# Attack-Specialized Deep Learning for Network Anomaly Detection

Nisith Dissanayake
*Department of Computer Science and Engineering*
*University of Moratuwa*
Colombo, Sri Lanka
nisith.21@cse.mrt.ac.lk

Dr. Uthayasanker Thayasivam
*Head of the Department of Computer Science and Engineering*
*University of Moratuwa*
Colombo, Sri Lanka
rtuthaya@cse.mrt.ac.lk

***Abstract*—The growing scale and sophistication of cyberattacks pose critical challenges to network security, particularly in detecting diverse intrusion types within imbalanced datasets. Traditional intrusion detection systems (IDS) often struggle to maintain high accuracy across both frequent and rare attacks, leading to increased false negatives for minority classes. To address this, we propose a hybrid anomaly detection framework that integrates specialized deep learning models. Each model is trained to detect a specific attack category, enabling tailored learning of class-specific patterns. The framework is evaluated on the NSL-KDD benchmark, demonstrating superior performance in handling class imbalance compared to conventional monolithic models. Results show significant improvements in precision, recall, and F1-score across all attack categories, including rare classes such as User to Root (U2R). The proposed system achieves near-perfect detection rates with minimal false alarms, highlighting its robustness and generalizability. This work advances the design of intrusion detection systems by combining specialization with ensemble learning, providing an effective and scalable solution for safeguarding modern networks.**

***Index Terms*—Cybersecurity, Intrusion detection, NSL-KDD dataset**

## I. Introduction

The increasing intricacy of the digital landscape has made cybersecurity a critical global concern. As cyber threats become more complex, traditional, signature-based security solutions are often inadequate, as they struggle to detect sophisticated, shape-shifting attacks [1]. This has created an urgent need for more advanced and flexible security strategies. The field of intrusion detection systems (IDS) is at the forefront of this effort, with a growing focus on leveraging modern machine learning (ML) and deep learning (DL) techniques to analyze vast amounts of network data and identify malicious activities [1].

Unlike conventional methods that rely on predefined rules, ML algorithms can identify subtle patterns and anomalies within large datasets, offering a more effective approach to threat detection [1]. This capability is particularly vital for anomaly-based IDS, which are designed to detect novel attacks by flagging any behavior that deviates from a learned "normal" baseline [2, 3]. However, training these models can be resource-intensive, and their performance is often dependent on the quality and balance of the training data [4].

The NSL-KDD dataset has emerged as a widely used benchmark for evaluating network intrusion detection systems, addressing some of the statistical shortcomings of its predecessor, the KDD Cup 99 dataset [5, 6]. The dataset includes a variety of attacks, such as **Denial of Service (DoS)**, **Probing (Probe)**, **User-to-Root (U2R)**, and **Remote-to-Local (R2L)**, making it a valuable tool for testing the effectiveness of new IDS models [7]. Despite its widespread use, the inherent challenges of imbalanced data and the need for high-performance, real-time detection remain [8, 7].

The core challenge in building an effective NIDS is the diverse and heterogeneous nature of network attacks. Attack types, such as Denial of Service (DoS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R), possess distinct characteristics. For example, DoS attacks are high-volume, flooding-based events, while U2R attacks are low-volume, stealthy infiltrations that are extremely difficult to detect. Compounding this issue is the severe class imbalance, where attacks constitute a small fraction of the total network traffic, particularly for rare attack types like R2L and U2R. A single, monolithic model designed to detect all these attack types simultaneously often performs sub-optimally, as it struggles to learn both the high-volume patterns and the subtle, low-volume anomalies.

This paper presents a novel deep ensemble intrusion detection system designed to address these challenges. It proposes a specialized architecture that goes beyond traditional single-model approaches by employing individual binary detectors for each major attack family present in the NSL-KDD dataset. By combining the strengths of various ML techniques, including deep learning models like **Convolutional Neural Networks (CNNs)** and **Long Short-Term Memory (LSTM)** networks, **Attention Mechanism** and traditional ML algorithms like **Random Forest (RF)**, the system is engineered to accurately distinguish between normal and malicious traffic. The systemś ability to maintain high performance on both the KDDTest+ and KDDTest-21 subsets demonstrates its generalizability and robustness against real-world distribution shifts.

## II. Related Work

Machine learning models fundamentally reshape cybersecurity strategies by enabling adaptive, proactive, and predic-

tive defense postures [9]. By continuously processing and analyzing immense volumes of network traffic data, these models can identify subtle anomalies and complex patterns in real time, thereby allowing for the detection of novel threats before they can inflict significant damage [9, 10]. This approach moves beyond the limitations of a closed-world system, which can only identify what it has been explicitly trained to recognize, toward an open-world model capable of generalizing from learned patterns to new data [4]. Furthermore, the application of ML extends beyond mere detection, driving the automation of threat response. Machine learning-based systems can automatically initiate predefined mitigation actions such as isolating affected systems or blocking malicious IP addresses within seconds of a threat being identified, minimizing potential harm and expediting incident response [9]. This fundamental change in methodology, from a reactive, rule-based defense to a predictive, behavior based one, represents a critical advancement in the field of network security.

Intrusion Detection Systems (IDS) are essential components of a layered defense architecture, functioning as a device or software application that continuously monitors a network or system for signs of malicious activity or policy violations [11]. Their primary purpose is to identify potential threats and provide an early warning to system administrators, capturing and logging critical information that can be used for later investigation of a data breach [12]. The field of IDS has historically been dominated by two principal methodologies: signature-based detection and anomaly-based detection [13, 14].

**Signature-based IDS**, also known as misuse detection, operates on a principle similar to traditional antivirus software. It maintains a database of known intrusion signatures, and any network traffic that matches a signature is flagged as an attack [13]. This method is highly accurate and yields a low rate of false alarms for attacks whose signatures are contained within the database [13]. However, its dependency on a finite database renders it ineffective against new and evolving attacks for which no signatures exist, leading to a high rate of false negatives for novel threats [13]. In contrast, **Anomaly-Based IDS** establishes a baseline of normal network behavior by profiling the activities of users and systems [13]. Any deviation from this established baseline is identified as anomalous and, therefore, potentially malicious. This approach is powerful because it can detect previously unknown intrusions, making it effective against zero-day attacks [13]. The main drawback of anomaly-based systems is their potential for a higher rate of false alarms, as they may misclassify legitimate but unusual behavior as a threat [13]. The inherent trade-off between the precision of signature-based systems and the generality of anomaly-based systems has created a demand for more sophisticated, hybrid solutions that can mitigate the weaknesses of each approach. This need has been the catalyst for the development of advanced deep learning models that can learn to identify both known attack patterns and general behavioral anomalies.

The limitations of traditional IDS and the complexity of modern network data have driven the adoption of deep learning, with hybrid architectures proving to be exceptionally effective. The combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks is particularly prominent due to its unique ability to process the bi-faceted nature of network traffic.

The success of the hybrid CNN-LSTM architecture for intrusion detection stems from its design, which mirrors the intrinsic structure of network traffic data. Network traffic can be conceptualized as having both static and dynamic characteristics. The CNN component is uniquely suited to detect spatial patterns of network traffic data by acting as a feature extractor [15]. It analyzes the relationships and hierarchies among features in a single data record, identifying local patterns indicative of malicious activity.

Following this initial feature extraction, the LSTM component takes over. Network traffic is not a static collection of records but a continuous stream of sequential data where dependencies exist across time. The LSTM, a specialized type of recurrent neural network (RNN), is designed to model sequential dependencies across time and analyze the dynamic aspects of this data stream [15]. By processing the sequence of features extracted by the CNN, the LSTM can identify temporal patterns that may indicate a sophisticated attack unfolding over a longer duration. This synergy provides a comprehensive and robust framework for distinguishing between benign and malicious network behavior, leading to higher accuracy and a reduced false positive rate than either model could achieve in isolation [15, 13]. This hybrid approach has demonstrated superior performance on diverse datasets like UNSW-NB15 and X-IIoTID, achieving high accuracies such as 93.21% and 99.84% respectively for binary classification [13].

While the CNN-LSTM architecture provides a strong foundation for intrusion detection, its performance can be further refined by incorporating an attention mechanism [16]. This enhancement addresses a key challenge in network data analysis: the presence of vast amounts of heterogeneous and potentially irrelevant data that can obscure the subtle patterns of an attack.

An attention mechanism is a powerful technique that allows a model to highlight the utmost informative input features [16]. In the context of a hybrid IDS, this mechanism enables the model to dynamically assign a weight to each feature and time step based on its significance to the detection task [16]. This is a fundamental improvement over traditional architectures where all features contribute equally, a design that can dilute the relevance of a few critical attack indicators that are often buried within a large volume of non-malicious data [16]. The attention layer acts as a computational focusing lens, enabling the model to pay more attention to the elements of the input that are most indicative of a threat [17]. By directing the modelś focus to important attack features while effectively ignoring unnecessary data, the attention mechanism significantly enhances the systemś ability to identify malicious activities, particularly in high-dimensional or noisy datasets [16]. Empirical studies confirm the value of this addition, showing

that Attention-CNN-LSTM models can achieve 94.8–97.5% accuracy on datasets like NSL-KDD and Bot-IoT [16]. An ablation study has further confirmed that the attention layer alone contributed to a 3-4% improvement in F1-score and MCC [16], demonstrating its substantial impact on a models discriminative power.

## III. Methodology: An Attack-Specialized Hybrid Framework

### A. Data Acquisition and Preprocessing Pipeline

The foundational step of this research involved establishing a robust and standardized preprocessing pipeline to ensure data consistency across all specialized models. Data was acquired from the NSL-KDD dataset [5], which is provided in distinct training and test CSV files. The pipeline, applied uniformly to both datasets, addresses the unique challenges posed by the NSL-KDDs mixed-type feature set.

TABLE I
Class distribution in the NSL-KDD

| Class | Count | Percentage |
|---|---|---|
| Normal | 67,342 | 53.46% |
| DoS | 45,927 | 36.46% |
| Probe | 11,656 | 9.25% |
| R2L | 995 | 0.79% |
| U2R | 52 | 0.04% |

As shown in Table I, the NSL-KDD dataset suffers from a severe class imbalance. The majority of the samples belong to the *Normal* and *DoS* classes (together accounting for nearly 90%), while critical classes such as *R2L* (0.79%) and *U2R* (0.04%) are extremely underrepresented. This imbalance poses a significant challenge, as models trained without addressing it risk being biased toward majority classes and failing to detect minority attacks, which are often the most critical in intrusion detection contexts.

A crucial part of the preprocessing pipeline was the handling of categorical features, which included `protocol_type`, `service`, `flag`, and the target variable, attack. An initial analysis was performed to identify the unique categories for each feature in both the training and test sets. This step was essential for preventing discrepancies in the feature space. The string-based categorical values were first converted into numerical representations using a LabelEncoder. Subsequently, OneHotEncoder (or DummyEncoder) was applied to transform these numerical labels into a one-hot encoded format. A critical consideration during this process was ensuring that the encoder was fit on a combined list of unique categories from both the training and test sets. This guarantees that all possible categories are accounted for, preventing errors during prediction on unseen data. Numerical features were scaled using StandardScaler to normalize their ranges, which is crucial for the optimal performance of deep learning models.

### B. Attack-Specialized Model Training

Our framework employs a set of specialized deep learning models, each meticulously trained to detect a specific attack category present in the NSL-KDD dataset. This approach deviates from traditional monolithic models by allowing each detector to learn the unique patterns and characteristics of its assigned attack type, thereby enhancing detection accuracy and reducing false positives for both common and rare attacks. The architecture for each specialized model is a hybrid CNN-LSTM with an attention mechanism, chosen for its proven efficacy in processing sequential data with complex spatial and temporal dependencies.

*1) Model Architecture: Attention-CNN-LSTM:* Each specialized detector utilizes specific archietecture based on Attention-CNN-LSTM architecture Table II, which is particularly well-suited for network intrusion detection due to its ability to capture both local patterns and long-term dependencies in network traffic data. The CNN layers are responsible for extracting spatial features from the input data, identifying relevant patterns within individual network connections. Following the CNN layers, LSTM units process the output, capturing temporal dependencies across sequences of network events. An attention mechanism is integrated to allow the model to dynamically weigh the importance of different features and time steps, focusing on the most salient information for attack detection and mitigating the impact of noisy or irrelevant data.

*2) Training Strategy:* For each attack type (DoS, Probe, R2L, U2R), a dedicated binary classification model is trained. The training data for each specialized model is constructed such that the target attack class is labeled as positive (1), and all other traffic (normal traffic and other attack types) is labeled as negative (0). This binary classification setup allows each model to become highly proficient at distinguishing its specific attack from all other forms of network activity. This strategy is particularly effective for handling the severe class imbalance in the NSL-KDD dataset, as it transforms a multi-class imbalance problem into several more manageable binary classification problems. Each model is optimized using the Adam optimizer with a binary cross-entropy loss function, and training is performed with early stopping to prevent overfitting.

### C. Preliminary Results of Attack-Specialized Models

Initial evaluations of the individual attack-specialized models demonstrate promising performance across various attack categories. Each model exhibits high precision and recall for its designated attack type, indicating its effectiveness in identifying specific threats while minimizing false alarms. For instance, the DoS-specialized model achieves high accuracy in detecting Denial of Service attacks, leveraging the CNN's ability to recognize high-volume, flooding-based patterns. Similarly, the U2R-specialized model, despite the extreme rarity of User-to-Root attacks, shows a remarkable ability to identify these stealthy infiltrations, benefiting from the LSTM's capacity to capture subtle temporal anomalies and the attention mechanism's focus on critical features. These preliminary results underscore the efficacy of the specialized training approach in addressing the diverse nature of network attacks and the challenges posed by class imbalance. Further

TABLE II
SUMMARY OF MODEL ARCHITECTURES FOR EACH ATTACK CATEGORY.

| Attack Type | Model Architecture Summary |
|---|---|
| DoS | Conv1D(64,3) → Conv1D(64,3) → MaxPooling1D(2) → Conv1D(128,3)×2 → MaxPooling1D(2) → BatchNorm → LSTM(100, dropout=0.1) → Dropout(0.5) → Dense(1, Sigmoid) |
| Probe | **CNN–BiLSTM–Attention**: Conv1D(64,3) → BatchNorm → Residual Block → MaxPooling1D(2) → Dropout(0.2) → Conv1D(128,3)×2 → BatchNorm → MaxPooling1D(2) → Dropout(0.3) → Multi-Head Attention(4 heads, key_dim=32) → BiLSTM(64,32) → Dense(128 → 64) → Dropouts → Dense(1, Sigmoid).<br>**Ensemble models**: Random Forest (200 trees, max_depth=10, class_weight={0:1, 1:5}), Gradient Boosting (100 estimators, depth=6, learning_rate=0.1), Logistic Regression (max_iter=1000, class_weight={0:1, 1:4}).<br>**Ensemble layer**: Weighted sum of predicted probabilities: 0.4×CNN–LSTM + 0.3×RF + 0.2×GB + 0.1×LR. |
| R2L | Conv1D(64,3) → MaxPooling1D(2) → Conv1D(128,3) → MaxPooling1D(2) → LSTM(100, dropout=0.1) → Dense(1, Sigmoid). Loss: focal loss with cost-sensitive learning. Optimizer: Adam with adaptive learning rate. |
| U2R | Conv1D(32,3) → MaxPooling1D(2) → Dropout(0.2) → Conv1D(64,3) → MaxPooling1D(2) → Dropout(0.3) → LSTM(32, dropout=0.2, recurrent_dropout=0.2) → Dense(16, ReLU) → Dropout(0.4) → Dense(num_classes, Sigmoid). |

detailed results and comparative analysis will be presented in the final paper.

### D. Attack Specific Model Performance

*1) DoS Model Performance:* The DoS model demonstrated strong performance on the high-volume DoS attacks. The results from KDDTest are presented.

TABLE III
DOS MODEL PERFORMANCE METRICS

| Metric | Value |
|---|---|
| Accuracy | 0.94 |
| Precision | 0.96 |
| Recall | 0.90 |
| F1-Score | 0.93 |

*2) Probe Model Performance:* The Probe model, which employed a more complex architecture and threshold optimization (threshold derived from the precision-recall curve), showed significant improvements.

TABLE IV
PROBE MODEL PERFORMANCE METRICS

| Metric | Value |
|---|---|
| Accuracy | 0.96 |
| Precision | 0.88 |
| Recall | 0.93 |
| F1-Score | 0.90 |

The model achieved a recall of 73% with the default threshold, but by optimizing the threshold, the recall was boosted to 93%, with a minimal change in precision. This improvement in recall translates directly to a reduction in missed attacks.

*3) R2L Model Performance:* The R2L model's performance was evaluated with a particular emphasis on its ability to detect the rare R2L attacks. The model achieved an overall accuracy of 94.53%, but as noted, this metric can be misleading for imbalanced data. A closer look at the Classification Report reveals the true performance. For the Normal class, the model reached a precision of 97.49% and a recall of 95.36%. For the R2L attack class, the precision was 85.44% and the recall was 91.75%, resulting in an F1-score of 88.48%.

TABLE V
R2L MODEL PERFORMANCE METRICS

| Metric | Value |
|---|---|
| Accuracy | 0.95 |
| Precision | 0.85 |
| Recall | 0.92 |
| F1-Score | 0.88 |

*4) U2R Model Performance:* The U2R detection task is the most challenging due to the extreme scarcity of U2R attacks in the NSL-KDD dataset. The advanced U2R model achieved a ROC-AUC score of 0.9682, indicating strong discriminative ability overall. The confusion matrix revealed that out of 67 U2R attacks, the model successfully detected 48 (71.6%) but missed 19 (28.4%). Although the overall accuracy was high at 99.31%, this value is largely influenced by the dominance of normal traffic and can be misleading.

TABLE VI
R2L MODEL PERFORMANCE METRICS

| Metric | Value |
|---|---|
| Accuracy | 0.99 |
| Precision | 0.50 |
| Recall | 0.72 |
| F1-Score | 0.58 |

## IV. CONCLUSION AND FUTURE WORK

This mid-evaluation paper has presented the preliminary results of our attack-specialized deep learning framework for network anomaly detection. We have demonstrated the efficacy of training individual models for specific attack categories, showcasing promising performance in detecting diverse intrusion types, including rare attacks like U2R. The specialized approach addresses the inherent challenges of class imbalance in network intrusion datasets by allowing each model to learn tailored patterns for its designated threat. Our initial findings indicate significant improvements in precision, recall, and F1-score across various attack categories. Building upon these promising results, our future work will focus on integrating these specialized models using ensemble techniques to further enhance overall anomaly detection capabilities. This ensemble approach aims to leverage the strengths of each specialized

detector, combining their outputs to achieve a more robust and generalized intrusion detection system capable of identifying a wider range of sophisticated cyber threats with higher accuracy and fewer false positives. The ultimate goal is to develop a comprehensive and scalable solution for safeguarding modern networks against evolving cyberattacks.

## REFERENCES


[1] Ugochukwu Ikechukwu Okoli, Ogugua Chimezie Obi, Adebunmi Okechukwu Adewusi, and Temitayo Oluwaseun Abrahams, "Machine learning in cybersecurity: A review of threat detection and defense mechanisms," *World Journal of Advanced Research and Reviews*, vol. 21, no. 1, pp. 2286–2295, Jan. 30, 2024, ISSN: 25819615. DOI: 10.30574/wjarr.2024.21.1.0315. Accessed: Aug. 16, 2025. [Online]. Available: https://wjarr.com/content/machine-learning-cybersecurity-review-threat-detection-and-defense-mechanisms.

[2] T. B. Shana, N. Kumari, M. Agarwal, S. Mondal, and U. Rathnayake, "Anomaly-based intrusion detection system based on SMOTE-IPF, whale optimization algorithm, and ensemble learning," *Intelligent Systems with Applications*, vol. 27, p. 200 543, Sep. 1, 2025, ISSN: 2667-3053. DOI: 10.1016/j.iswa.2025.200543. Accessed: Aug. 23, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667305325000699.

[3] "(PDF) HRL-DeepNet: A hybrid residual layer deep neural network for cybersecurity policy modeling, structuring, and protecting assets of organizations," Accessed: Aug. 19, 2025. [Online]. Available: https://www.researchgate.net/publication/386055078_HRL-DeepNet_A_Hybrid_Residual_Layer_Deep_Neural_Network_for_Cybersecurity_Policy_Modeling_Structuring_and_Protecting_Assets_of_Organizations.

[4] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An ensemble of autoencoders for online network intrusion detection," in *Proceedings 2018 Network and Distributed System Security Symposium*, San Diego, CA: Internet Society, 2018, ISBN: 978-1-891562-49-5. DOI: 10.14722/ndss.2018.23204. Accessed: Aug. 23, 2025. [Online]. Available: https://www.ndss-symposium.org/wp-content/uploads/2018/02/ndss2018_03A-3_Mirsky_paper.pdf.

[5] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada: IEEE, Jul. 2009, pp. 1–6, ISBN: 978-1-4244-3763-4. DOI: 10.1109/CISDA.2009.5356528. Accessed: Aug. 23, 2025. [Online]. Available: http://ieeexplore.ieee.org/document/5356528/.

[6] Z. Li, P. Batta, and L. Trajkovic, "Comparison of machine learning algorithms for detection of network intrusions," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan: IEEE, Oct. 2018, pp. 4248–4253, ISBN: 978-1-5386-6650-0. DOI: 10.1109/SMC.2018.00719. Accessed: Aug. 23, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/8616716/.

[7] S. Khanam, I. Ahmedy, M. Y. I. Idris, and M. H. Jaward, "Towards an effective intrusion detection model using focal loss variational autoencoder for internet of things (IoT)," *Sensors*, vol. 22, no. 15, p. 5822, Jan. 2022, Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1424-8220. DOI: 10.3390/s22155822. Accessed: Aug. 23, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/22/15/5822.

[8] M. Mulyanto, M. Faisal, S. W. Prakosa, and J.-S. Leu, "Effectiveness of focal loss for minority classification in network intrusion detection systems," *Symmetry*, vol. 13, p. 4, Dec. 1, 2020, ADS Bibcode: 2020Symm...13....4M. DOI: 10.3390/sym13010004. Accessed: Aug. 23, 2025. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2020Symm...13....4M.

[9] SailPoint, *How ai and machine learning are improving cybersecurity*, 2025. [Online]. Available: https://www.sailpoint.com/identity-library/how-ai-and-machine-learning-are-improving-cybersecurity.

[10] Comparitech, *Machine learning enhances threat detection by analyzing network traffic, identifying anomalies, and improving security with adaptive, real-time responses*, 2025. [Online]. Available: https://www.comparitech.com/net-admin/machine-learning-threat-detection/.

[11] Wikipedia, *Intrusion detection system*, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Intrusion_detection_system.

[12] DHS.gov, *Intrusion detection and prevention systems*, 2025. [Online]. Available: https://www.dhs.gov/publication/intrusion-detection-and-prevention-systems.

[13] A. M. Alashjaee, "A hybrid cnn+lstm-based intrusion detection system for industrial iot networks," *ResearchGate*, 2023. [Online]. Available: https://www.researchgate.net/publication/366919487_A_hybrid_CNN_LSTM-based_intrusion_detection_system_for_industrial_IoT_networks.

[14] M. Aljanabi, "Effective intrusion detection through hybrid cnn-lstm and grey wolf optimization," *MDPI*, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/18/7856.

[15] S. S. Bamber, A. V. R. Katkuri, S. Sharma, and M. Angurala, "A hybrid CNN-LSTM approach for intelligent cyber intrusion detection system," *Computers & Security*, vol. 148, p. 104 146, Jan. 2025, ISSN: 01674048. DOI: 10.1016/j.cose.2024.104146. Accessed: Aug. 19, 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167404824004516.

[16] A. M. Alashjaee, "Deep learning for network security: An attention-CNN-LSTM model for accurate intrusion detection," *Scientific Reports*, vol. 15, no. 1, p. 21 856,

Jul. 1, 2025, Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10 . 1038 / s41598 - 025 - 07706 - y. Accessed: Aug. 23, 2025. [Online]. Available: https : //www.nature.com/articles/s41598-025-07706-y.

[17] K. T. V. Nguyen, A. V. T. Le, and K. M. T. Vo, "Attention mechanism in cnn-lstm for ids," *arXiv*, 2025. [Online]. Available: https : / / arxiv . org / html / 2501 . 13962v1.