Contemporary Research Analysis Journal

e-ISSN: 3050-5909
p-ISSN: 3050-5895

Open Access

Volume 03 Issue 06 June 2026

CrossRef DOI: 10.55677/CRAJ/10-2026-Vol03I06

Page no: 471-479

Enhancing Network Intrusion Detection Systems Using Deep Learning-Based Anomaly Detection Models

Okoli C. Johnson[1], Bukola A. Akindipe[2]

[1]University of Birmingham, United Kingdom

[2]North Carolina A & T University

ABSTRACT: Deep learning has revolutionised pattern recognition by integrating feature extraction and classification into end-to-end models, eliminating the need for manual feature engineering. This survey paper reviews recent advancements in deep learning–based anomaly detection for network intrusion detection systems (NIDS). The review focuses on three major classes of models: convolutional neural networks (CNNs), recurrent architectures (particularly LSTMs), and autoencoders. Each class demonstrates distinct strengths in capturing spatial, temporal, and latent representations of network traffic. The paper synthesises key studies applying these models to anomaly-based intrusion detection, compares reported performance across commonly used benchmark datasets, and discusses their effectiveness in detecting previously unseen attacks. Finally, it highlights ongoing challenges such as class imbalance and concept drift, and outlines future research directions, including adversarial training and online adaptation, to enhance deep-learning-driven NIDS.

KEYWORDS: Intrusion detection, anomaly detection, deep learning, convolutional neural networks, recurrent neural networks, autoencoders, cybersecurity.

Cite the Article: *Johnson, O.C., Akindipe, B. A. (2026). Enhancing Network Intrusion Detection Systems Using Deep Learning-Based Anomaly Detection Models. Contemporary Research Analysis Journal, 3(6), 471– 479. https://doi.org/10.55677/CRAJ/10-2026-Vol03I06*

Publication Date: June 13, 2026

**Corresponding Author: Okoli C. Johnson*

1. INTRODUCTION

Network intrusion detection systems (NIDS) monitor live network traffic to identify and prevent malicious activities. Hence, these intrusion detection systems fall into two broad categories: signature-based systems, which detect known attacks by matching predefined patterns, and anomaly-based systems, which help identify deviations from normal network behaviour [4]. However, anomaly-based NIDS are particularly valuable for detecting novel and zero-day attacks since they do not rely on prior signatures; however, they are often associated with higher false-positive rates [4]. Thus, developing effective anomaly detection remains challenging because network traffic is inherently high-dimensional, dynamic, and noisy [5].

However, machine learning (ML) techniques such as Naïve Bayes, support vector machines (SVMs), and random forests have been used for NIDS. Although these methods can deliver satisfactory results, they often require extensive feature engineering and data preprocessing [6], [15], [12]. Additionally, as network traffic grows in scale and variety, traditional ML models find it harder to generalize well and remain accurate [7]. Thus, deep learning methods can automatically learn layered feature representations directly from raw data [2]. However, models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and autoencoders have shown strong representation-learning capabilities and have been increasingly applied to intrusion detection tasks [5], [6], [17]. Therefore, this paper examines how CNNs, long short-term memory (LSTM) networks, and autoencoder models have been applied to anomaly-based NIDS, with emphasis on recent studies and reported state-of-the-art results in detection accuracy, scalability, and robustness. This paper provides an in-depth review of deep-learning-based anomaly detectors for NIDS, covering individual and hybrid architectures, benchmark datasets, reported performance results, and real-world deployment challenges.

2. REVIEW OF LITERATURE

This review considers studies on deep learning-based anomaly detection for NIDS, with emphasis on CNN, LSTM, autoencoder, and hybrid architectures. The selected studies were drawn from peer-reviewed articles and widely cited works relevant to benchmark

datasets such as KDD'99, NSL-KDD, UNSW-NB15, and CIC-IDS2017. The review prioritises studies that report detection performance, model architecture, dataset usage, and practical deployment challenges.

Contextual Review of Anomaly-Based NIDS and Deep Learning

Anomaly-based Network Intrusion Detection Systems (NIDS) are designed to identify malicious activities by learning normal network traffic behaviour and flagging any significant deviations as potential intrusions. Unlike signature-based approaches, which are limited to detecting known attack patterns, anomaly-based systems can identify zero-day and previously unseen threats [15]. Despite this advantage, traditional statistical and machine learning–based anomaly detectors often suffer from high false positive rates, scalability challenges, and difficulty in adapting to dynamic traffic patterns [16]. These limitations have created a growing need for more robust, adaptive, and scalable detection techniques.

To address these challenges, recent research has increasingly applied deep learning techniques such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and autoencoders to anomaly-based NIDS. Recent studies have demonstrated that CNNs, LSTMs, and autoencoder-based models can effectively learn hierarchical and temporal representations of network traffic. For instance, Shone et al. [17] showed that deep autoencoder-based feature learning significantly improves intrusion detection accuracy, while other works have highlighted the complementary strengths of CNNs for spatial feature extraction and LSTMs for temporal modelling. Furthermore, hybrid CNN–LSTM architectures have been shown to improve detection accuracy and reduce false alarm rates in several studies on modern datasets such as CICIDS2017 and UNSW-NB15 [18]. Nevertheless, model interpretability, data imbalance, and computational overhead remain important challenges for real-time deployment in deep-learning-based anomaly detection [19]. These challenges are particularly significant in NIDS, where models must balance detection accuracy, efficiency, and transparency under dynamic network conditions [22], [23].

NIDS based on anomaly detection and Deep Learning.

Network traffic can be characterised by features (such as packet sizes, inter-arrival times, and protocol flags) that are monitored by a NIDS. Traditional anomaly-based IDS creates a model of "normal" traffic and flags any significant deviations as potential attacks. These systems can identify new or unknown threats by learning normal behaviour patterns [9]. In practice, a NIDS classifies each network connection or flow as "normal" or "attack" (binary classification) or may also identify the specific attack category (multiclass classification).

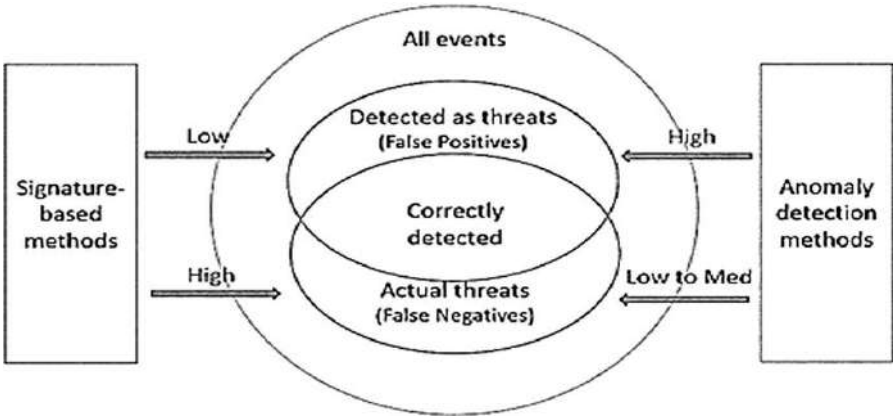

Figure 2.1 Comparison between signature-based and anomaly-based intrusion detection systems (Naqash et al; 2022)

Figure 2.1 illustrates the relationship between signature-based and anomaly-based intrusion detection methods in terms of detection outcomes. Signature-based systems rely on predefined attack patterns, which enables them to maintain a low false positive rate (i.e., fewer benign events flagged as threats). However, they often fail to detect previously unseen or modified attacks, resulting in a high false negative rate (i.e., actual threats go undetected) [20]. However, anomaly detection methods build a model of normal network behaviour and classify deviations as suspicious. This allows them to achieve higher detection of actual threats, including zero-day attacks, but at the cost of an increased false positive rate, since legitimate but unusual activities may be flagged incorrectly [21]. The overlap region in figure 2.1 highlights the subset of events that are correctly detected (true positives), showing that each approach captures part of the threat landscape but leaves gaps in accuracy. Recent studies therefore emphasize hybrid and deep learning–based NIDS, which combine both approaches to reduce false negatives while controlling false positives, thereby achieving a better balance between detection accuracy and efficiency [22], [23].

Recent literature emphasizes that traditional ML algorithms often struggle to fully capture the complexity of modern network traffic. Wang et al. [5] therefore proposed HAST-IDS, a hierarchical deep-learning IDS that automatically learns spatial and temporal traffic features using CNN and LSTM components, reducing dependence on manual feature engineering.

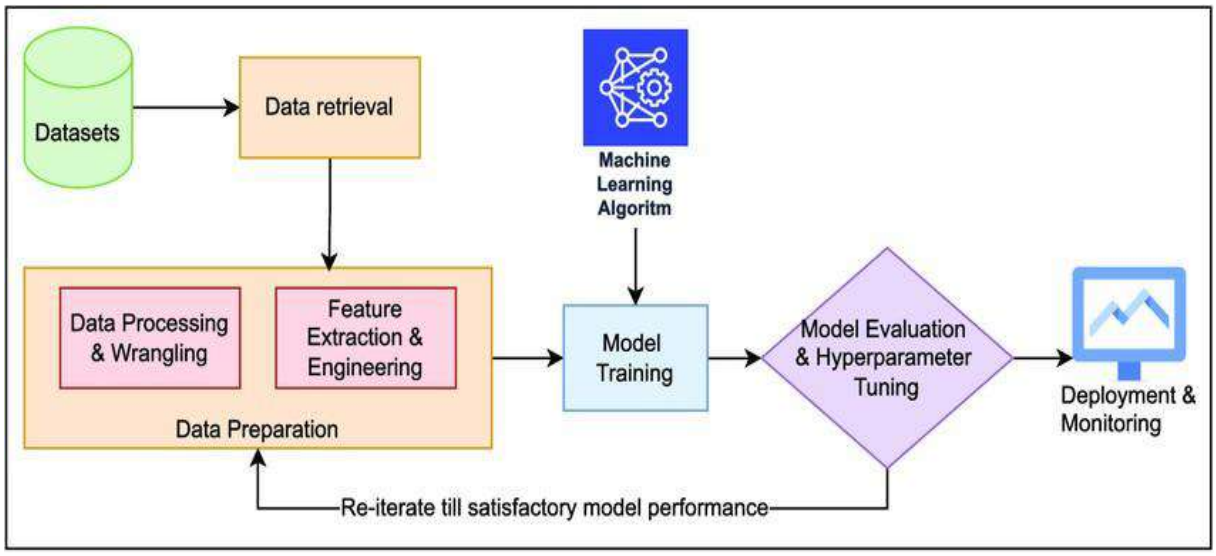

Figure 2.2. General machine-learning pipeline adapted for NIDS development (adapted from Barrak et al. [24]).

The pipeline shown in Figure 2.2 represents the general machine-learning process commonly adapted for NIDS development. The process begins with data retrieval and preparation, followed by feature engineering, model training, evaluation, hyperparameter tuning, deployment, and continuous monitoring.

Deep learning, by contrast, builds internal representations directly from raw features, which can improve detection accuracy and reduce false alarms[2], [1]. Indeed, studies have shown that deep models (CNNs, RNNs, autoencoders) can learn richer feature sets and generalize better to unseen attacks than traditional classifiers[2], [8].

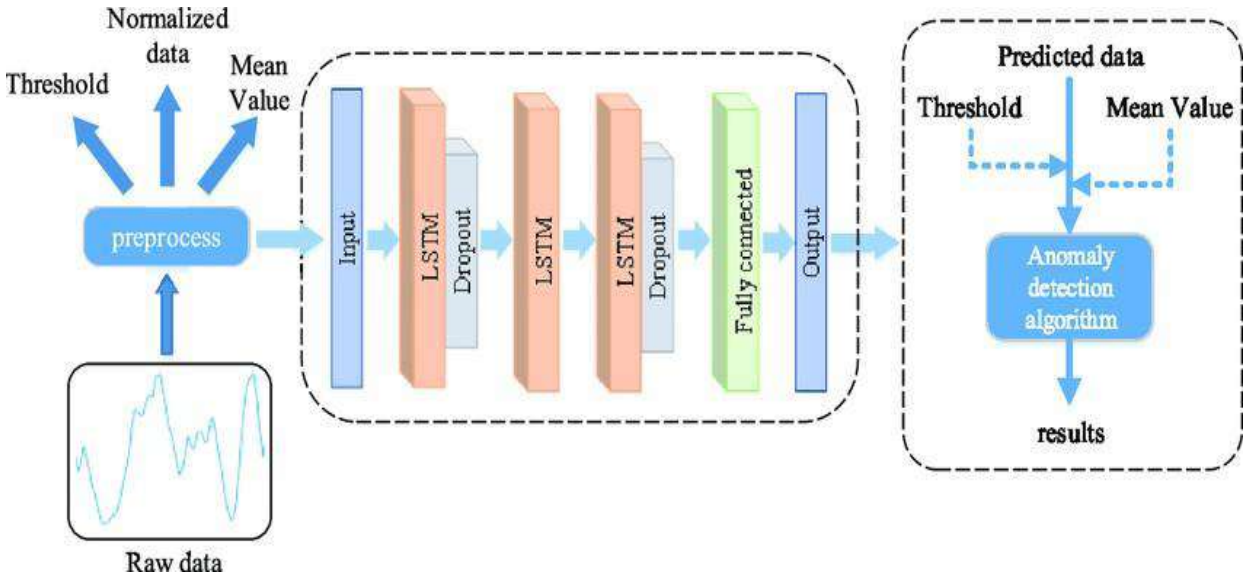

Figure 2.3. LSTM-based anomaly-detection framework for sequential network traffic (adapted from Tan et al. [31]).

Figure 2.3 presents an LSTM-based anomaly-detection framework. Raw traffic data is preprocessed and normalised before being passed through the LSTM network, which learns temporal patterns and generates predicted values. Anomalies are identified when deviations between the predicted and observed values exceed a predefined threshold.

The effectiveness of such anomaly-detection frameworks is commonly assessed using established benchmark datasets. Studies show that IDS research now often uses benchmark datasets to evaluate models. The KDD'99 dataset (derived from DARPA'98 data) was a standard, but it contains many duplicate records and outdated attack types [10], [11]. NSL-KDD is a cleaned version that removes redundancies [10], [12]. More recent datasets like UNSW-NB15 (2015) and CIC-IDS2017 include realistic modern attack scenarios (see Table 1 below). Evaluations of deep anomaly detectors are often reported on NSL-KDD, UNSW-NB15, and other public datasets [13], [10]. While these benchmarks enable comparison, caution is necessary: for example, results on KDD'99 may overestimate real-world performance due to data artefacts [10]. In this review, we cite results on these common datasets, recognising their limitations.

Table 1. Benchmark Datasets for NIDS Research

Dataset	Year	Records	Features	Attack Types	Remarks
KDD'99	1999	~4.8M	41	Four classes	Outdated, redundant
NSL-KDD	2009	125k	41	Four classes	Cleaned KDD'99
UNSW-NB15	2015	2.5M	49	Nine classes	Modern realistic data
CIC-IDS2017	2017	3M+	80+	Multiple	Realistic and diverse

Author compilation 2025

Convolutional Neural Networks (CNNs) for Anomaly Detection

Convolutional Neural Networks (CNNs) are deep architectures originally designed for image and spatial data analysis, where they automatically extract hierarchical features through convolutional filters applied to local regions of the input [7], [14]. Within the intrusion detection domain, CNNs have been adapted by representing network traffic features in structured forms, such as two-dimensional matrices or sequential feature vectors, enabling the application of 2D or 1D convolutions. By exploiting local correlations and weight sharing, CNNs reduce model complexity while learning discriminative patterns relevant for intrusion detection tasks [7].

Several studies have demonstrated the effectiveness of CNN-based intrusion detection models. Al-Turaiki and Altwaijry [11] applied CNN architectures to anomaly-based network intrusion detection using benchmark datasets, including NSL-KDD and UNSW-NB15, and reported competitive accuracy and recall. Similarly, Li et al. [26] transformed NSL-KDD traffic features into image representations and employed CNN architectures for intrusion detection, reporting accuracies of approximately 77–79% on NSL-KDD.

Beyond standalone CNNs, comparative studies have shown that deep learning–based feature extraction can outperform conventional neural architectures on legacy benchmarks such as KDD'99. For example, Shone et al. [17] demonstrated that a deep autoencoder-based model significantly outperformed traditional classifiers, achieving accuracy above 97% on KDD'99. More recent works further indicate that hybrid CNN–LSTM architectures can surpass individual CNN or RNN models by jointly capturing spatial and temporal characteristics of network traffic, with several studies reporting detection accuracies exceeding 96% on benchmark datasets [6], [12].

Despite their advantages, CNNs also face notable challenges. When network data lacks a natural spatial structure, constructing a suitable representation can be non-trivial. Comparative evaluations indicate that recurrent models may surpass CNNs in scenarios where temporal dependencies are critical, as observed in [18], where RNNs achieved higher accuracy than CNNs on NSL-KDD. Consequently, CNNs are often integrated into hybrid models to leverage both spatial feature extraction and temporal modelling. For instance, attention mechanisms have been incorporated into deep learning models to improve feature selection and detection performance, illustrating the growing trend toward hybrid and attention-based architectures in intrusion detection [18], [22], [23].

Hence, CNNs contribute significantly to intrusion detection by offering automatic feature extraction, parameter efficiency through weight sharing, and competitive performance on benchmark datasets [15], [17]. However, their limitations in capturing sequential dependencies suggest that CNNs are most effective when combined with complementary models, reinforcing the shift toward hybrid deep learning frameworks for next-generation intrusion detection.

Table 2 below emphasizes the principal deep learning architectures most frequently employed in anomaly-based network intrusion detection system (NIDS) research. In operational industrial settings, these architectures are commonly augmented with supplementary paradigms, including transformer-based sequence models, graph neural networks for relational traffic analysis, self-supervised representation learning, and ensemble-based detection pipelines. Although such approaches are progressively being integrated into practical applications, they are often constructed upon or combined with the foundational CNN, LSTM, and autoencoder architectures summarized in Table 2. Consequently, the taxonomy presented herein captures the core model families that serve as the basis for both academic investigation and a substantial number of real-world NIDS implementations.

Table 2. Principal Deep Learning Model Families Used in Anomaly-Based NIDS

Model Type	Learning Paradigm	Strength	Limitation	Typical Datasets
CNN	Supervised	Spatial feature extraction	Weak temporal modelling	NSL-KDD, UNSW
LSTM	Supervised	Temporal modelling	Slow training	NSL-KDD
Autoencoder	Unsupervised	Zero-day detection	Threshold tuning	KDD'99
CNN–LSTM	Hybrid	Spatial + temporal	Computational cost	UNSW, CIC

Recurrent Neural Networks and LSTM for Anomaly Detection

Recurrent neural networks (RNNs) are widely employed for modelling sequential data due to their ability to propagate information across time steps, making them well-suited for analysing network traffic sequences. However, traditional or "vanilla" RNNs face the well-documented challenge of vanishing gradients when learning long temporal dependencies [20]. To address this limitation, Long Short-Term Memory (LSTM) networks were introduced, incorporating gated memory cells that enable the capture of long-range temporal patterns [21], [3]. In network intrusion detection systems (NIDS), LSTMs have demonstrated the ability to learn evolving patterns such as multi-stage attack behaviours and temporal dependencies in protocol traffic.

Several empirical studies highlight the effectiveness of LSTM-based intrusion detection models. Yin et al. [32] proposed an RNN-based intrusion detection model and evaluated it on the NSL-KDD dataset, demonstrating the suitability of recurrent architectures for modelling network traffic. Similarly, Nawaz et al. [10] reported strong detection performance using a deep learning-based intrusion detection approach. More recent works, including hybrid CNN–LSTM and autoencoder–LSTM models, further show that recurrent architectures remain useful for capturing temporal dependencies in network traffic [12], [27], [35].

Within anomaly-based detection, LSTMs are particularly effective due to their capacity to model streaming data. Other studies have combined LSTMs with autoencoders [27] or attention mechanisms [18], enhancing detection in complex contexts such as Internet of Things (IoT) networks where temporal dependencies are critical. Although evaluated outside the NIDS domain, hybrid models integrating Transformer-based encoders, CNNs, and LSTMs have demonstrated strong performance, highlighting the broader potential of combining contextual, spatial, and sequential feature learning [29].

Despite these successes, practical challenges remain. Training recurrent networks is computationally intensive and time-consuming, especially for long sequences, and requires significant hyperparameter tuning (e.g., sequence length, learning rate, and number of hidden layers) [30]. Moreover, LSTM-based IDS models often rely on supervised learning, necessitating large amounts of labelled sequential traffic data, which is costly to obtain. Nonetheless, their ability to capture temporal dependencies makes LSTMs a valuable architecture in NIDS, particularly when attack patterns are embedded in sequential traffic characteristics.

Hybrid CNN–LSTM Models: To take advantage of spatial and temporal correlations, hybrid CNN-LSTM architectures are employed by many researchers. As an example, Vinaykumar et al. (2019) established that a CNN-LSTM hybrid was superior to a pure CNN to classify multiclass intrusions [6]. The CNN component is used to extract spatial features of the packet windows, and the LSTM processes the resulting sequence of features to extract order information. Wu and Guo [35] proposed LuNet, a hierarchical CNN–RNN deep neural network for network intrusion detection. In this model, CNN layers are used to learn local traffic features, while recurrent layers capture temporal behaviour. The model was evaluated on NSL-KDD and UNSW-NB15 and showed competitive performance against baseline intrusion-detection methods [35]. To conclude, recurrent architectures and particularly LSTMs have become important to model the time aspect of network traffic, and their combination with CNNs has produced strong IDS performance in several deep-learning and hybrid NIDS studies [5], [6], [12], [35].

Autoencoders for Anomaly Detection

Autoencoders (AEs) are neural networks which are used to recreate their inputs. In anomaly detection, the point of interest is to train an AE with normal network traffic as an input, such that it learns a latent representation of benign behaviour in a compressed code [27]. At inference time, inputs that are drastically different from the norm used in training (i.e. anomalous attacks) will reconstruct poorly, producing a large reconstruction error. The model flags inputs with reconstruction errors above a selected threshold as anomalies [27]. This data-driven unsupervised method does not require labelled attack data, making it attractive for intrusion detection where new attack types continually emerge [17], [33], [34], [36].

Recent publications indicate widespread applications of autoencoder variations to IDS. Simple dense AEs, sparse autoencoders, stacked autoencoders, stochastic autoencoders, and ensemble autoencoders have been investigated for learning complex traffic patterns [17], [33], [34], [36]. Sparse autoencoders impose a sparsity penalty on the hidden representation and promote the network to learn bottleneck features. Denoising autoencoders introduce noise to the inputs in training, which enhances robustness. Ensemble methods use multiple autoencoders and combine their outputs to achieve more stable detection [34].

Many autoencoder-based IDS approaches can be trained using normal or mostly normal traffic, making them attractive for anomaly detection settings where labelled attack data is limited. However, their computational cost depends on the architecture, dataset size, and deployment environment. As an illustration, Shone et al. (2017) produced a non-symmetric deep autoencoder (NDAE) which stacked several AE layers to learn features unsupervised, and then applied a random forest classifier to the encoder features [17]. On KDD99, using five classes, 97.85 percent accuracy was obtained on this system [17]. In the same manner, Al-Qatf et al. (2018) applied a sparse autoencoder to extract features using traffic and finally introduced them to an SVM to classify [33]. Their NSL-KDD experiments indicated that the AE+SVM combination was superior to the traditional ML (Decision Trees, Naive Bayes, etc.), with an accuracy of approximately 85% when it came to binary detection [33].

Other AE-based IDS approaches include ensemble autoencoders, such as Kitsune by Mirsky et al. [34], and stochastic reconstruction-error thresholding methods proposed by Aygun and Yavuz [36]. Because many AE-based IDS approaches are trained in an unsupervised or semi-supervised manner, they can model normal traffic patterns and identify deviations without requiring

extensive labelled attack data. The disadvantage is that it might be challenging to select a reasonable error threshold and AEs might not work effectively when training data has undetected anomalies.

Key Points:

- Autoencoders learn compact representations of normal or mostly normal network traffic and detect anomalies using reconstruction error [17], [33], [34], [36].
- They are useful for anomaly-based IDS because they can support unsupervised or semi-supervised detection when labelled attack data is limited [17], [33], [34].
- Sparse, stacked, stochastic, and ensemble autoencoder variants have been used to improve feature learning, robustness, and detection stability [17], [33], [34], [36].
- Their main limitations include threshold selection, sensitivity to noisy training data, hidden anomalies in the training set, and reduced adaptability when normal traffic behaviour changes over time [33], [34], [36].
- Recent hybrid approaches combine autoencoders with CNNs or LSTMs to capture both feature relationships and temporal traffic behaviour, supporting the broader move toward more adaptive and robust anomaly-based NIDS [27], [28].

Autoencoders are thus a powerful tool for anomaly-based IDS. However, they are just one component; recent work often combines AEs with RNNs or CNNs (see next section) to model both feature interrelationships and temporal dynamics.

Hybrid Deep Models and Anomaly Detection

The majority of contemporary intrusion detection models integrate CNNs, RNNs/LSTMs, and autoencoders to take advantage of their complementary strengths. For example, Singh and Jang-Jaccard [28] proposed an autoencoder-based unsupervised intrusion-detection model using multi-scale convolutional and recurrent components to capture both spatial and temporal traffic patterns. Their MSCNN-LSTM-AE model was evaluated on NSL-KDD, UNSW-NB15, and CICDDoS2019, demonstrating the benefit of combining CNN, LSTM, and autoencoder components for network anomaly detection [28].

Similarly, LSTM-CNN hybrid models have been explored for anomaly detection, combining convolutional layers for local feature extraction with LSTM layers for temporal modelling [5], [12], [35]. LSTM-CNN is a model which employs CNN layers (usually 1D convolution) to pre-process every time window, followed by LSTM steps to enhance noise resistance. The findings indicate that the spatial and temporal features can be further combined to enhance the detection of anomalies, in the case of CNN and LSTM layers hybridisation.

Other hybrid approaches include attention-based and generative deep learning models, which have been explored to improve feature learning, anomaly discrimination, and detection robustness. However, these approaches remain beyond the main focus of this review, which centres on CNN, LSTM, autoencoder, and CNN–LSTM-based NIDS architectures [5], [18], [22], [23].

In summary, the trend in deep anomaly detection for NIDS is toward architectural fusion: combining CNNs for feature extraction, LSTMs for sequence modelling, and autoencoders for unsupervised representation. Empirical studies consistently show that such multi-component systems achieve higher accuracy and lower false alarms than any single-model approach [17].

Evaluation and Benchmark Results

Deep learning-based NIDS are typically evaluated on standard datasets. Reported performance varies by task (binary vs. multiclass) and dataset. Here we highlight representative results from the literature:

- CNN vs. CNN-LSTM: Vinaykumar et al. (2019) evaluated CNN, RNN, LSTM, and their combinations on KDD'99. They found the CNN–LSTM hybrid achieved 98.7% accuracy on the 5-class problem and 96.4% on binary classification, outperforming the single CNN [6]. The standalone CNN exceeded the other models in the binary task.
- LSTM-RNN: Nawaz et al. (2025) reported that their optimised LSTM-RNN achieved 97.54% accuracy and 98.95% detection rate on the NSL-KDD dataset [10], highlighting the potency of LSTM-based NIDS. Earlier work by Yin et al. had also shown ~83% accuracy for LSTM on NSL-KDD, illustrating rapid progress. Autoencoders: Al-Qatf et al. (2018) combined a sparse autoencoder with SVM on
- NSL-KDD and achieved ~85% accuracy for both binary and five-class tasks [33], which was higher than several traditional methods. Shone et al. (2017) achieved 97.85% accuracy on KDD'99 (five classes) using a stacked autoencoder + RF [17]. Mirsky et al. (2019) and others have reported similarly high detection rates using AE variants.
- Multidata set comparisons: Some studies evaluate models across multiple datasets. Al-Turaiki and Altwaijry [11] tested a CNN-based model on both NSL-KDD and UNSW-NB15, reporting high accuracy and recall on both. Wu and Guo's LuNet [35] was validated on NSL-KDD and UNSW-NB15, showing improvements over several baseline methods.

These results demonstrate that deep models can achieve very high detection metrics on benchmark data. However, it is important to interpret such numbers cautiously. As noted earlier, the KDD'99 dataset contains redundant records, so even very high accuracy may not generalise to real traffic[10]. The newer NSL-KDD and UNSW-NB15 datasets are more realistic but still represent limited

scenarios. Accordingly, recent reviews urge testing models on diverse, up-to-date traffic (e.g., CIC-IDS2017, UNSW, etc.) and considering metrics beyond accuracy, such as precision, recall, and false-positive rates, when evaluating NIDS.

Common Datasets and Metrics

Network intrusion research uses several standard datasets:

- KDD'99: One of the oldest NIDS datasets, derived from DARPA'98. Contains ~4.8M records with 41 features, labelled into normal and 4 attack categories[11]. It suffers from many duplicate records and outdated attack types[10], [11], making it less reliable today.
- NSL-KDD: A cleaned subset of KDD'99 that removes redundant entries to avoid bias[10], [12]. It contains several subsets for train/test and is widely used as a benchmark for anomaly detection.
- UNSW-NB15: Collected in 2015 by the Australian Centre for Cyber Security. Contains ~2.5M records, 49 features, and nine attack categories (e.g., DoS, fuzzers, worms)[7]. It reflects modern traffic patterns and has become a standard for NIDS evaluation.
- CIC-IDS2017: A dataset of real, daily network traffic with modern attack scenarios (DoS, botnets, DDoS, etc.)[8]. It is large and diverse, used to test NIDS generalisation to realistic traffic.

Evaluation metrics commonly reported include accuracy, precision, recall (detection rate), F1-score, and false-positive rate. For anomaly detection, the Area Under the ROC Curve (AUC) is also frequently used. Deep IDS papers typically present confusion matrices or aggregate metrics to compare models. As a rule, achieving both high recall (detecting attacks) and low false-positive rate (avoiding normal traffic alarms) is crucial.

Challenges and Future Directions

Despite their successes, deep learning–based anomaly detectors face several open challenges before widespread deployment in real networks:

- Data imbalance and labelling: Network attacks are rare, leading to highly skewed datasets. Deep models may be overfit to most normal traffic unless exceptional care is taken (e.g., resampling, weighted loss). Moreover, labelled data is costly; many attacks are unlabelled or evolving. Future work on semi-supervised and unsupervised learning (e.g., self-supervised methods) is needed to leverage unlabelled traffic.
- Concept drift: Network traffic patterns can change over time (new protocols, usage patterns, or attack types). Static models risk becoming outdated. Future NIDS should incorporate online learning or periodic retraining. Some propose adaptive architectures that adjust parameters in real time based on traffic characteristics[21].
- False positives: Anomaly detectors often flag benign but unusual behaviour as malicious. Too many false alarms can render an NIDS impractical. Research is needed on threshold-setting, confidence estimation, and human-in-the-loop systems to reduce false positives. In one future vision, adversarial training could be used to teach the model to distinguish subtle normal variations from true attacks[22].
- Adversarial robustness: As deep models are susceptible to adversarial inputs, NIDSs themselves may be attacked by crafted traffic. Future work should integrate adversarial machine learning techniques into NIDS design, as suggested in recent surveys[12]. For example, adversarial training or generative models could harden detectors against evasion tactics.
- Resource constraints: Deploying deep NIDS on high-speed or resource-limited networks (e.g. IoT) is challenging due to computational and memory requirements. Future approaches include model compression, pruning, or specialised hardware. Researchers also envision lightweight NIDS frameworks that dynamically adjust model complexity based on real-time network load[21], [23]. For IoT and edge environments, methods like one-time learning or federated training may be explored. In distributed or federated NIDS settings, adversarial robustness is also important because malicious clients may poison shared model updates, requiring robust aggregation and proactive defence mechanisms such as RECESS [37].
- Explainability: Deep anomaly detectors are often black boxes. In security applications, administrators may need interpretable explanations for alerts. Integrating attention mechanisms or using techniques like Layer-wise Relevance Propagation could help make deep NIDS decisions more transparent. Developing explainable NIDS remains an active area.
- Emerging techniques: New deep learning trends may benefit intrusion detection. For example, transformer-based models (with self-attention) could capture long-term patterns in traffic more efficiently than LSTMs. Graph neural networks might model the network topology of flows. Quantum machine learning, though early stage, is suggested to speed up training on large traffic datasets[24].

Hence, future deep-learning-enhanced NIDS should be adaptive, robust, and efficient. As recent reviews note, promising directions include combining adversarial and transfer learning to handle novel attacks [22], generating more realistic datasets for evaluation [25], and optimizing models for evolving networks (e.g., dynamic, IoT environments) [23].

CONCLUSION

Deep learning has significantly advanced network intrusion detection by enabling powerful anomaly-based models [10], [16]. Convolutional networks can automatically extract spatial features from traffic data, LSTM networks capture temporal dependencies in flows, and autoencoders model normal traffic to detect anomalies. This review has examined how CNNs, LSTMs, and

autoencoders have been applied to NIDS, often in hybrid combinations. Empirical studies consistently show that deep architectures can outperform traditional machine learning approaches in detection accuracy and adaptability to unknown attacks. For example, CNNs and LSTMs can improve detection precision through learned feature representations, while autoencoder-based approaches reduce reliance on labelled attack data.

Nevertheless, real-world deployment of deep IDS requires careful attention to data quality, system design, and evaluation methodology. Datasets such as KDD'99 can produce overly optimistic results because of redundancy and outdated attack patterns [10], while many deep IDS models are still evaluated mainly in simulated or benchmark environments. To improve operational reliability, future research must address imbalanced data, evolving threats, adversarial evasion, explainability, and resource constraints [22], [25].

Building on these challenges, future NIDS research could benefit from moving beyond static benchmark evaluation toward adaptive, explainable, and adversarially robust detection systems. In particular, integrating deep anomaly detection with AI-driven cyber-range environments offers a promising direction, as attack scenarios can be dynamically generated, tested, and used to retrain detection models under more realistic conditions. This could support the development of NIDS that are not only accurate on benchmark datasets but also more resilient against evolving and adversarial threats in real-world networks.

DECLARATIONS

All authors declare that they have no conflicts of interest.

REFERENCES

1. Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y., 2016. *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.
2. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, *521*(7553), pp.436-444.
3. Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.
4. Denning, D.E., 1987. An intrusion-detection model. *IEEE Transactions on software engineering*, (2), pp.222-232.
5. Wang, W., Sheng, Y., Wang, J., Zeng, X., Ye, X., Huang, Y. and Zhu, M., 2017. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. *IEEE access*, *6*, pp.1792-1806.
6. Vinayakumar, R., Alazab, M., Soman, K.P., Poornachandran, P., Al-Nemrat, A. and Venkatraman, S., 2019. Deep learning approach for intelligent intrusion detection system. *IEEE access*, *7*, pp.41525-41550.
7. Moustafa, N. and Slay, J., 2015, November. 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.
8. Sharafaldin, I., Lashkari, A.H., and Ghorbani, A.A., 2018. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, *1*(2018), pp.108-116.
9. Rani, M., and Gagandeep, 2022. Effective network intrusion detection by addressing class imbalance with deep neural networks multimedia tools and applications. *Multimedia Tools and Applications*, *81*(6), pp.8499-8518.
10. Nawaz, M. H., Ahsan, A., Khan, I. U., Wang, Y., Ahmad, M., & Akhtar, M. S. (2025). Mitigating Message Injection Attacks in Internet of Vehicles Using Deep Learning Based Intrusion Detection System. *ICCK Transactions on Advanced Computing and Systems*, *1*(4), 208-221.
11. Al-Turaiki, I. and Altwaijry, N., 2021. A convolutional neural network for improved anomaly-based network intrusion detection. *Big Data*, *9*(3), pp.233-252.
12. Abdallah, M., An Le Khac, N., Jahromi, H. and Delia Jurcut, A., 2021, August. A hybrid CNN-LSTM based approach for anomaly detection systems in SDNs. In *Proceedings of the 16th International Conference on Availability, Reliability and Security* (pp. 1-7).
13. Aljanabi, M. and Kumaran, N., 2024. Effective Intrusion Detection through Hybrid CNN-LSTM and Grey Wolf Optimization for Feature Selection in Complex Network Environments. *GK International Journal of Advanced Research in Engineering and Technology*, *1*(1), pp.22-32.
14. Gwon, H., Lee, C., Keum, R. and Choi, H., 2019. Network intrusion detection based on LSTM and feature embedding. *arXiv preprint arXiv:1911.11552*.
15. Bhuyan, M.H., Bhattacharyya, D.K. and Kalita, J.K., 2013. Network anomaly detection: methods, systems, and tools. *Ieee communications surveys & tutorials*, *16*(1), pp.303-336.
16. Diro, A.A. and Chilamkurti, N., 2018. Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Generation Computer Systems*, *82*, pp.761-768.
17. Shone, N., Ngoc, T.N., Phai, V.D. and Shi, Q., 2018. A deep learning approach to network intrusion detection. *IEEE transactions on emerging topics in computational intelligence*, *2*(1), pp.41-50.
18. Kim, J., Shin, N., Jo, S.Y. and Kim, S.H., 2017, February. Method of intrusion detection using deep neural network. In *2017 IEEE international conference on big data and smart computing (BigComp)* (pp. 313-316). IEEE.

19. Duong, H.T., Le, V.T. and Hoang, V.T., 2023. Deep learning-based anomaly detection in video surveillance: A survey. *Sensors*, *23*(11), p.5024.
20. Al-Zewairi, M., Almajali, S., & Ayyash, M. (2020). Unknown security attack detection using shallow and deep ANN classifiers. *Electronics*, *9*(12), 2006.
21. Javaid, A., Niyaz, Q., Sun, W. and Alam, M., 2016, May. A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)* (pp. 21-26).
22. Kwon, D., Kim, H., Kim, J., Suh, S.C., Kim, I., and Kim, K.J., 2019. A survey of deep learning-based network anomaly detection. *Cluster Computing*, *22*(Supple 1), pp.949-961.
23. Khalaf, L.I., Alhamadani, B., Ismael, O.A., Radhi, A.A., Ahmed, S.R. and Algburi, S., 2024, May. Deep learning-based anomaly detection in network traffic for cyber threat identification. In *Proceedings of the Cognitive Models and Artificial Intelligence Conference* (pp. 303-309).
24. Barrak, A., Petrillo, F. and Jaafar, F., 2022. Serverless on machine learning: A systematic mapping study. *IEEE Access*, *10*, pp.99337-99352.
25. Naqash, T., Shah, S.H. and Islam, M.N.U., 2022. Statistical analysis-based intrusion detection system for ultra-high-speed software-defined network. *International Journal of Parallel Programming*, *50*(1), pp.89-114.
26. Li, Z., Qin, Z., Huang, K., Yang, X. and Ye, S., 2017. Intrusion Detection Using Convolutional Neural Networks for Representation Learning. In: Neural Information Processing, Lecture Notes in Computer Science, vol. 10638, pp.858–866. Springer. doi:10.1007/978-3-319-70139-4_87.
27. Narmadha, S., & Balaji, N. V. (2025). Improved network anomaly detection system using optimized autoencoder− LSTM. *Expert Systems with Applications*, *273*, 126854.
28. Singh, A., & Jang-Jaccard, J. (2022). Autoencoder-based unsupervised intrusion detection using multi-scale convolutional recurrent networks. *arXiv preprint arXiv:2204.03779*.
29. Shafi, S. M., & Chinnappan, S. K. (2024). Hybrid transformer-CNN and LSTM model for lung disease segmentation and classification. *PeerJ Computer Science*, *10*, e2444.
30. Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521-13617.
31. Tan, Y., Hu, C., Zhang, K., Zheng, K., Davis, E. and Park, J., 2020. LSTM-based anomaly detection for non-linear dynamical system. IEEE Access. doi: 10.1109/ACCESS.2020.2999065.
32. Yin, C., Zhu, Y., Fei, J. and He, X., 2017. A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5, pp.21954–21961. doi: 10.1109/ACCESS.2017.2762418.
33. Al-Qatf, M., Lasheng, Y., Al-Habib, M. and Al-Sabahi, K., 2018. Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. IEEE Access, 6, pp.52843–52856. doi: 10.1109/ACCESS.2018.2869577.
34. Mirsky, Y., Doitshman, T., Elovici, Y. and Shabtai, A., 2018. Kitsune: An ensemble of autoencoders for online network intrusion detection. Proceedings of the Network and Distributed System Security Symposium (NDSS). doi: 10.14722/ndss.2018.23211.
35. Wu, P. and Guo, H., 2019. LuNet: A deep neural network for network intrusion detection. arXiv preprint arXiv:1909.10031.
36. Aygun, R.C. and Yavuz, A.G., 2017. Network anomaly detection with stochastically improved autoencoder based models. In 4th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud) and 3rd IEEE International Conference on Scalable and Smart Cloud (SSC), pp.193–198.
37. Yan, H., Zhang, W., Chen, Q., Li, X., Sun, W., Li, H. and Lin, X., 2023. RECESS Vaccine for Federated Learning: Proactive Defense Against Model Poisoning Attacks. arXiv preprint arXiv:2310.05431.