

University of midocean

Faculty of Computer Science & Information Technology

Master of Science in Artificial Intelligence

Machine Learning-Based Intrusion Detection Systems
in Network Security: A Comparative Literature Review

Assignment 1 — Literature Review

Course Code	AI800
Course Title	Research Methodology in Computer Science
Lecturer	Dr. Mohammed Anbar
Student Name	Ibrahim Saber
Student ID	261002585
Trimester	First Trimester 2026
Submission Date	14 March 2026
Weighting	20%

1. Research Title

Machine Learning-Based Intrusion Detection Systems in Network Security: A Comparative Literature Review

This literature review investigates the application of machine learning (ML) and deep learning (DL) techniques to network intrusion detection systems (IDS). As cyber threats grow in sophistication and volume, traditional signature-based IDS are increasingly insufficient. The primary objective of this review is to examine, compare, and critically analyse six peer-reviewed studies that propose or evaluate ML/DL-based IDS solutions, thereby identifying trends, methodological strengths, weaknesses, and open research challenges in this rapidly evolving domain.

2. Literature Review

The following six papers were selected because they represent seminal or highly cited contributions to the field of ML-based network intrusion detection. The selection spans from 2009 to 2019 to capture the progression of methodological sophistication in this domain, from foundational dataset analysis to advanced deep learning architectures deployed on big-data platforms.

2.1 Paper 1: A Detailed Analysis of the KDD CUP 99 Data Set

Reference:

Tavallae, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A Detailed Analysis of the KDD CUP 99 Data Set. Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), pp. 1–6.

Summary:

This paper provides a rigorous statistical examination of the KDD CUP 99 dataset, which was — and remains — one of the most widely used benchmarks for evaluating intrusion detection systems. The authors discovered that approximately 78% of the records in the training set are duplicates and around 75% of the test set records are duplicated, severely biasing evaluation metrics. As a remedy, they proposed and released the NSL-KDD dataset, which eliminates redundant records, balances difficulty levels across attack categories, and provides a more reliable benchmark for future IDS research. The methodology involved statistical comparison of class distributions, frequency analysis of duplicate records, and evaluation of several classifiers (C4.5, NaiveBayes, and NBTree) on both KDD CUP 99 and NSL-KDD to demonstrate improved generalisation on the cleaner dataset.

Critical Analysis:

Strengths: The paper addresses a fundamental flaw that undermined years of prior IDS research, making it an indispensable methodological contribution. The release of NSL-KDD as a public resource added significant practical value to the research community. The statistical evidence presented is transparent and reproducible. **Weaknesses:** The paper does not introduce a new intrusion detection algorithm; it is purely a dataset critique. Moreover, both KDD CUP 99 and NSL-KDD were constructed from simulated network traffic rather than real-world environments, which limits ecological validity. The dataset is also now outdated and does not reflect modern attack vectors such as advanced persistent threats (APTs) or Internet-of-Things (IoT) intrusions.

2.2 Paper 2: A Deep Learning Approach for Intrusion Detection Using RNNs

Reference:

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.

Summary:

Yin et al. proposed an intrusion detection model based on Recurrent Neural Networks (RNN), specifically leveraging the Long Short-Term Memory (LSTM) architecture to capture temporal dependencies in network traffic sequences. The model was trained and evaluated on the NSL-KDD dataset. The authors conducted binary classification (normal vs. attack) and multi-class classification (identifying specific attack types: DoS, Probe, R2L, U2R) experiments. Results demonstrated that the RNN-IDS outperformed several traditional classifiers, including Support Vector Machines (SVM) and Random Forests, achieving detection accuracy of up to 99.61% in binary classification and 83.28% in multi-class classification. The study also provided a sensitivity analysis of hyper-parameters such as the number of hidden units and learning rate.

Critical Analysis:

Strengths: The application of LSTM to model temporal patterns in network traffic is a meaningful innovation over stateless classifiers. The paper provides thorough hyper-parameter analysis, enhancing reproducibility. The comparison against multiple baseline classifiers gives credible evidence of superiority. **Weaknesses:** Evaluation is entirely confined to NSL-KDD; no real-world traffic dataset or live network deployment is tested, raising questions about external validity. The computational overhead of LSTM models is acknowledged but not rigorously quantified in terms of real-time feasibility. Additionally, multi-class accuracy of 83.28% for rare attack categories (R2L, U2R) may be insufficient for production environments where these attacks carry the highest risk.

2.3 Paper 3: A Deep Learning Approach to Network Intrusion Detection

Reference:

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.

Summary:

Shone et al. introduced a novel non-symmetric deep autoencoder (NDAE) combined with a Random Forest classifier to form a hybrid IDS. The NDAE performs unsupervised feature learning by encoding network traffic into compact latent representations without requiring decoding, reducing computational burden compared to conventional stacked autoencoders (SAE). The refined feature representations are then passed to a Random Forest for final classification. Experiments on both KDD CUP 99 and NSL-KDD demonstrated classification accuracy of 97.85% on KDD CUP 99 and 85.42% on NSL-KDD, with training times reduced by approximately 15% compared to SAE-based approaches. The paper is notable for combining unsupervised pre-training with supervised classification in an efficient pipeline.

Critical Analysis:

Strengths: The NDAE design is computationally efficient and represents a genuine architectural contribution. The hybrid unsupervised-supervised approach reduces dependence on labelled data during the feature extraction phase, which is practically important as labelled intrusion data is scarce.

The comparative experiments across two datasets strengthen the validity of claims. Weaknesses: The study still relies exclusively on benchmark datasets that do not represent real-world traffic diversity. The Random Forest component introduces a degree of interpretability loss. Performance on NSL-KDD (85.42%) is notably lower than on KDD CUP 99 (97.85%), suggesting the model may be over-fitted to characteristics of the older, noisier dataset. Encrypted network traffic — a dominant characteristic of modern networks — is not considered.

2.4 Paper 4: Deep Learning Approach for Intelligent Intrusion Detection System

Reference:

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.

Summary:

This study presents a systematic evaluation of multiple deep neural network (DNN) architectures — including multi-layer perceptrons (MLP), convolutional neural networks (CNN), RNN/LSTM, and hybrid CNN-LSTM models — for intrusion detection. The evaluation was performed on both the KDD CUP 99 and UNSW-NB15 datasets to ensure cross-dataset generalisability. The authors also explored transfer learning as a mechanism for improving detection on underrepresented attack classes. Key findings indicated that CNN-LSTM hybrid architectures achieved the highest overall accuracy (up to 99.44% on KDD CUP 99 and 88.79% on UNSW-NB15). The paper further examined the effect of network depth, dropout regularisation, and batch normalisation on detection performance.

Critical Analysis:

Strengths: The breadth of this study — comparing seven DNN architectures across two datasets — is exceptional and provides the research community with a valuable empirical reference. The inclusion of UNSW-NB15, a more modern dataset, enhances practical relevance. Weaknesses: Deep neural networks function as black boxes, offering no explanatory insight into why a particular traffic pattern is classified as an intrusion, which is a significant concern for security practitioners requiring audit trails. The computational cost of training and inference for complex CNN-LSTM models is high, making real-time deployment on constrained hardware challenging. Encrypted and obfuscated traffic remains outside the scope of the evaluation.

2.5 Paper 5: Survey of Intrusion Detection Systems — Techniques, Datasets and Challenges

Reference:

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges. *Cybersecurity*, 2(1), 20.

Summary:

Khraisat et al. conducted a comprehensive systematic literature review of 124 IDS research papers published between 2000 and 2019. The survey established a taxonomy categorising IDS into signature-based, anomaly-based, and hybrid approaches, and further subcategorised the methods by the machine learning algorithms employed (e.g., decision trees, neural networks, SVM, ensemble methods). The authors critically compared datasets including KDD CUP 99, NSL-KDD, UNSW-NB15, CICIDS2017, and others, evaluating their suitability for representative benchmarking. Key findings highlighted persistent challenges: lack of standardised evaluation metrics, absence of real-world validated datasets, limited focus on IoT-specific intrusions, and poor handling of zero-day attacks.

Critical Analysis:

Strengths: The paper's comprehensive scope and structured taxonomy make it an authoritative reference for the IDS research community. Its identification of the dataset standardisation problem and evaluation inconsistencies is particularly valuable for guiding future research. The survey is methodologically rigorous, with a clearly defined inclusion/exclusion criteria for paper selection. **Weaknesses:** As a survey, it does not contribute original experimental data or a novel algorithm. The rapid pace of development in this field means portions of the findings may become outdated quickly. The review also does not deeply examine privacy implications of ML-based IDS or the adversarial robustness of reviewed methods.

2.6 Paper 6: Intrusion Detection Using Big Data and Deep Learning Techniques

Reference:

Faker, O., & Dogdu, E. (2019). Intrusion Detection Using Big Data and Deep Learning Techniques. Proceedings of the 2019 ACM Southeast Conference (ACM SE '19), pp. 86–93.

Summary:

Faker and Dogdu proposed an IDS pipeline that integrates big data processing — using Apache Spark — with deep learning models (CNN and RNN/LSTM) to address the scalability limitations of single-machine IDS frameworks. The pipeline was evaluated on the UNSW-NB15 dataset. By distributing data preprocessing and model inference across a Spark cluster, the authors achieved significant reductions in processing time for high-volume traffic streams. CNN achieved the best classification accuracy of 98.36%, while the RNN achieved 97.13%. The paper demonstrated that integrating big data frameworks with DL models is feasible and efficient, paving the way for production-grade, scalable IDS deployment.

Critical Analysis:

Strengths: Addressing the scalability challenge is a critical practical contribution, as real-world networks generate traffic volumes that overwhelm single-node systems. The use of UNSW-NB15 — a relatively modern dataset — lends more credibility to the findings than studies using only KDD CUP

99. The Spark integration demonstrates a clear pathway to production deployment. Weaknesses: The evaluation is confined to a single dataset, limiting generalisability. The complexity of setting up and maintaining a Spark-DL infrastructure poses a significant barrier for small and medium-sized organisations. The paper does not address model explainability or the performance of the system against adversarial attacks designed to evade deep learning classifiers.

3. Connections and Conflicts Between Papers

An analysis of the six reviewed papers reveals several important thematic connections and methodological tensions:

3.1 Convergent Findings:

- **Dataset dependency:** Five of the six papers (Tavallae et al., Yin et al., Shone et al., Vinayakumar et al., and Faker & Dogdu) use KDD CUP 99 or NSL-KDD, reinforcing both the centrality of these benchmarks and — as Tavallae et al. cautioned — the risk of over-optimistic performance metrics due to dataset limitations.
- **Superiority of deep learning:** Yin et al. (2017), Shone et al. (2018), Vinayakumar et al. (2019), and Faker & Dogdu (2019) all demonstrate that deep learning architectures outperform traditional ML classifiers, establishing a clear research consensus on this point.
- **Benchmark inadequacy:** Khraisat et al. (2019) and Tavallae et al. (2009), though separated by a decade, both emphasise that the available benchmark datasets do not adequately represent real-world network conditions. This shared concern is echoed implicitly by studies that achieve high accuracy on benchmarks but do not validate results in live environments.

3.2 Conflicting Findings and Tensions:

- **Performance inconsistency across datasets:** Shone et al. report 97.85% accuracy on KDD CUP 99 but only 85.42% on NSL-KDD. Vinayakumar et al. achieve 99.44% on KDD CUP 99 but 88.79% on UNSW-NB15. These discrepancies highlight that model performance is dataset-dependent and challenge the reliability of accuracy as a universal performance measure.
- **Scalability vs. accuracy trade-off:** Faker & Dogdu explicitly prioritise scalability through big data integration, whereas Yin et al. and Shone et al. focus on maximising classification accuracy without considering deployment at scale. These differing priorities reflect a fundamental tension in IDS research between model optimisation and operational practicality.
- **Interpretability gap:** All experimental papers employ deep learning, which sacrifices interpretability for performance. However, neither the experimental papers nor the survey (Khraisat et al.) propose satisfactory solutions to this problem, leaving a significant gap that may hinder adoption in regulated environments.

4. Summary Table of Reviewed Papers

Year	Author(s)	Methodology	Key Findings	Weaknesses / Limitations	Citation
2009	Tavallae et al.	Statistical analysis & pre-processing of KDD CUP 99 dataset	Identified severe redundancy in training/test sets affecting classifier accuracy	No new machine learning model proposed; study limited to dataset critique without broader network environments	Tavallae et al. (2009)
2017	Yin et al.	Recurrent Neural Network (RNN / LSTM) on NSL-KDD dataset	RNN-IDS outperformed SVM and Random Forest in binary & multi-class classification; accuracy up to 99.61%	Evaluated only on NSL-KDD; no real-time traffic validation; high computational cost limits deployment	Yin et al. (2017)
2018	Shone et al.	Non-symmetric Deep Auto-encoder (NDAE) + Random Forest (unsupervised + supervised hybrid)	Achieved 97.85% accuracy on KDD CUP 99 with significantly reduced training time vs. stacked autoencoders	Results not generalised beyond benchmark datasets; sensitivity to network traffic variability not tested	Shone et al. (2018)
2019	Vinayakumar et al.	Deep Neural Networks (DNN) with various architectures; evaluated on KDD CUP 99 & UNSW-NB15	Deep learning achieved superior detection rates; scalable across multiple benchmark datasets	Black-box nature of DNN reduces interpretability; high resource requirements; encrypted traffic not addressed	Vinayakumar et al. (2019)
2019	Khraisat et al.	Systematic literature review of 124 IDS papers; taxonomy of techniques & datasets	Comprehensive taxonomy of anomaly, signature, and hybrid IDS; identified gaps in real-time and IoT security	Survey-based; no original experimental data; findings may become outdated rapidly in the evolving threat landscape	Khraisat et al. (2019)
2019	Faker & Dogdu	Big Data framework (Apache	Big Data + DL pipeline improved	Relies on UNSW-NB15 only; complexity of	Faker & Dogdu (2019)

Year	Author(s)	Methodology	Key Findings	Weaknesses / Limitations	Citation
		Spark) + Deep Learning (CNN, RNN) on UNSW-NB15	scalability and detection speed; CNN showed best accuracy (98.36%)	Spark-DL integration raises deployment barriers for small organisations	

5. Synthesis of Information

The six reviewed papers collectively paint a coherent narrative of progress, persistent challenges, and emerging research directions in ML-based network intrusion detection systems.

The field's trajectory is clearly defined: from the foundational recognition that benchmark datasets are fundamentally flawed (Tavallaei et al., 2009), to the empirical validation of deep learning superiority over traditional classifiers (Yin et al., 2017; Shone et al., 2018), followed by a comprehensive cross-architecture evaluation (Vinayakumar et al., 2019), a synthesising survey identifying persistent gaps (Khraisat et al., 2019), and culminating in a scalable, production-oriented framework (Faker & Dogdu, 2019). This progression reflects a maturing research community moving from proof-of-concept experimentation towards deployment-ready solutions.

Three cross-cutting themes emerge from the synthesis. First, the dataset challenge remains unresolved: despite NSL-KDD's improvement over KDD CUP 99 and the introduction of UNSW-NB15 and CICIDS2017, no universally accepted benchmark that captures contemporary attack patterns — including APTs, encrypted traffic attacks, and IoT vulnerabilities — exists. This gap inhibits meaningful cross-study comparisons and may lead to overestimated detection capabilities in practice.

Second, there is a recurring tension between model complexity and operational practicality. Deep learning models, particularly LSTM and CNN-LSTM hybrids, achieve impressive accuracy but demand substantial computational resources. Faker and Dogdu's big-data integration approach represents the most direct effort to bridge this gap, yet it introduces its own infrastructure complexity that many organisations cannot sustain.

Third, interpretability and adversarial robustness represent critical blind spots across the reviewed literature. None of the experimental studies assess how their models perform under adversarial attacks — a significant oversight, given that sophisticated attackers are aware of ML-based defences and may craft inputs to evade them. Similarly, the black-box nature of deep neural networks conflicts with the requirements of security audits and regulatory frameworks in sectors such as finance and healthcare.

In conclusion, the reviewed literature demonstrates that ML and deep learning are powerful tools for intrusion detection, but the field must prioritise the development of realistic, standardised datasets, computationally efficient and interpretable models, and rigorous adversarial evaluation frameworks before these approaches can be fully trusted in high-stakes production environments.

6. References

- Faker, O., & Dogdu, E. (2019). Intrusion detection using big data and deep learning techniques. *Proceedings of the 2019 ACM Southeast Conference*, pp. 86–93. <https://doi.org/10.1145/3299815.3314439>
- Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, 2(1), 20. <https://doi.org/10.1186/s42400-019-0038-7>
- 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>
- Tavallae, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, pp. 1–6. <https://doi.org/10.1109/CISDA.2009.5356528>
- 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>
- 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>