IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

AI-Enabled Cyber Threat Detection Using Behavioral Analytics

Prof. Renuka Vinayak Kalyankar[1] and Prof. Prathmesh Bharat Yelmame[2]
Assistant Professor, JSPM's Jayawant Institute of Management Studies, Tathawade, Pune[1,2]
Savitribai Phule Pune University, Pune
renukakalyankar20@gmail.com and prathm.py@gmail.com

Abstract: *Cyber threats increasingly exploit legitimate credentials, trusted applications, and ordinary network protocols, making signature-only security controls insufficient for identifying subtle misuse. This paper examines how artificial intelligence (AI) combined with behavioural analytics can strengthen cyber threat detection by learning patterns of activity across identities, endpoints, applications, and networks. The study adopts a structured narrative review and conceptual research design based exclusively on secondary sources, including peer-reviewed literature, established benchmark-dataset publications, and cybersecurity guidance from recognised institutions. It synthesises anomaly detection, supervised classification, sequence modelling, graph-based analysis, and explainable AI, and proposes an integrated architecture that combines telemetry normalisation, behavioural baselining, contextual risk scoring, analyst review, and feedback governance. The analysis indicates that behavioural analytics is valuable where malicious activity departs from an entity's historical pattern or from the expected behaviour of comparable peers. However, the literature also demonstrates that high benchmark accuracy does not automatically translate into operational effectiveness. Dataset age, class imbalance, concept drift, adversarial manipulation, privacy constraints, alert fatigue, and limited explainability can undermine deployment. Accordingly, the paper recommends time-aware and entity-aware evaluation, cross-environment testing, precision–recall and false-positive reporting, calibrated risk scores, human-in-the-loop response, and continuous model governance. The paper does not claim new experimental results; its findings are analytical conclusions derived from the reviewed evidence. It concludes that AI-enabled behavioural detection should complement, rather than replace, conventional controls and should be implemented as a measurable, accountable component of a broader security operations capability.*

Keywords: Artificial intelligence; behavioural analytics; cyber threat detection; anomaly detection; intrusion detection; machine learning; cybersecurity

I. INTRODUCTION

Digital organisations generate extensive records of user authentication, endpoint processes, application transactions, network flows, cloud operations, and device communications. These records create an opportunity to identify suspicious behaviour before it develops into a major incident. At the same time, their volume, velocity, and diversity make manual inspection impractical. Cybersecurity teams therefore increasingly rely on analytics to prioritise events and identify patterns that rule-based systems may overlook. AI-enabled behavioural analytics applies statistical learning and machine-learning methods to establish expected patterns of activity and identify deviations that may indicate compromise.

Traditional signature-based detection remains important because known malware hashes, malicious domains, and recognised attack patterns can be matched efficiently. Its limitation is that it depends on previously observed indicators or explicit rules. Attackers can change infrastructure, use legitimate administrative tools, compromise valid accounts, or

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, November 2026 Impact Factor: 8.2

move slowly to avoid fixed thresholds. Anomaly detection offers a complementary approach: rather than asking only whether an event matches a known malicious signature, it asks whether the event is unusual in its context. Chandola et al. (2009) describe anomaly detection as the identification of patterns that do not conform to expected behaviour, while Buczak and Guven (2016) explain how data-mining and machine-learning methods can support intrusion detection.

Behaviour, however, is not synonymous with maliciousness. A senior administrator may legitimately access many systems; a travelling employee may authenticate from a new location; and a software deployment may generate a sudden increase in network traffic. A detection system that treats every deviation as an attack will produce excessive false positives. Effective behavioural detection must therefore account for identity role, asset criticality, time, peer groups, business processes, threat intelligence, and the confidence of the model. Sommer and Paxson (2010) caution that machine-learning-based network intrusion detection faces substantial challenges when laboratory assumptions meet operational networks.

This paper analyses the potential and limitations of AI-enabled cyber threat detection using behavioural analytics. It develops a conceptual architecture and a practical evaluation framework rather than reporting a new experiment. The central argument is that detection quality depends not only on algorithm selection but also on the quality of telemetry, the relevance of behavioural context, the validity of evaluation, and the governance of automated decisions. The paper is organised around literature, objectives, research gap, methodology, analytical discussion, findings, and conclusion.

II. REVIEW OF LITERATURE

2.1 Foundations of anomaly and behavioural detection.

Chandola et al. (2009) provide a broad taxonomy of anomaly-detection methods, distinguishing point, contextual, and collective anomalies. This distinction is directly relevant to cybersecurity. A single failed login may be ordinary noise; the same event may become suspicious when it occurs at an unusual hour, from an unfamiliar device, and is followed by privilege escalation. Collective anomalies emerge when a sequence or group of individually plausible events forms a suspicious pattern. The taxonomy clarifies why event-level thresholds alone are insufficient for sophisticated threat detection.

Buczak and Guven (2016) review data-mining and machine-learning approaches for cyber intrusion detection, covering supervised, unsupervised, and hybrid methods. Supervised models can classify known attack categories when labelled examples are representative. Unsupervised and semi-supervised methods are useful where labels are incomplete, particularly for previously unseen behaviour. Their review also highlights the importance of feature selection and dataset quality. These findings support a layered design in which known threats and deviations from normal activity are analysed together rather than forcing one model to solve every detection task. Sarker et al. (2020) further discuss cybersecurity data science from a machine-learning perspective, reinforcing the importance of treating data preparation, feature representation, and the operational security task as connected design decisions.

2.2 Operational limits of machine learning.

Sommer and Paxson (2010) argue that network intrusion detection is unusually difficult because operational traffic is heterogeneous, the environment changes, and the cost of false alarms is high. Their critique remains important: a model that performs well on a static benchmark may fail when deployed on a different network, user population, or application mix. This challenges the assumption that high accuracy alone demonstrates readiness. In practical environments, class imbalance can make accuracy misleading because benign traffic dominates. Precision, recall, false-positive rate, calibration, and detection delay should also be considered.

2.3 Deep learning and adaptive detection.

Deep-learning methods can learn representations from sequences, flows, and high-dimensional telemetry. Mirsky et al. (2018), through Kitsune, demonstrate an online network intrusion detection approach based on an ensemble of autoencoders, designed to operate with limited computational resources. Such methods can detect deviations without requiring a fully labelled catalogue of attacks. Nevertheless, an autoencoder reconstruction error is not itself an explanation of malicious intent. Threshold selection, drift, and the need for analyst interpretation remain material

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, November 2026 Impact Factor: 8.2

concerns. Recent surveys, including Salem et al. (2024), describe a broad range of AI-driven detection approaches and emphasise continuing challenges involving datasets, scalability, and deployment.

2.4 Datasets and evaluation.

Public datasets make reproducible comparison possible, but they also shape what researchers can detect. Moustafa and Slay (2015) introduced UNSW-NB15 to provide modern normal and attack traffic with a range of features and attack categories. Sharafaldin et al. (2018) presented CICIDS2017, including benign traffic and several attack scenarios intended to reflect contemporary network activity at the time of collection. Ring et al. (2019) systematically reviewed network-based intrusion detection datasets and documented differences in capture environments, labelling, features, and suitability. These studies show that dataset choice is not a neutral technical detail: it affects the threats represented and the validity of comparisons.

2.5 Explainability and trustworthy AI.

Ribeiro et al. (2016) introduced LIME as a model-agnostic method for explaining individual predictions, while Lundberg and Lee (2017) proposed SHAP as a unified framework for feature attribution. In security operations, explanations can help analysts understand why an event was prioritised—for example, unusual access time, a new device, or an atypical data-transfer volume. Yet explanations must be interpreted carefully: feature attribution describes model behaviour, not necessarily causal proof of an attack. Biggio and Roli (2018) further demonstrate that machine-learning systems can be vulnerable to adversarial manipulation, a concern that applies to behavioural models whose inputs can be influenced by attackers.

2.6 Institutional guidance and security operations.

NIST SP 800-94 describes intrusion detection and prevention systems and their role in monitoring and responding to suspicious activity (Scarfone & Mell, 2007). NIST SP 800-61 Rev. 2 places detection within the wider incident-response lifecycle, linking analysis to containment, eradication, recovery, and lessons learned (Cichonski et al., 2012). The NIST Cybersecurity Framework 2.0 similarly positions continuous monitoring and risk management within organisational cybersecurity outcomes (National Institute of Standards and Technology [NIST], 2024). These sources imply that behavioural analytics should be integrated into operational workflows, not treated as an isolated classifier.

Overall, the literature supports the value of machine learning for identifying suspicious patterns but also warns against over-reliance on benchmark scores. Classical methods can be efficient and interpretable; deep models can capture complex patterns but may be difficult to explain; unsupervised methods can surface unknown behaviours but often generate noisy alerts. A robust system therefore needs complementary models, context enrichment, disciplined evaluation, and analyst oversight.

III. RESEARCH OBJECTIVES

The principal objective is to examine how AI-enabled behavioural analytics can improve the detection and prioritisation of cyber threats in organisational information systems. The specific objectives are:

- To explain the role of behavioural baselines and contextual anomaly detection across identity, endpoint, application, and network telemetry.
- To compare major AI approaches—including supervised learning, unsupervised anomaly detection, deep learning, and hybrid models—in terms of suitability and limitations.
- To identify data, evaluation, privacy, explainability, and operational challenges affecting deployment. To propose an integrated conceptual architecture and a validation framework for trustworthy behavioural threat detection.
- To derive practical recommendations for security teams implementing AI-assisted monitoring and incident response.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

IV. RESEARCH GAP

Existing research provides substantial coverage of anomaly-detection algorithms, intrusion-detection datasets, and AI model families. However, several issues remain insufficiently connected. First, many studies concentrate on network-flow classification, while operational threat detection requires correlating identity, endpoint, cloud, application, and network events. Second, static benchmark evaluation can conceal whether a model generalises across time, sites, devices, and user populations. Third, many studies report accuracy without sufficient attention to alert volume, analyst workload, calibration, detection delay, and the cost of false positives. Fourth, model explainability and human decision-making are often discussed separately from detection performance. Fifth, adaptive models may respond to drift but can also be manipulated or poisoned if feedback and retraining are not governed.

The gap addressed by this paper is therefore integrative and methodological: there is a need for a practical framework that connects multi-source behavioural evidence, contextual risk scoring, explainability, operational evaluation, and governance. The paper does not claim that no prior architecture combines these elements. Rather, it synthesises established research into a coherent conceptual model and identifies the evidence that should be collected before operational deployment.

V. RESEARCH METHODOLOGY

5.1 Research design and approach.

The study uses a qualitative, structured narrative review and conceptual synthesis. It is explanatory and analytical: it compares established methods, identifies recurring limitations, and develops a proposed architecture. It is not a systematic review with a registered protocol, nor is it an experimental evaluation. Accordingly, no primary dataset, participant sample, survey, interview, or measured model result is claimed.

5.2 Data sources.

The evidence base comprises peer-reviewed research on anomaly detection, intrusion detection, machine learning, deep learning, explainability, and adversarial machine learning; benchmark dataset publications; and institutional cybersecurity guidance. Key sources include Chandola et al. (2009), Sommer and Paxson (2010), Buczak and Guven (2016), Mirsky et al. (2018), Ring et al. (2019), and NIST publications. Sources were selected for relevance to behavioural detection, methodological influence, dataset provenance, or operational guidance. This is a purposive source selection, not a statistically representative sample of all publications.

5.3 Primary and secondary data distinction.

No primary data were collected. All claims about prior methods and datasets are based on secondary sources or the original publications describing those datasets. Figures 1–3 are original conceptual diagrams prepared for this paper; they are not reproductions of published figures. Tables 1–3 are analytical summaries created by the author from the cited literature. They do not report newly measured performance.

5.4 Analytical procedure.

The literature was synthesised using four dimensions: (a) behavioural evidence and telemetry; (b) learning approach and detection capability; (c) evaluation and generalisation; and (d) operational trustworthiness, including explainability, privacy, and response governance. Methods were compared according to their data requirements, ability to identify known or novel patterns, interpretability, and deployment constraints. The conceptual architecture was then developed by mapping these dimensions into a continuous detection workflow.

5.5 Methodological limitations.

A narrative review is subject to selection and interpretation bias and does not provide the exhaustive coverage or reproducibility of a fully specified systematic review. The underlying studies use different datasets, labels, attack definitions, feature sets, and train-test procedures; their numerical results are therefore not pooled or ranked here. The paper's conclusions should be understood as a synthesis and set of design recommendations, not proof that a particular model achieves a specified detection rate.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

VI. DATA ANALYSIS AND DISCUSSION

6.1 Behavioural telemetry as the foundation.

Behavioural analytics begins with the selection of observable events. Identity telemetry includes authentication time, source location, device identity, failed logins, MFA events, privilege changes, and access to applications. Endpoint telemetry includes process creation, executable lineage, file access, registry or configuration changes, and device posture. Network telemetry includes flow duration, bytes transferred, destination diversity, DNS requests, protocol usage, and connection frequency. Cloud and application telemetry adds API calls, administrative actions, object access, and data-sharing events. Each source offers only a partial view; correlation increases the likelihood that a sequence can be interpreted correctly.

Figure 1 presents the proposed high-level architecture. Telemetry is normalised before modelling so that timestamps, entity identifiers, asset labels, and event semantics can be compared. Context enrichment adds role, asset criticality, known maintenance windows, and threat intelligence. The AI layer generates anomaly and risk scores, while the response layer supports investigation and controlled action. This architecture reflects the operational emphasis of NIST incident-response guidance and the limitations identified in machine-learning intrusion detection (Cichonski et al., 2012; Sommer & Paxson, 2010).

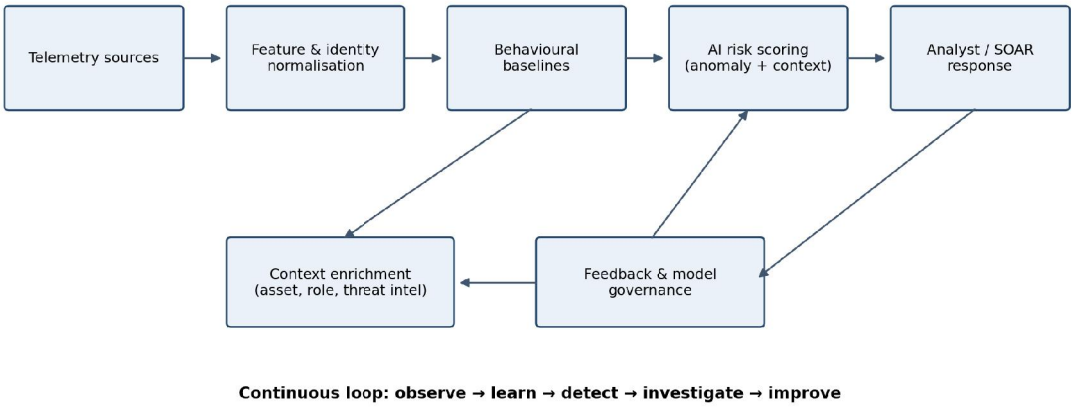

Figure 1 Proposed architecture for AI-enabled behavioural cyber threat detection

Source: Author's conceptual synthesis, informed by Buczak and Guven (2016), Cichonski et al. (2012), and Sommer and Paxson (2010).

6.2 Establishing behavioural baselines.

A baseline describes expected activity for an entity or a relevant peer group over a defined period. Individual baselines are useful for detecting changes in a specific account or host; peer-group baselines help distinguish legitimate variation between roles. For example, a database administrator and a sales employee should not be judged against the same access profile. Baselines can be time-conditioned to account for working hours, weekly patterns, and seasonal activity. A robust system should also represent uncertainty: a new employee or newly deployed service may not yet have enough history for a reliable individual baseline.

Figure 2 illustrates how identity, endpoint, and network evidence can be compared across time and peer groups before risk is aggregated. The figure is a conceptual synthesis of anomaly taxonomies and multi-source intrusion detection research (Chandola et al., 2009; Buczak & Guven, 2016). A key design principle is that one unusual feature should rarely trigger a severe response by itself. A new login location may be benign; a new location combined with impossible travel, unfamiliar device posture, privilege escalation, and unusual data access is more concerning.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

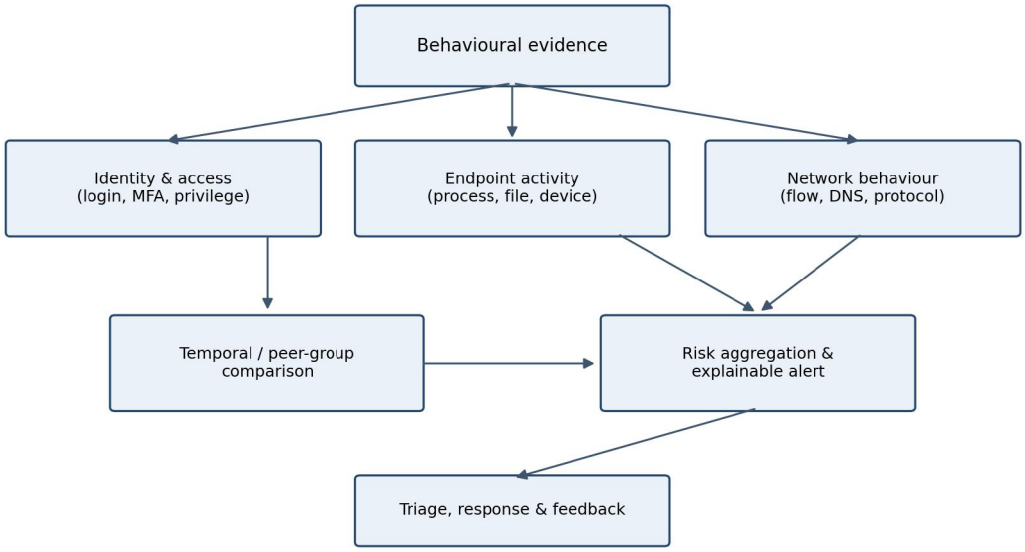

Figure 2 multi-source behavioural evidence and contextual risk aggregation

Source: Author's conceptual synthesis based on Chandola et al. (2009) and Buczak and Guven (2016).

6.3 Comparing AI approaches.

Supervised classification learns from labelled benign and malicious examples. It is useful for recurring attack patterns and can provide strong discrimination when labels are accurate and representative. Its weakness is dependence on labelled coverage: a model trained on known attack families may not recognise novel tactics. Unsupervised methods, such as clustering and isolation-based approaches, can identify rare or atypical events without complete labels. Their challenge is that rare does not necessarily mean malicious. Deep learning can model sequences and complex feature interactions, but may require more data and compute, and may be difficult to explain. Hybrid systems combine these methods with rules, threat intelligence, or analyst feedback to balance coverage and precision.

Table 1 compares the principal approaches. The comparison is qualitative and synthesises the strengths and limitations discussed in the literature; it is not a benchmark ranking (Buczak & Guven, 2016; Mirsky et al., 2018; Sommer & Paxson, 2010).

Table 1 Comparative overview of AI approaches for behavioural threat detection

Approach	Typical use	Strengths	Limitations
Supervised learning	Classifying labelled attacks and benign events	Effective for represented attack classes; measurable decision boundaries	Requires reliable labels; weak coverage of novel behaviours
Unsupervised anomaly detection	Finding rare or unusual activity without full labels	Can surface previously unlabelled deviations	High false-positive potential; unusual does not mean malicious
Deep learning / sequence models	Learning complex temporal and feature relationships	Can model high-dimensional and sequential behaviour	Data- and compute-intensive; explanations may be difficult
Hybrid / ensemble systems	Combining rules, classifiers, anomaly scores and context	Balances known-threat detection and anomaly discovery	Integration complexity; calibration and maintenance required

Note. Author's synthesis of the cited literature; no new empirical measurements are reported.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

6.4 Risk scoring and alert prioritisation.

Behavioural systems should distinguish anomaly scores from risk scores. An anomaly score expresses how unusual an event is relative to a model or baseline. A risk score should additionally incorporate asset importance, account privilege, threat context, event sequence, and confidence. For example, unusual activity on a low-impact test machine may warrant monitoring, whereas similar behaviour on a privileged identity controlling production infrastructure may justify urgent investigation. Scores should be calibrated against operational outcomes where reliable labels exist, and thresholds should be chosen according to the organisation's tolerance for missed threats and analyst capacity.

A practical risk aggregation process can combine multiple evidence categories without assuming that all signals are independent. Correlated alerts should be grouped into incidents to reduce duplication. Security teams should preserve the underlying events and model version so that an analyst can reconstruct why a score was assigned. Explainability methods such as LIME and SHAP can help identify influential features, but explanations should be presented as model evidence rather than definitive causal statements (Lundberg & Lee, 2017; Ribeiro et al., 2016).

6.5 Dataset validity and model evaluation.

The quality of a behavioural detector cannot be established by a single random train-test split. If records from the same host, user, or time period appear in both training and testing sets, the model may exploit repeated patterns and produce optimistic results. Evaluation should therefore include chronological splits, entity-separated splits, and, where possible, cross-site or cross-dataset testing. The benchmark publications UNSW-NB15 and CICIDS2017 are useful for reproducibility, but neither can represent every modern enterprise environment (Moustafa & Slay, 2015; Sharafaldin et al., 2018). Ring et al. (2019) further show that datasets differ in their collection and labelling characteristics.

Table 2 summarises commonly used evaluation dimensions. Accuracy alone can be misleading under class imbalance. Precision measures the proportion of alerts that are true positives; recall measures the proportion of relevant malicious events detected. The false-positive rate is especially important because a large number of benign events may overwhelm analysts. F1-score balances precision and recall but does not reflect the relative operational cost of different errors. Detection latency, throughput, memory use, and analyst workload should also be measured where deployment is intended.

Table 2 Recommended evaluation dimensions for behavioural detection

Metric / dimension	What it measures	Why it matters
Precision	Share of alerts that are true positives	Indicates alert usefulness and analyst burden
Recall	Share of relevant malicious events detected	Indicates coverage and missed-threat risk
False-positive rate	Benign events incorrectly flagged	Important for alert volume in imbalanced traffic
F1-score	Harmonic mean of precision and recall	Useful summary, but hides different error costs
Detection latency	Time from observable activity to alert	Determines whether response can limit impact
Cross-time / cross-site validation	Generalisation beyond the training context	Tests drift and environmental dependence
Resource use & throughput	Compute, memory, and events processed per unit time	Determines operational feasibility
Calibration & analyst workload	Reliability of risk scores and effort to resolve alerts	Connects model output to security operations

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-41212

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

6.6 Concept drift and adaptive learning.

Behaviour changes as organisations introduce new applications, change working patterns, migrate to cloud services, or restructure access privileges. A model trained on historical behaviour may therefore become less accurate over time. Drift monitoring should track feature distributions, score distributions, alert rates, and confirmed incident outcomes. Retraining should not occur automatically from every analyst disposition: attackers may deliberately generate misleading activity or exploit weak labels. A governed process should validate feedback, preserve training lineage, compare candidate models with the incumbent, and support rollback.

6.7 Adversarial resilience and data poisoning.

Attackers may attempt to evade behavioural detection by gradually changing activity, mimicking legitimate users, or manipulating the data used for learning. Biggio and Roli (2018) explain that adversarial machine learning creates threats across training and inference. For behavioural analytics, this implies that models should not trust a single telemetry source, should detect sensor gaps and unusual logging changes, and should maintain deterministic security controls for high-risk actions. AI should support, not replace, access control, endpoint protection, network segmentation, and incident-response procedures.

6.8 Explainability, privacy, and human oversight.

Behavioural monitoring can reveal sensitive information about employee routines, locations, communications, and work patterns. Organisations should define a lawful purpose, minimise collected data, restrict access, establish retention limits, and assess privacy impacts. Role-based access and pseudonymisation may reduce exposure while preserving analytical utility. Analysts should receive concise explanations that identify the events, features, baseline deviations, and contextual factors contributing to an alert. High-impact actions—such as disabling accounts or isolating critical systems—should generally require human approval or carefully tested policy controls. NIST guidance places detection within a managed response lifecycle, reinforcing the need for accountable decision-making (Cichonski et al., 2012; NIST, 2024).

6.9 Proposed validation workflow.

Figure 3 outlines a validation process from data collection to shadow deployment. Before automated response is enabled, the model should run in observation mode and its alerts should be compared with analyst assessments. Validation should include representative benign variation, known attack scenarios, rare-event handling, adversarial tests, and operational constraints. A release decision should consider detection quality, false-positive burden, latency, resource consumption, explainability, privacy, and rollback readiness. This approach is consistent with the operational concerns identified by Sommer and Paxson (2010) and the incident-response lifecycle described by Cichonski et al. (2012).

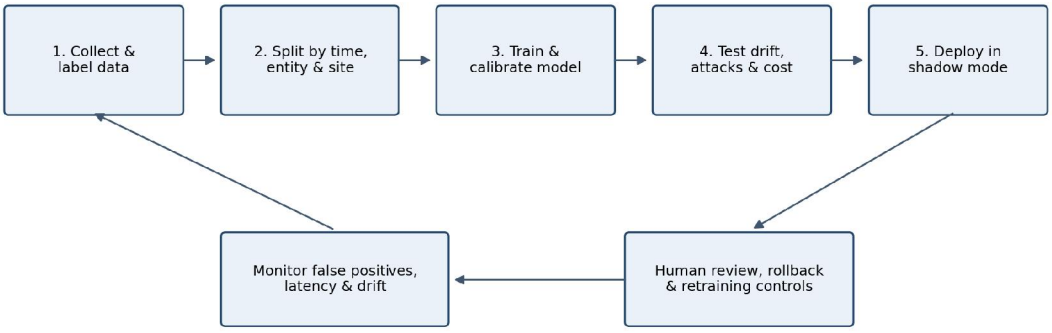

Figure 3 Proposed model validation and controlled deployment workflow

Source: Author's conceptual synthesis based on Sommer and Paxson (2010) and Cichonski et al. (2012).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

6.10 Integrated comparison of design choices.

Table 3 maps common deployment challenges to practical controls. The table emphasises that technical model choice is only one part of the solution: data governance, analyst workflow, and continuous validation are equally important. These recommendations synthesise the limitations discussed in the literature and institutional guidance (Buczak & Guven, 2016; NIST, 2024; Ring et al., 2019).

Table 3 Operational challenges and recommended controls

Challenge	Potential consequence	Recommended control
Noisy or incomplete telemetry	Missed context and unreliable baselines	Data-quality monitoring, source health checks, time synchronization
Class imbalance and label scarcity	Misleading accuracy and weak minority-class detection	Precision–recall analysis, careful sampling, analyst-validated labels
Concept drift	Rising false positives or missed attacks	Drift monitoring, governed retraining, champion–challenger tests
Adversarial evasion / poisoning	Manipulated scores or corrupted learning	Robustness testing, trusted data pipelines, multi-signal corroboration
Limited explainability	Low analyst trust and slow triage	Event-level evidence, feature attribution, model/version traceability
Privacy and proportionality	Excessive monitoring and data exposure	Purpose limitation, minimisation, retention controls, access governance
Unsafe automation	Disruption to legitimate operations	Shadow mode, approval gates, playbooks, rollback and audit trails

.

6.11 Discussion

The evidence suggests that behavioural analytics is most valuable when threats use legitimate identities or tools and therefore do not produce a clear signature. Its value increases when multiple weak signals are correlated across systems and interpreted in organisational context. Yet the same flexibility creates uncertainty: legitimate operational changes can resemble attacks, and malicious actors can imitate normal behaviour. The strongest design is consequently layered. Deterministic rules and known indicators handle well-understood threats; supervised models classify known patterns; anomaly detectors identify deviations; contextual engines prioritise risk; and analysts validate ambiguous cases. No single model should be expected to detect every threat.

The literature also supports a shift in evaluation emphasis. Rather than reporting only accuracy on a benchmark, researchers and practitioners should report the data provenance, label quality, split strategy, attack coverage, class distribution, threshold selection, precision, recall, false-positive rate, latency, resource use, and performance under drift. Where possible, they should include cross-environment validation and analyst workload. This evidence would make comparisons more meaningful and reduce the risk of deploying a model that is impressive in a laboratory but ineffective in a security operations centre.

VII. CONCLUSION

AI-enabled cyber threat detection using behavioural analytics offers a practical way to identify suspicious activity that may evade conventional signature-based controls. By learning patterns across identities, endpoints, applications, and networks, such systems can surface deviations and sequences that merit investigation. The literature reviewed in this paper demonstrates the promise of supervised learning, unsupervised anomaly detection, deep learning, and hybrid approaches, while also exposing important limitations involving data quality, operational generalisation, false positives, drift, adversarial threats, explainability, and privacy.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, November 2026 Impact Factor: 8.2

The paper proposed a conceptual architecture in which telemetry is normalised, enriched with organisational context, evaluated against behavioural baselines, and processed by AI models to generate prioritised alerts. It also proposed a validation workflow that emphasises chronological and entity-aware testing, shadow deployment, human review, and model governance. The analytical findings indicate that a trustworthy system must be judged by more than classification accuracy: it must demonstrate acceptable alert burden, timely detection, resilience to changing behaviour, interpretable evidence, and responsible handling of personal and operational data.

The study is limited by its narrative-review design and the absence of primary experimental validation. Future research should implement the proposed architecture in a controlled enterprise or cyber-range environment, compare model families under consistent data splits, assess cross-domain generalisation, and measure analyst workload and response outcomes. Research should also investigate privacy-preserving collaborative learning, robust drift adaptation, and explanations that are useful to security analysts. In conclusion, AI-enabled behavioural analytics is best understood as a decision-support capability within a layered cybersecurity programme: its effectiveness depends on sound data, contextual interpretation, disciplined evaluation, and accountable human governance.

REFERENCES

1. 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
2. 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
3. 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
4. Cichonski, P., Millar, T., Grance, T., & Scarfone, K. (2012). Computer security incident handling guide (NIST Special Publication 800-61 Rev. 2). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-61r2
5. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (pp. 4765–4774).
6. Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. In Proceedings of the Network and Distributed System Security Symposium (NDSS). https://doi.org/10.14722/ndss.2018.23204
7. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In 2015 Military Communications and Information Systems Conference (MilCIS) (pp. 1–6). IEEE. https://doi.org/10.1109/MilCIS.2015.7348942
8. National Institute of Standards and Technology. (2024). The NIST Cybersecurity Framework (CSF) 2.0 (NIST CSWP 29). https://doi.org/10.6028/NIST.CSWP.29
9. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778
10. 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
11. Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: A comprehensive review of AI-driven detection techniques. Journal of Big Data, 11, Article 105. https://doi.org/10.1186/s40537-024-00957-y
12. Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from a machine learning perspective. Journal of Big Data, 7, Article 41. https://doi.org/10.1186/s40537-020-00318-5

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, November 2026
Impact Factor: 8.2

13. Scarfone, K., & Mell, P. (2007). Guide to intrusion detection and prevention systems (IDPS) (NIST Special Publication 800-94). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-94
14. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP) (pp. 108–116). https://doi.org/10.5220/0006639801080116
15. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In 2010 IEEE Symposium on Security and Privacy (pp. 305–316). IEEE. https://doi.org/10.1109/SP.2010.25