, Jordan Vincent, Samuel Moore, et al. (2025). Alert correlation for intelligent threat detection and response. Intelligent Systems with Applications, 28, 200606. https://doi.org/10.1016/j.iswa.2025.200606
28. Shahroz Tariq, Mohan Baruwal Chhetri, Surya Nepal, Cecile Paris (2025). Alert Fatigue in Security Operations Centres: Research Challenges and Opportunities. ACM Computing Surveys, 57(9), 138. https://doi.org/10.1145/3723158
29. Mohan Baruwal Chhetri, Shahroz Tariq, Ronal Singh, Fatemeh Jalalvand, Cecile Paris, Surya Nepal (2024). Towards HumanAI Teaming to Mitigate Alert Fatigue in Security Operations Centres. ACM Transactions on Internet Technology, 24(3), 122. https://doi.org/10.1145/3670009
30. Elinor Tsen, Ryan K L Ko, Sergeja Slapničar (2025). The effect of organizational cyber resilience on cyber incident outcomes. Journal of Cybersecurity, 11(1), tyaf040. https://doi.org/10.1093/cybsec/tyaf040

UNDER PEER REVIEW IN IJAR