MANTECH Publications

Journal of Converging Technologies in Intelligent
Volume 1, Issue 3, September-December 2025

# *Intelligent Cybersecurity Frameworks Using Deep Learning and Behavioral Anomaly Detection Techniques: A Systematic Review of Architectures, Benchmark Evidence, and Deployment Challenges*

***Ankita R. Sen[1], Subhajit D. Pal[2]***

***ABSTRACT***

*Modern enterprise networks, cloud platforms, and Internet of Things deployments generate traffic and event volumes that exceed the capacity of signature-based defences, while adversaries increasingly rely on zero-day exploits, credential abuse, and slow, low-volume intrusions that evade static rules. Deep learning (DL) and behavioral anomaly detection have therefore become central to intelligent cybersecurity research. This review-based study critically synthesises peer-reviewed literature published between 2015 and 2025 on convolutional, recurrent, autoencoder, Transformer, and graph-based detectors, together with user and entity behavior analytics (UEBA), and evaluates them against a five-dimension rubric covering known-attack detection, zero-day sensitivity, label independence, explainability, and deployment efficiency. A layered framework that fuses a supervised DL detector with an unsupervised behavioral baseline through a risk-fusion engine and an explainability layer is proposed. The synthesis shows that supervised models report near-saturated accuracy on random-split benchmarks but degrade markedly on harder test sets, that reconstruction-based and sequence-based models offer better sensitivity to unseen threats, and that hybrid architectures balance these trade-offs most effectively. Dataset quality, concept drift, adversarial robustness, and alert explainability emerge as the dominant barriers to operational adoption, and priorities for future research are outlined.*

***KEYWORDS:*** *Cybersecurity, Deep learning, Behavioral anomaly detection, Intrusion detection, UEBA, Autoencoder, Transformer, Explainable AI, Zero-day attacks, Network security*


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

## INTRODUCTION

The attack surface of contemporary digital infrastructure is expanding faster than the defences built to protect it. Cloud migration, remote work, containerised services, and billions of Internet of Things (IoT) endpoints generate continuous streams of network flows, authentication events, and system logs, while adversaries exploit this complexity through zero-day vulnerabilities, credential theft, and multi-stage campaigns that unfold over weeks. Conventional intrusion detection systems (IDS) depend on signatures and hand-written rules, which identify known threats with high precision but are structurally blind to novel attacks [1, 2]. Anomaly-based detection promises to flag previously unseen behaviour by modelling what is normal, yet early statistical implementations produced false-alarm rates that made them impractical in real networks [1, 3].

Deep learning (DL) has changed this landscape by learning hierarchical representations directly from raw or minimally engineered inputs, removing much of the manual feature design that constrained classical machine learning [4, 5]. Convolutional, recurrent, autoencoder, Transformer, and graph architectures now process packet payloads, flow statistics, log sequences, and communication graphs at scales that rule-based tools cannot match. In parallel, behavioral anomaly detection, including user and entity behavior analytics (UEBA), shifts the unit of analysis from the individual packet or signature to the long-term behavioral profile of users, hosts, and services. This shift makes it possible to detect insider misuse and compromised credentials, which often generate no malicious payload at all.

Despite rapid growth in publications, the field lacks a consolidated view of which DL and behavioral techniques remain effective once evaluation moves beyond curated laboratory benchmarks. Reported accuracies above 99% coexist with practitioner experience of alert fatigue, model decay, and brittle behaviour under distribution shift [6, 7]. The problem addressed in this paper is therefore stated as follows: how can deep learning and behavioral anomaly detection be combined into a single framework that detects unknown threats reliably, produces explainable alerts, and remains deployable under realistic data, latency, and resource constraints? The paper reviews the evidence, evaluates competing approaches against a common rubric, and proposes a layered framework that reconciles their strengths.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

## LITERATURE REVIEW

Research on learned intrusion detection spans more than two decades. The surveyed work is organised into four strands: the transition from rule-based to learned detectors, deep architectures for threat detection, behavioral anomaly modelling, and the datasets and evaluation practices that govern how progress is measured.

### From Signature Rules To Learned Detectors

Early surveys documented the adoption of data mining and machine learning for intrusion detection and identified recurring obstacles, including scarce labelled data, high false-positive rates, and limited generalisation [4]. Much early work relied on the KDD Cup 1999 dataset, whose analysis later revealed that roughly 78% of training records and 75% of test records were duplicates, biasing classifiers towards frequent records and inflating reported accuracy [8]. Deep models were subsequently applied to the cleaner NSL-KDD derivative. Kim et al. showed that an LSTM-based classifier could exploit the sequential character of connection records [9], and Yin et al. reported roughly 83% binary accuracy for a recurrent network on the demanding KDDTest+ subset [10]. Shone et al. combined a stacked non-symmetric deep autoencoder with a random forest and obtained about 85% accuracy in five-class classification on the same subset [11]. Vinayakumar et al. benchmarked deep networks across several datasets and reported advantages over classical learners [12], while Ferrag et al. compared a range of DL approaches on contemporary datasets and highlighted the gap between benchmark performance and deployment readiness [5].

### Deep Learning Architectures For Threat Detection

Convolutional neural networks (CNNs) capture local patterns in traffic features or payload bytes and are computationally efficient. Recurrent networks, particularly long short-term memory (LSTM) units [13], model temporal dependencies in event streams. Autoencoders learn compact representations of normal data and flag inputs with high reconstruction error; Kitsune extended this idea to an online ensemble of lightweight autoencoders that operates without labelled attacks and can run on low-cost hardware [14]. DeepLog applied LSTM next-event prediction to system logs and flagged deviations from learned execution patterns, reporting an F-measure of about 96% on HDFS logs [15]. Transformer architectures replace recurrence with self-attention [16], capturing long-range dependencies across events at the cost of heavier computation, while graph neural networks build on graph convolution [17] to


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

represent hosts and connections as nodes and edges, which suits the detection of lateral movement. Table 1 summarises these architectures and Figure 2 organises them into a taxonomy.

***Table 1. Deep Learning Architectures Reviewed for Cyber-Threat Detection***

| Architecture | Typical Input | Principal Strength | Principal Limitation |
|---|---|---|---|
| CNN [5] | Flow feature vectors, payload bytes | Efficient local pattern extraction | Weak temporal modelling; label dependent |
| LSTM / GRU [13, 15] | Event and log sequences | Models temporal dependencies | Sequential computation; long histories needed |
| Autoencoder / VAE [14] | Normal-only traffic or logs | Learns normality without attack labels | Threshold sensitive; may reconstruct attacks |
| Transformer [16] | Long event sequences | Captures long-range dependencies | High compute and memory cost |
| Graph neural network [17] | Host–connection graphs | Models lateral movement | Graph construction overhead |

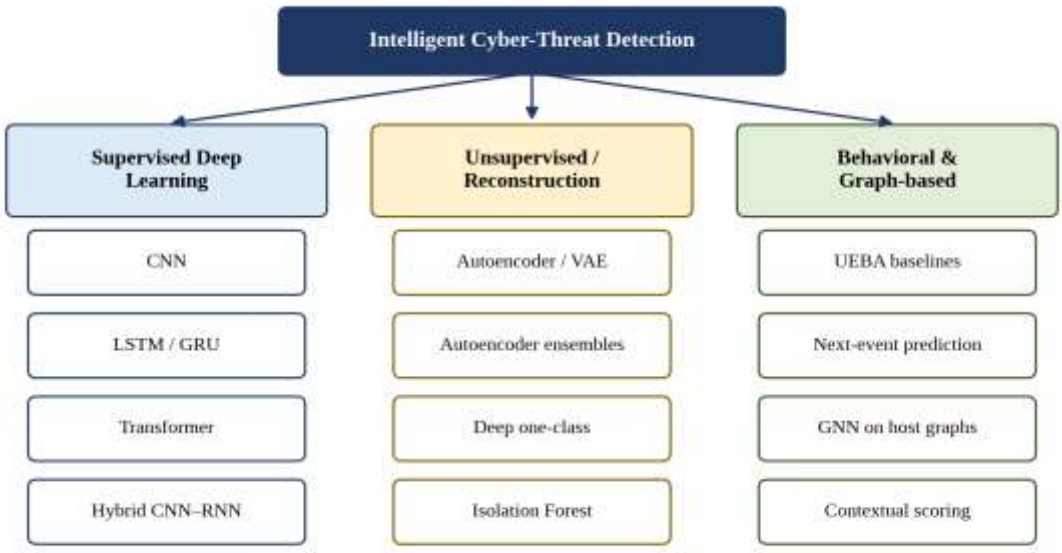



***Figure 2. Taxonomy of deep learning and behavioral anomaly detection techniques for cyber-threat detection reviewed in this study.***


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

### Behavioral Anomaly Detection

Anomalies are conventionally classified as point, contextual, or collective, and the contextual category is central to cybersecurity because an action may be normal for one user or time window and malicious for another [2]. Classical unsupervised detectors such as Isolation Forest isolate anomalies through random partitioning without requiring labels [18], whereas deep one-class objectives learn a compact description of normal data [19]. Surveys of deep anomaly detection catalogue reconstruction-, prediction-, and distance-based approaches and stress that performance depends heavily on how well the training data represent normal behaviour [20, 21]. UEBA systems apply these ideas to per-user and per-host baselines, scoring deviations in login times, data volumes, resource access, and command sequences, and are an established approach to insider-threat and credential-abuse detection.

### Datasets And Evaluation Practice

Dataset choice strongly conditions reported results. Ring et al. surveyed network-based intrusion datasets and documented wide variation in realism, labelling, and attack coverage [7]. CICIDS2017 [22] and UNSW-NB15 [23] replaced the obsolete KDD family with modern traffic and attack types, yet later analysis of CICIDS2017 identified labelling errors and flow-construction artefacts that affect classifier results [24]. Arp et al. showed that pitfalls such as sampling bias, data snooping, and unrealistic laboratory settings pervade security-oriented machine learning and can make published results unreliable in practice [6].

## RESEARCH GAP

Three gaps emerge from the literature. First, most studies evaluate a single architecture on a single benchmark with random train-test splits, so comparisons across architectures and datasets under a common rubric are scarce [5, 7]. Second, supervised DL detectors and behavioral anomaly engines are typically developed in isolation, leaving principled fusion of their scores largely unaddressed. Third, explainability, adversarial robustness, and the base-rate problem [3] are treated as peripheral concerns rather than design requirements. No consolidated framework currently links architecture choice to these operational criteria.

## OBJECTIVES

- To critically review the deep learning architectures (CNN, recurrent, autoencoder, Transformer, and graph-based) used for cyber-threat and anomaly detection.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

- To analyse behavioral anomaly detection and UEBA techniques and their complementarity with supervised DL detectors.
- To compare the paradigms against a five-dimension rubric covering known-attack detection, zero-day sensitivity, label independence, explainability, and deployment efficiency.
- To propose a layered intelligent cybersecurity framework that fuses supervised and behavioral evidence with explainable outputs.
- To identify limitations, deployment barriers, and priorities for future research.

## METHODOLOGY

### Review Design And Search Strategy

This study follows a review-based analytical design guided by the evidence-synthesis principles of Kitchenham and Charters [25] and the transparent-reporting principles of PRISMA 2020 [26]. Literature was sought in IEEE Xplore, the ACM Digital Library, ScienceDirect, Scopus, and SpringerLink for the period January 2015 to December 2025, with foundational earlier work retained for context. Search strings combined DL terms (deep learning, neural network, autoencoder, LSTM, Transformer, graph neural network) with security terms (intrusion detection, anomaly detection, user behavior analytics, cyber security). Studies were included if they were peer reviewed, described the detection architecture reproducibly, and reported evaluation on a public benchmark or a documented operational dataset. Studies were excluded if they were purely theoretical, lacked experimental evaluation, were not available in English, or duplicated earlier work. Citation chaining from key surveys supplemented database searching.

### Analytical Evaluation Framework

Each architectural family was assessed against five dimensions chosen to reflect operational rather than purely statistical performance: D1, detection of known attack classes; D2, sensitivity to previously unseen (zero-day) behaviour; D3, label independence, where a higher score indicates less reliance on annotated attack data; D4, explainability of individual alerts; and D5, deployment efficiency in terms of computation, memory, and latency. Scores from 1 (weak) to 5 (strong) were assigned through structured qualitative judgement of the reported evidence. They are comparative indicators derived from the literature and not measured


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

experimental values. Benchmark figures quoted in the text are those reported by the original authors under their own protocols and are therefore not directly comparable across studies.

**Proposed Layered Framework**

Informed by the review, a five-layer framework is proposed (Figure 1). The data acquisition layer ingests network flows, endpoint telemetry, authentication logs, and cloud audit events. The preprocessing layer performs normalisation, sliding-window sequence construction, categorical encoding, and entity-graph assembly. The detection layer runs two complementary engines in parallel: a supervised detector, for example a CNN–BiLSTM or Transformer encoder, that recognises known attack patterns, and an unsupervised behavioral engine that scores deviations from per-entity baselines using autoencoder reconstruction error. For an entity e at time t, the behavioral anomaly score is standardised as $z(e,t) = (\varepsilon(e,t) - \mu_e) / \sigma_e$, where $\varepsilon$ is the reconstruction error and $\mu_e$ and $\sigma_e$ are the entity's running mean and standard deviation.

The risk-fusion layer combines the evidence as $R(e,t) = \alpha \cdot p(e,t) + \beta \cdot S(z(e,t)) + \gamma \cdot c(e)$, where p is the calibrated supervised attack probability, S is the logistic function, c(e) encodes asset criticality, and $\alpha$, $\beta$, $\gamma$ are non-negative weights summing to one that are tuned on validation data. Finally, the explainability and response layer attaches feature attributions to each alert and routes high-risk events to analysts or automated playbooks, whose confirmations feed a continuous retraining loop with drift monitoring.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

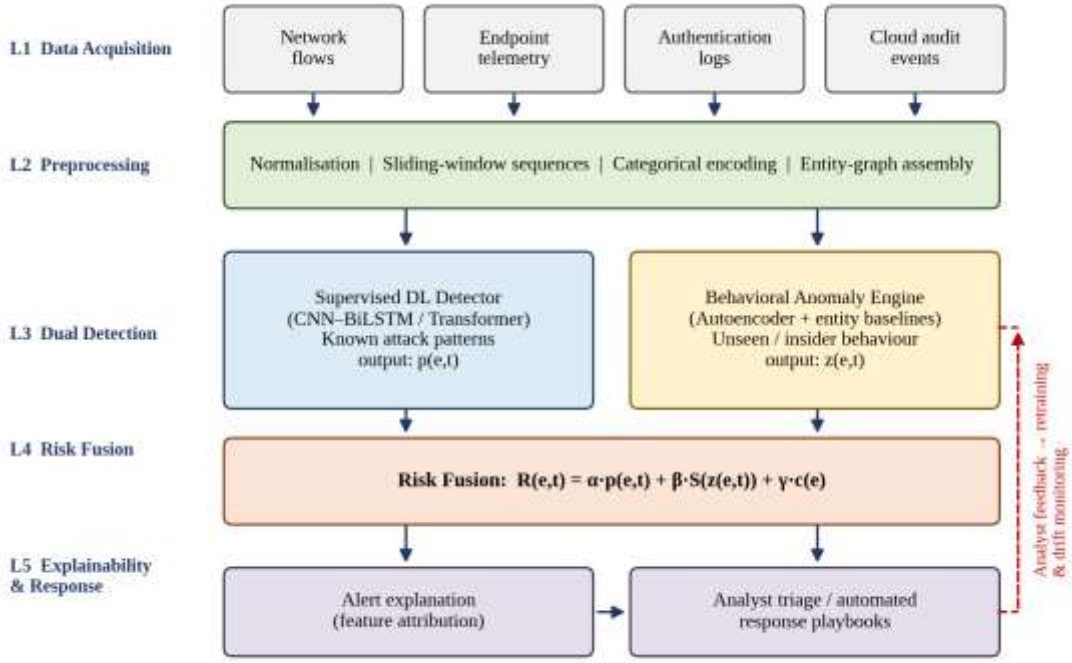



***Figure 1. Proposed five-layer intelligent cybersecurity framework combining a supervised deep learning detector and an unsupervised behavioral anomaly engine through a risk-fusion layer, with explainability and analyst feedback.***

## RESULTS AND FINDINGS

The synthesis yields four principal findings, supported by Tables 2 and 3 and Figure 3.

### Benchmark Saturation And The Generalisation Gap

Table 2 collects representative evidence. On modern random-split benchmarks, supervised DL detectors routinely report accuracies above 95%, which has led several authors to describe these datasets as close to saturated [5, 12]. In contrast, models evaluated on the deliberately difficult KDDTest+ subset reach only about 83–85% [10, 11], showing how sharply performance falls when test records differ in distribution from training records. This contrast, together with the duplication and labelling problems documented in KDD and CICIDS2017 [8, 24], indicates that headline accuracy overstates operational capability.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

***Table 2. Representative Benchmark Evidence Reported in the Reviewed Literature***

| Study | Data | Technique | Reported Outcome |
|---|---|---|---|
| Yin et al. [10] | NSL-KDD (KDDTest+) | Recurrent neural network | ≈83% accuracy (binary) |
| Shone et al. [11] | NSL-KDD (KDDTest+) | Deep autoencoder + random forest | ≈85% accuracy (five-class) |
| Du et al. [15] | HDFS system logs | LSTM next-event prediction | ≈96% F-measure |
| Mirsky et al. [14] | Network attack captures | Autoencoder ensemble | Unsupervised online detection on low-cost hardware |
| Vinayakumar et al. [12] | Several public IDS datasets | Deep neural networks | Reported gains over classical learners |

**Zero-Day Sensitivity Favours Reconstruction And Sequence Models**

Autoencoder ensembles and next-event prediction models detect attacks without needing examples of them, because they flag departures from learned normality [14, 15]. Supervised CNN and Transformer detectors are most accurate on classes seen in training but are inherently limited by the attack coverage of their labelled data. Table 3 captures this pattern: autoencoders score highest on label independence (5) and strongly on zero-day sensitivity (4), whereas CNNs and Transformers score highest on known-attack detection (5).

***Table 3. Rubric-Based Comparison of Detection Paradigms (Scores 1–5 on Dimensions D1–D5)***

| Paradigm | D1 | D2 | D3 | D4 | D5 | Mean |
|---|---|---|---|---|---|---|
| CNN | 5 | 2 | 1 | 2 | 4 | 2.8 |
| LSTM / GRU | 4 | 3 | 2 | 3 | 3 | 3.0 |
| Autoencoder | 3 | 4 | 5 | 3 | 4 | 3.8 |
| Transformer | 5 | 3 | 2 | 3 | 2 | 3.0 |
| GNN | 4 | 4 | 3 | 2 | 2 | 3.0 |
| Hybrid framework | 5 | 4 | 4 | 4 | 3 | 4.0 |


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

**No Single Paradigm Dominates**

Figure 3 plots the rubric scores. Each single-paradigm detector shows at least one weak dimension: CNNs depend on labels, Transformers carry heavy computational cost, GNNs require graph-construction pipelines, and autoencoders trade peak accuracy on known attacks for flexibility. The proposed hybrid framework attains the most balanced profile, with a mean score of 4.0 compared with 2.8–3.8 for the single-paradigm detectors. The autoencoder ranks second at 3.8 owing to its label independence and efficiency but is weaker on known-attack detection. The hybrid scores are a design-level assessment and not an implemented result.

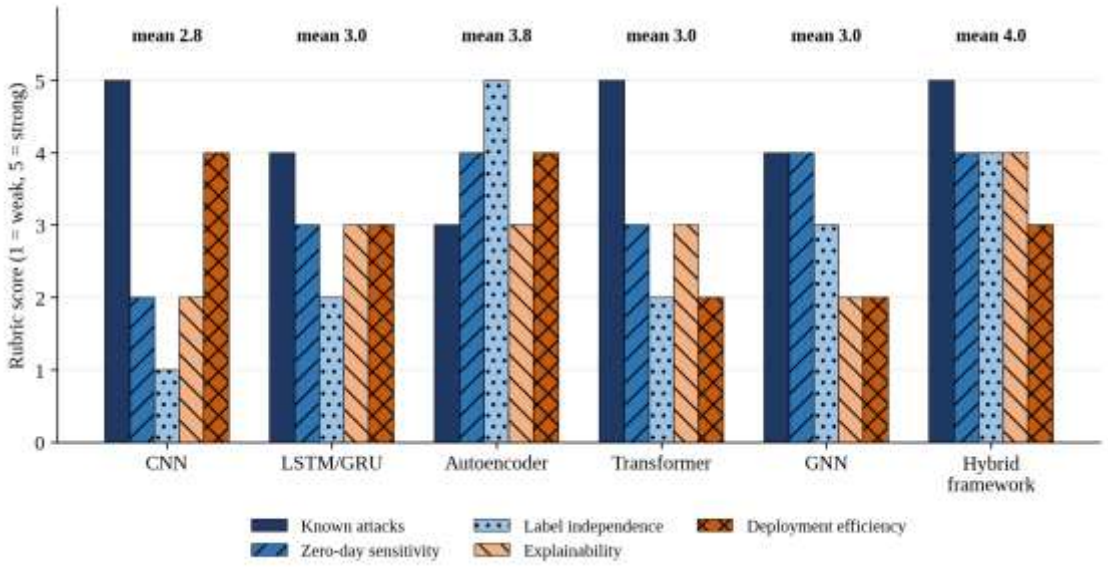



***Figure 3. Rubric scores (1 = weak, 5 = strong) of detection paradigms across five operational dimensions; scores are qualitative review judgements, not measured values.***

**Operational Barriers Recur Across Studies**

Four barriers recur in the reviewed studies. First, the base-rate effect: because malicious events are rare, even a small false-positive rate produces an overwhelming number of false alerts, so very low false-alarm rates, on the order of one in a hundred thousand events, are essential [3]. Second, concept drift: behavioral baselines and traffic distributions shift as users, software, and threats change, degrading static models. Third, adversarial vulnerability: small crafted perturbations can cause misclassification in neural networks [27], and attackers can shape traffic to evade learned detectors. Fourth, opacity: analysts need to know why an alert fired, and post-hoc attribution methods such as SHAP can provide feature-level explanations [28].


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

## DISCUSSION

The results indicate that architecture selection should be driven by the threat model and operating constraints rather than by benchmark accuracy alone. Supervised DL detectors suit high-volume perimeter and cloud-gateway traffic where known attack families dominate and labelled data exist, whereas reconstruction-based and sequence-based behavioral engines suit environments where novel or insider threats are the main concern and labelled attacks are rare. The proposed framework formalises this complementarity: the supervised engine contributes precision on known threats, the behavioral engine contributes coverage of the unknown, and the fusion layer converts both into a single prioritised risk score. Scientifically, the review reframes progress in the field from raising benchmark accuracy to improving robustness, explainability, and base-rate-aware evaluation.

The practical applications are broad. In security operations centres, fused risk scores with attribution can reduce triage burden and support prioritisation. In banking and financial services, entity-level baselines can reveal account takeover and insider fraud. In critical infrastructure and smart-grid environments, lightweight autoencoders can run at the edge where labelled attacks are almost nonexistent. In cloud and zero-trust deployments, continuous behavioral risk scoring can inform per-request access decisions in line with NIST zero-trust architecture guidance [29]. In IoT networks, ensembles of small autoencoders suit resource-constrained gateways [14].

Several tensions deserve emphasis. Higher model capacity improves known-attack detection but increases latency and weakens explainability, so Transformer-scale models may be better placed in offline threat hunting than in inline blocking. Privacy regulation and organisational boundaries restrict the pooling of security telemetry, which makes federated training attractive [30], although it introduces its own poisoning risks. Finally, evaluation protocols should follow the guidance of Arp et al., using temporally realistic splits, cross-dataset testing, and explicit false-alarm budgets [6].

## LIMITATIONS

- The review is restricted to English-language, peer-reviewed publications from selected databases, and publication bias towards positive results may overstate the performance of DL detectors.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

- Benchmark figures are reported under heterogeneous protocols, datasets, and preprocessing choices, so they cannot be compared directly or pooled statistically.
- Rubric scores are structured qualitative judgements and not measured quantities, and the proposed framework is evaluated conceptually; it has not yet been implemented and tested on operational traffic.
- Public datasets age quickly and contain known labelling artefacts, which limits how far the findings generalise to modern enterprise environments.

## FUTURE SCOPE

- Implementation and empirical validation of the proposed framework on temporally split, multi-dataset benchmarks and on operational telemetry, with explicit false-alarm budgets.
- Adversarial robustness research, including adversarial training and certified defences tailored to traffic and log features.
- Federated and privacy-preserving training across organisations, with defences against model poisoning.
- Continual and drift-aware learning that updates behavioral baselines without absorbing attacker activity as normal.
- Human-centred evaluation of explanation quality with security analysts, and cautious integration of large language models for alert summarisation and investigation support.

## CONCLUSION

Deep learning and behavioral anomaly detection are complementary rather than competing foundations for intelligent cybersecurity. This review found that supervised DL detectors deliver the strongest accuracy on known attack classes but generalise poorly beyond their training distribution, that reconstruction-based and sequence-based behavioral models provide the most useful sensitivity to unseen threats, and that no single paradigm satisfies all five operational dimensions. A layered framework that fuses supervised and behavioral evidence, calibrates risk against asset criticality, and exposes explanations to analysts offers the most balanced profile. Realising this potential depends less on new architectures than on better datasets, base-rate-aware evaluation, adversarial robustness, and drift management, and closing these gaps is the central priority for the next generation of intelligent cybersecurity research.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

**REFERENCES**

1. Sommer R, Paxson V. Outside the closed world: On using machine learning for network intrusion detection. *Proceedings of the IEEE Symposium on Security and Privacy.* 2010; 305–316.
2. Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. *ACM Computing Surveys.* 2009; 41(3): 1–58.
3. Axelsson S. The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security.* 2000; 3(3): 186–205.
4. Buczak AL, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials.* 2016; 18(2): 1153–1176.
5. Ferrag MA, Maglaras L, Moschoyiannis S, Janicke H. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications.* 2020; 50: 102419.
6. Arp D, Quiring E, Pendlebury F, et al. Dos and don'ts of machine learning in computer security. *Proceedings of the 31st USENIX Security Symposium.* 2022; 3971–3988.
7. Ring M, Wunderlich S, Scheuring D, Landes D, Hotho A. A survey of network-based intrusion detection data sets. *Computers & Security.* 2019; 86: 147–167.
8. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the KDD CUP 99 data set. *Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications.* 2009; 1–6.
9. Kim J, Kim J, Thu HLT, Kim H. Long short term memory recurrent neural network classifier for intrusion detection. *Proceedings of the International Conference on Platform Technology and Service.* 2016; 1–5.
10. Yin C, Zhu Y, Fei J, He X. A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access.* 2017; 5: 21954–21961.
11. Shone N, Ngoc TN, Phai VD, Shi Q. A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence.* 2018; 2(1): 41–50.
12. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Al-Nemrat A, Venkatraman S. Deep learning approach for intelligent intrusion detection system. *IEEE Access.* 2019; 7: 41525–41550.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

13. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation.* 1997; 9(8): 1735–1780.
14. Mirsky Y, Doitshman T, Elovici Y, Shabtai A. Kitsune: An ensemble of autoencoders for online network intrusion detection. *Proceedings of the Network and Distributed System Security Symposium.* 2018.
15. Du M, Li F, Zheng G, Srikumar V. DeepLog: Anomaly detection and diagnosis from system logs through deep learning. *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security.* 2017; 1285–1298.
16. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in Neural Information Processing Systems.* 2017; 30: 5998–6008.
17. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *Proceedings of the International Conference on Learning Representations.* 2017.
18. Liu FT, Ting KM, Zhou ZH. Isolation forest. *Proceedings of the IEEE International Conference on Data Mining.* 2008; 413–422.
19. Ruff L, Vandermeulen R, Goernitz N, et al. Deep one-class classification. *Proceedings of the 35th International Conference on Machine Learning.* 2018; 4393–4402.
20. Pang G, Shen C, Cao L, van den Hengel A. Deep learning for anomaly detection: A review. *ACM Computing Surveys.* 2021; 54(2): 1–38.
21. Chalapathy R, Chawla S. Deep learning for anomaly detection: A survey. *arXiv preprint* arXiv:1901.03407. 2019.
22. Sharafaldin I, Lashkari AH, Ghorbani AA. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *Proceedings of the 4th International Conference on Information Systems Security and Privacy.* 2018; 108–116.
23. Moustafa N, Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems. *Proceedings of the Military Communications and Information Systems Conference.* 2015; 1–6.
24. Engelen G, Rimmer V, Joosen W. Troubleshooting an intrusion detection dataset: The CICIDS2017 case study. *Proceedings of the IEEE Security and Privacy Workshops.* 2021; 7–12.
25. Kitchenham B, Charters S. *Guidelines for performing Systematic Literature Reviews in Software Engineering (EBSE Technical Report).* Keele University and University of Durham. 2007.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved

26. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021; 372: n71.
27. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. *Proceedings of the International Conference on Learning Representations.* 2015.
28. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems.* 2017; 30: 4765–4774.
29. Rose S, Borchert O, Mitchell S, Connelly S. *Zero Trust Architecture (NIST Special Publication 800-207).* National Institute of Standards and Technology, Gaithersburg. 2020.
30. McMahan HB, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics.* 2017; 1273–1282.

***Author for Correspondence:**
**Ankita R. Sen**
**E-mail:** ankita.sen.cse@gmail.com

[1]Assistant Professor, Department of Computer Science and Engineering, Haldia Institute of Technology, West Bengal
[2]Assistant Professor, Department of Computer Science and Engineering, Haldia Institute of Technology, West Bengal

Received Date: September 04, 2025
Accepted Date: September 19, 2025
Published Date: September 30, 2025

**Citation:** Ankita R. Sen, Subhajit D. Pal. Intelligent Cybersecurity Frameworks Using Deep Learning and Behavioral Anomaly Detection Techniques: A Systematic Review of Architectures, Benchmark Evidence, and Deployment Challenges. Journal of Converging Technologies in Intelligent. 2025; 1(3): 147-161.


Page 147-161 © MANTECH PUBLICATIONS 2025. All Rights Reserved