IJRRETAS

INTERNATIONAL JOURNAL FOR RAPID RESEARCH

IN ENGINEERING TECHNOLOGY & APPLIED SCIENCE

Volume: 11

Issue: 9

Month of publication: September 2025

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 11 Issue 9, September, 2025 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

AI-Driven Computer Network Security Using Neural Network-Based Defence Systems for Threat Detection and Future Security Enhancement

[1]Deepak Tiwari, [2]Dr. Vijay Singh

[1]Research Scholar, Department of Computer Science, Amaltas University, Dewas

[2]Supervisor, Department of Computer Science, Amaltas University, Dewas

Abstract

Computer networks have become the connective tissue of modern economic, industrial, and social activity, and the same expansion that has made them indispensable has also widened the surface available to adversaries. Signature-driven firewalls, static access control lists, and rule-based intrusion detection engines, which formed the backbone of network defence for nearly three decades, are increasingly unable to keep pace with polymorphic malware, encrypted command-and-control channels, distributed botnets, and attacks generated or optimised by machine learning itself. The paper synthesises evidence from surveys, benchmark studies, and architecture-specific investigations covering multilayer perceptrons, deep belief networks, stacked and non-symmetric autoencoders, convolutional neural networks, recurrent and bidirectional long short-term memory models, and deep reinforcement learning agents. Three comparative tables consolidate the benchmark datasets that underpin the field, the reported behaviour of the principal neural architectures, and the adversarial threats that complicate their deployment. The review finds that deep architectures consistently outperform shallow classifiers on curated benchmarks, that representation learning has largely displaced manual feature engineering, and that autoencoder-based and recurrent models are particularly effective for unsupervised and sequence-sensitive detection. It also finds a persistent gap between laboratory accuracy and operational reliability, driven by dataset obsolescence, class imbalance, concept drift, adversarial evasion, opacity of decision-making, and computational cost. The paper concludes with a discussion of federated learning, adversarially robust training, explainable security analytics, and autonomous response as the most consequential directions for future research.

Keywords: artificial intelligence; neural networks; deep learning; network security; intrusion detection; anomaly detection; adversarial machine learning; cyber defence

1. Introduction

1.1 The Changing Character of Network Threats

The threat landscape confronting contemporary computer networks bears little resemblance to the one that shaped the first generation of defensive technology. Early network attacks were largely opportunistic, reused across targets, and expressed through recognisable byte sequences that could be catalogued and matched. The modern equivalent is adaptive, financially or geopolitically motivated, and frequently designed to blend into ordinary traffic. Adversaries employ polymorphic and metamorphic code that mutates between infections, they tunnel command-and-control traffic through encrypted and otherwise legitimate protocols, and they

stage intrusions across weeks or months in ways that defeat any detector reasoning about single packets or single sessions.

Scale compounds the difficulty. Enterprise and carrier networks now generate volumes of telemetry that exceed what any human analyst team can inspect, and the growth of cloud-native infrastructure, software-defined networking, and mobile access has dissolved the perimeter that older architectures assumed. The Internet of Things has been particularly consequential: billions of constrained devices, often shipped with default credentials and rarely patched, have been absorbed into networks that were never designed to accommodate them, and the resulting security challenges cut across the perception, network, and application layers simultaneously (Al-Garadi et al., 2020). Botnets assembled from such devices have demonstrated that the aggregate capacity of weak endpoints can exceed that of well-defended servers.

A further shift is that attackers have themselves become consumers of machine learning. Automated vulnerability discovery, generative production of phishing content, and the deliberate crafting of inputs that mislead classifiers have moved from research demonstrations to practical tooling. The consequence is a defensive environment in which detection systems must not only recognise threats but must also remain reliable while under active attempt at manipulation (Biggio & Roli, 2018).

1.2 The Limits of Conventional Defence Mechanisms

Traditional network defence rests on two complementary strategies. Signature-based or misuse detection compares observed activity against a curated library of known malicious patterns, and it remains highly accurate and inexpensive for threats that have already been catalogued. Its weakness is structural rather than incidental: a signature cannot exist before the attack it describes, so zero-day exploits and novel variants pass unimpeded, and the maintenance burden of the signature library grows without bound as the catalogue expands (Buczak & Guven, 2016).

Anomaly-based detection attempts to address this by modelling normal behaviour and flagging deviations from it. In principle this permits the discovery of previously unseen attacks; in practice, early statistical and threshold-based implementations produced false positive rates that operational teams found intolerable. Network traffic is non-stationary, legitimate behaviour changes continuously as applications and usage patterns evolve, and a static model of normality degrades quickly. The result has been a well-documented reluctance to deploy anomaly detectors in production environments despite their theoretical advantages (Apruzzese et al., 2018).

Classical machine learning offered a partial remedy. Support vector machines, decision trees, random forests, naive Bayes classifiers, and k-nearest-neighbour methods were applied extensively to intrusion detection and demonstrated that data-driven models could generalise beyond fixed signatures. Their principal constraint is dependence on manual feature engineering. The quality of a shallow classifier is bounded by the quality of the features supplied to it, and constructing those features requires domain expertise that must be renewed whenever protocols, applications, or attack techniques change. Shallow models also struggle

with the high dimensionality and severe class imbalance characteristic of network traffic, where malicious samples may constitute a negligible fraction of total observations (Buczak & Guven, 2016; Ferrag et al., 2020).

1.3 Neural Networks as the Computational Core of Modern Defence

Neural networks address the feature engineering bottleneck directly. By composing successive layers of non-linear transformations, a deep network learns a hierarchy of representations from raw or lightly processed input, discovering discriminative structure that a human designer might not anticipate. Applied to network security, this capability allows models to operate on packet payloads, flow records, and system call sequences without an intervening stage of hand-crafted descriptors, and to capture interactions among features that shallow methods treat as independent (Javaid et al., 2016; Shone et al., 2018).

Different architectures contribute distinct inductive biases. Deep feedforward networks and deep belief networks provide general-purpose hierarchical classification. Autoencoders learn compressed representations of normal behaviour and reconstruct inputs, making reconstruction error a natural anomaly signal that requires no labelled attack data (Mirsky et al., 2018). Convolutional networks exploit local spatial structure, which transfers effectively to byte-level representations of traffic and malware. Recurrent networks and their gated variants model temporal dependency, which matters for attacks that manifest as a sequence of individually unremarkable events (Kim et al., 2016; Yin et al., 2017). Deep reinforcement learning extends the paradigm from classification to sequential decision-making, allowing an agent to learn defensive policies through interaction with an adversarial environment (Nguyen & Reddi, 2023).

The empirical record accumulated since 2015 is broadly favourable. Across the standard benchmarks, deep architectures have repeatedly matched or exceeded the accuracy of tuned shallow classifiers, with the margin widest on multi-class problems and on rare attack categories where representation learning appears to help most (Vinayakumar et al., 2019). Whether that advantage survives the transition from benchmark to production is the central open question of the field.

2. Literature Review

2.1 Data-Driven Foundations and the Benchmark Problem

Any assessment of learning-based network defence must begin with the data on which it is evaluated, because the field's conclusions are conditioned on a small number of shared benchmarks whose limitations propagate into every result derived from them.

The comprehensive survey by Buczak and Guven (2016) established the reference framing for data-driven intrusion detection. Reviewing methods spanning association rules, clustering, Bayesian networks, decision trees, support vector machines, and neural networks, the authors documented the trade-offs among accuracy, computational cost, and interpretability, and noted the difficulty of comparing published results when studies use different datasets, preprocessing pipelines, and evaluation metrics. That comparability problem has not been resolved in the intervening decade.

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 11 Issue 9, September, 2025 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

The dominance of the KDD Cup 1999 dataset and its cleaned derivative NSL-KDD is the most visible symptom. These datasets were generated from network conditions of the late 1990s, contain attack categories that no longer reflect contemporary practice, and exhibit redundancy and distributional artefacts that inflate reported performance. Recognising this, Moustafa and Slay (2015) constructed UNSW-NB15 by combining real modern benign traffic with synthesised contemporary attacks, producing forty-nine features across nine attack families and a substantially harder classification problem. Sharafaldin et al. (2018) subsequently released CIC-IDS2017, generated from profiled user behaviour over a realistic network topology and captured with full packet payloads, covering brute force, denial of service, distributed denial of service, web attacks, infiltration, and botnet activity.

Ring et al. (2019) surveyed the full landscape of network-based intrusion detection datasets and proposed evaluation criteria spanning recording environment, traffic realism, labelling accuracy, feature representation, and public availability. Their analysis makes clear that no single dataset satisfies all criteria and that many widely used resources fail several. Arp et al. (2022) sharpened the critique into a catalogue of recurring methodological errors in security machine learning, including sampling bias, spurious correlations, inappropriate temporal splits that leak future information into training, and evaluation on datasets whose class balance bears no relation to operational reality. Their central finding is that a substantial fraction of published performance improvements are artefacts of experimental design rather than genuine advances.

Table 1. Principal benchmark datasets used in neural network-based network security research

Dataset	Year	Features	Attack Categories	Principal Strengths	Principal Limitations
KDD Cup 1999	1999	41	4 (DoS, Probe, R2L, U2R)	Historical baseline; widely cited	Obsolete traffic; heavy record redundancy
NSL-KDD	2009	41	4	Redundancy removed; balanced splits	Underlying traffic still from the 1990s
UNSW-NB15	2015	49	9	Modern hybrid traffic; richer feature set	Partly synthetic; limited protocol diversity
CIC-IDS2017	2017	80+	7	Realistic topology; full packet capture	Large volume; severe class imbalance
Kitsune / Mirai	2018	Incremental	IoT reconnaissance and botnet	Real IoT capture; supports online evaluation	Narrow device and attack coverage

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 11 Issue 9, September, 2025 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

N-BaIoT / IoT captures	2018–2020	Varies	IoT botnet families	Device-level realism	Fragmented; inconsistent labelling

The consequence for the remainder of this review is interpretive. Reported accuracies should be read as evidence of relative architectural merit under controlled conditions, not as predictions of field performance.

2.2 Deep Neural Architectures for Threat Detection

The first substantial demonstrations of deep learning for intrusion detection appeared in the middle of the review period. Javaid et al. (2016) applied a self-taught learning approach in which a sparse autoencoder was trained on unlabelled traffic to learn a compact representation, after which a softmax classifier was fitted on a smaller labelled set. The design directly addressed the scarcity of labelled attack data, and evaluation on NSL-KDD showed that the learned representation outperformed the raw feature vector.

Recurrent architectures were introduced to capture the temporal structure that feedforward models discard. Kim et al. (2016) trained a long short-term memory classifier on sequences of connection records and demonstrated that gated recurrence could retain information across the extended intervals characteristic of multi-stage intrusions. Yin et al. (2017) provided a more systematic treatment, evaluating recurrent networks in both binary and multi-class settings and examining the effect of hidden unit count and learning rate. Their results showed clear gains over shallow classifiers, with the improvement concentrated in the low-frequency remote-to-local and user-to-root categories that had historically resisted detection.

Shone et al. (2018) introduced a non-symmetric deep autoencoder, using a stacked encoder without a matching decoder to reduce training cost, and combined the learned representation with a random forest classifier. The hybrid achieved strong accuracy on KDD Cup 1999 and NSL-KDD while substantially reducing training time relative to fully symmetric alternatives. Khan et al. (2019) proposed a two-stage deep learning model in which a stacked autoencoder with a softmax classifier first performs a coarse probability estimate, which is then appended as an additional feature for a second-stage classifier. The staged design improved detection of low-frequency attack classes on both KDD Cup 1999 and UNSW-NB15.

The most extensive comparative study of the period is that of Vinayakumar et al. (2019), who evaluated deep neural networks against a range of classical classifiers across multiple datasets and both host-based and network-based settings, and proposed a scalable hybrid framework capable of monitoring host and network telemetry in parallel. Their central finding was that deep architectures generalise more reliably across datasets than shallow models tuned to a single benchmark. Ferrag et al. (2020) complemented this with a systematic comparison of seven deep learning models, including deep neural networks, restricted Boltzmann machines, deep belief networks, convolutional networks, recurrent networks, deep Boltzmann machines, and deep autoencoders, evaluated under a common protocol on CSE-CIC-IDS2018 and the Bot-IoT dataset. The uniform experimental treatment makes their comparison one of the few in which architectural differences can be attributed with reasonable confidence.

International Journal for Rapid Research in Engineering Technology & Applied Science (IJRRETAS)

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume 11 Issue 9, September, 2025 OPEN ACCESS E-ISSN: 2455-4723 Impact Factor 7.95

Bidirectional processing was shown to yield further gains. Imrana et al. (2021) applied a bidirectional LSTM to NSL-KDD, allowing the model to condition on both preceding and following context within a sequence window. The bidirectional variant outperformed the unidirectional baseline and reduced the false alarm rate, with the largest improvements again occurring on the rare attack categories.

Table 2. Comparison of neural architectures applied to network threat detection

Architecture	Learning Mode	Representative Studies	Principal Strength	Principal Weakness
Deep neural network (MLP)	Supervised	Vinayakumar et al. (2019)	Simple, fast inference, strong baseline	Ignores temporal and spatial structure
Deep belief network / RBM	Unsupervised pretraining	Ferrag et al. (2020)	Effective with limited labels	Slow training; largely superseded
Autoencoder (stacked, non-symmetric)	Unsupervised / hybrid	Shone et al. (2018); Khan et al. (2019)	Learns compact features; no attack labels needed	Threshold selection is sensitive
Ensemble autoencoder (online)	Unsupervised	Mirsky et al. (2018)	Real-time operation on constrained hardware	Limited to anomaly, not classification
Convolutional neural network	Supervised	Ferrag et al. (2020)	Captures local byte and flow patterns	Requires fixed-size input encoding
LSTM / GRU	Supervised	Kim et al. (2016); Yin et al. (2017)	Models long-range temporal dependency	Sequential training; higher latency
Bidirectional LSTM	Supervised	Imrana et al. (2021)	Full-context modelling; lower false alarms	Not suited to strictly online detection
Deep reinforcement learning	Interactive	Nguyen and Reddi (2023)	Learns adaptive response policies	Needs safe environment; reward design is hard

2.3 Neural Approaches to Malware, Botnets, and Encrypted Traffic

Beyond intrusion detection narrowly construed, neural methods have been applied to related problems where the same representational advantages apply.

Online anomaly detection at the network edge was addressed by Mirsky et al. (2018), whose Kitsune system uses an ensemble of small autoencoders operating over incrementally maintained statistical features of the traffic stream. The design is unsupervised, requires no attack labels, adapts continuously to changing traffic, and was demonstrated on a Raspberry Pi, establishing that deep anomaly detection can run on commodity hardware at line rate. This addressed one of the more persistent objections to neural defence, namely that the computational cost precludes deployment where it is most needed.

The Internet of Things has proved a particularly active application area, both because constrained devices are attractive targets and because their traffic is comparatively regular and therefore amenable to behavioural modelling. Al-Garadi et al. (2020) surveyed machine and deep learning methods across the IoT protocol stack, organising the literature by layer and identifying resource constraints, heterogeneity, and the absence of standardised evaluation as the dominant obstacles. Their analysis emphasises that model selection in IoT contexts is governed as much by memory and energy budgets as by accuracy.

Privacy-preserving distributed training emerged as a response to the difficulty of centralising traffic data. Nguyen et al. (2019) proposed DÏoT, a federated self-learning anomaly detection system in which security gateways train device-type-specific behavioural models locally and share only model parameters with an aggregating server. The system detected compromised devices in a Mirai-infected deployment without requiring labelled attack data and without transmitting raw traffic, addressing both the labelling and the privacy constraint simultaneously. The federated formulation has since become one of the principal architectural patterns in the field.

Sequential decision-making under adversarial conditions is the domain of deep reinforcement learning. Nguyen and Reddi (2023) surveyed DRL applications across cyber-physical system protection, autonomous intrusion detection, and multi-agent game-theoretic defence simulation. Their assessment is that DRL is well suited to problems in which the defender must select among actions with delayed and uncertain consequences, such as dynamic firewall reconfiguration, honeypot placement, or resource allocation under attack, but that reward specification, sample efficiency, and the risk of learning policies that are themselves exploitable remain unresolved.

Running through this body of work is a tension identified early by Apruzzese et al. (2018), who examined the practical effectiveness of machine and deep learning for detecting malware, intrusions, and spam in operational settings. Their conclusion was that laboratory results systematically overstate field performance, that adversaries adapt in ways that static evaluation cannot capture, and that machine learning should be positioned as an augmentation of analyst capability rather than a replacement for it.

3. Conclusion

First, the shift from signature matching to learned representation is substantive rather than cosmetic. The decisive advantage of neural methods is not raw classification accuracy, on which well-tuned ensembles of shallow classifiers often remain competitive, but the

elimination of the manual feature engineering bottleneck. Because deep models learn discriminative structure directly from traffic, they can be retrained against new attack classes without a corresponding investment in expert feature design, and they can exploit interactions among features that hand-crafted descriptors discard (Javaid et al., 2016; Shone et al., 2018; Vinayakumar et al., 2019).

Second, architecture selection is governed by the structure of the detection problem rather than by any general ranking of models. Autoencoder-based methods are appropriate where labelled attack data is unavailable and where reconstruction error provides a natural anomaly signal, and they have been shown to operate online on constrained hardware (Mirsky et al., 2018). Recurrent and bidirectional recurrent models are appropriate where the attack manifests as a temporal sequence rather than a single observation, and they deliver their largest gains on the rare attack categories that shallow methods systematically miss (Kim et al., 2016; Yin et al., 2017; Imrana et al., 2021). Reinforcement learning is appropriate where the defensive problem is one of action selection under uncertainty rather than classification (Nguyen & Reddi, 2023).

Third, the evidentiary base is weaker than the volume of publication suggests. The field's benchmarks are aging, unevenly labelled, and unrepresentative of operational class balance, and methodological practices including temporally inconsistent splits and accuracy-centred evaluation systematically inflate reported performance (Ring et al., 2019; Arp et al., 2022). Comparisons across studies that use different datasets and preprocessing pipelines are not meaningful, and the small number of studies employing a uniform experimental protocol across multiple architectures carries disproportionate evidentiary weight (Ferrag et al., 2020).

4. Future Work

Several research directions follow from the limitations identified above, and they can be grouped into five areas.

Adversarially robust detection. The most urgent requirement is detection that degrades gracefully rather than catastrophically under manipulation. Adversarial training, certified robustness bounds, ensemble diversification, and detection of adversarial inputs themselves have all been proposed, but none has been evaluated systematically in the network security setting under realistic constraints. Network traffic differs from the image domain in which most adversarial research was conducted, because perturbations must preserve protocol validity and the functional effect of the attack. Characterising the feasible perturbation space for network features, and designing defences that exploit those constraints, is a well-defined and tractable programme (Papernot et al., 2016; Biggio & Roli, 2018).

Federated and privacy-preserving learning. Regulatory restriction and commercial sensitivity prevent the pooling of traffic data that centralised training assumes, and federated approaches offer a route around this by exchanging model updates rather than raw observations (Nguyen et al., 2019). Open problems include robustness of the aggregation step to poisoned updates from compromised participants, communication efficiency under bandwidth constraints, handling of non-identically-distributed data across heterogeneous participants, and the formal privacy guarantees that parameter sharing actually provides. Combining federated

training with differential privacy and with secure aggregation protocols is an active and promising direction.

Explainable security analytics. Detection without explanation is of limited operational value, because an analyst cannot triage an alert whose basis is opaque, and because regulatory frameworks increasingly require that automated decisions affecting access be justifiable. Attention mechanisms that expose which portions of a traffic sequence drove a classification, feature attribution methods adapted to the network domain, and hybrid architectures that pair a neural detector with an interpretable surrogate all merit systematic investigation. The evaluation question, namely how to measure whether an explanation is useful to a security analyst rather than merely plausible, is itself unresolved.

Continual learning and benchmark renewal. Models must adapt to distributional change without catastrophic forgetting of previously learned attack classes and without opening a retraining pipeline to poisoning. Online and incremental architectures, drift detection triggers, and rehearsal-based methods are candidates. This work depends on infrastructure: the field requires continuously generated, realistically labelled datasets that reflect contemporary protocols, encrypted traffic, cloud-native and IoT deployment patterns, and operational class balance, together with shared evaluation protocols enforcing temporally consistent splits and adversarial testing (Ring et al., 2019; Arp et al., 2022).

Autonomous and coordinated response. Extending learning-based defence from detection to response would allow reconfiguration, isolation, and containment to occur at machine speed. Deep reinforcement learning provides the natural formalism, and multi-agent formulations permit modelling of the interaction between defender and adversary as a repeated game (Nguyen & Reddi, 2023). The obstacles are practical and severe: reward functions must encode operational cost as well as security benefit, training requires high-fidelity simulation environments because exploration on production networks is unacceptable, and any autonomous responder becomes a target for adversaries seeking to induce damaging self-inflicted actions. Safe exploration, verifiable action bounds, and graduated human oversight are prerequisites for deployment.

Cutting across all five areas is the need for efficiency. Model compression, pruning, quantisation, and knowledge distillation determine whether a given architecture can run at the network edge or at carrier line rates, and a detector that cannot be deployed where the traffic is has no operational value regardless of its benchmark accuracy (Al-Garadi et al., 2020; Mirsky et al., 2018). Progress in the field will be measured less by incremental gains on established benchmarks than by demonstrated reliability under the conditions that actual networks impose.

References

1. Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for Internet of Things (IoT) security. *IEEE Communications Surveys & Tutorials, 22*(3), 1646–1685. https://doi.org/10.1109/COMST.2020.2988293

2. Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. In *2018 10th International Conference on Cyber Conflict (CyCon)* (pp. 371–390). IEEE. https://doi.org/10.23919/CYCON.2018.8405026
3. Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., Cavallaro, L., & Rieck, K. (2022). Dos and don'ts of machine learning in computer security. In *Proceedings of the 31st USENIX Security Symposium* (pp. 3971–3988). USENIX Association.
4. 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
5. 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
6. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications, 50*, 102419. https://doi.org/10.1016/j.jisa.2019.102419
7. Imrana, Y., Xiang, Y., Ali, L., & Abdul-Rauf, Z. (2021). A bidirectional LSTM deep learning approach for intrusion detection. *Expert Systems with Applications, 185*, 115524. https://doi.org/10.1016/j.eswa.2021.115524
8. Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (BICT)* (pp. 21–26). ICST. https://doi.org/10.4108/eai.3-12-2015.2262516
9. Khan, F. A., Gumaei, A., Derhab, A., & Hussain, A. (2019). A novel two-stage deep learning model for efficient network intrusion detection. *IEEE Access, 7*, 30373–30385. https://doi.org/10.1109/ACCESS.2019.2899721
10. Kim, J., Kim, J., Thu, H. L. T., & Kim, H. (2016). Long short term memory recurrent neural network classifier for intrusion detection. In *2016 International Conference on Platform Technology and Service (PlatCon)* (pp. 1–5). IEEE. https://doi.org/10.1109/PlatCon.2016.7456805
11. Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. In *Proceedings of the 2018 Network and Distributed System Security Symposium (NDSS)*. Internet Society. https://doi.org/10.14722/ndss.2018.23204
12. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. In *2015 Military Communications and Information Systems Conference (MilCIS)* (pp. 1–6). IEEE. https://doi.org/10.1109/MilCIS.2015.7348942

13. Nguyen, T. D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan, N., & Sadeghi, A.-R. (2019). DÏoT: A federated self-learning anomaly detection system for IoT. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)* (pp. 756–767). IEEE. https://doi.org/10.1109/ICDCS.2019.00080
14. Nguyen, T. T., & Reddi, V. J. (2023). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems, 34*(8), 3779–3795. https://doi.org/10.1109/TNNLS.2021.3121870
15. Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2016). The limitations of deep learning in adversarial settings. In *2016 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 372–387). IEEE. https://doi.org/10.1109/EuroSP.2016.36
16. 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
17. 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). SCITEPRESS. https://doi.org/10.5220/0006639801080116
18. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence, 2*(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792
19. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access, 7*, 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334
20. Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access, 5*, 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418