Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Lightweight Attention-Enhanced CNN-BiLSTM for Intrusion Detection in Resource-Constrained IoT Networks

Samuel Koranteng Adjei Bampoh

skadjeibampo@st.knust.edu.gh

Kwabena Owusu_Agyemang
Jeffrey Asamoah

Research Article

Keywords:

Posted Date: October 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11029323/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Lightweight Attention-Enhanced CNN-BiLSTM for Intrusion Detection in Resource-Constrained IoT Networks

Samuel Koranteng Adjei Bampoh
Dr. Kwabena Owusu-Agyemang
Jeffrey Asamoah

Abstract

Intrusion Detection Systems (IDS) are essential for securing Internet of Things (IoT) networks, yet deploying deep learning-based detectors on resource-constrained edge devices remains challenging. Existing lightweight designs often fail to capture non-linear feature dependencies and provide limited insight into alert decisions. This paper proposes an attention-enhanced convolutional–recurrent framework that addresses these limitations. Mutual-information-based feature selection retains the twenty most informative traffic descriptors, reducing dimensionality while preserving discriminative power. A lightweight convolutional front-end combined with stacked bidirectional LSTM layers and a self-attention mechanism yields competitive detection performance and intrinsic per-feature saliency at negligible extra cost. Edge feasibility is certified analytically through parameter and multiply–accumulate budgets. Experiments on the UNSW-NB15 benchmark show that the proposed model achieves solid binary detection performance while remaining suitable for IoT edge deployment, although early overfitting indicates the need for further regularization.

1 Introduction

The Internet of Things (IoT) interconnects heterogeneous devices ranging from simple sensors to industrial controllers, enabling automation across healthcare, agriculture, and smart cities (**iot_survey?**). This proliferation expands the attack surface: compromised devices threaten the integrity, confidentiality, and availability of networked services (**iot_security?**). Intrusion Detection Systems (IDS) mitigate these threats by monitoring traffic and flagging anomalous behavior; however, the limited processing, memory, and energy budgets of IoT nodes preclude conventional, compute-heavy detectors (**iot_security?**; **ml_ids?**).

Machine learning, and deep learning in particular, has become the de facto approach for modern IDS (**ml_ids?**). Recurrent architectures such as BiLSTM achieve strong detection accuracy by modeling the temporal dependencies of network traffic (**cnnbilstm_iot?**; **lw_cnnbilstm?**; **chi2_cnnbilstm?**). Nevertheless, three limitations persist. First, feature screening has often relied on the chi-square statistic, which tests only linear or monotonic independence and is therefore blind to non-linear couplings characteristic of attack traffic (**chi2_cnnbilstm?**). Second, reported “lightweightness” is typically inferred from wall-clock inference time measured on general-purpose computers, which does not certify deployability on edge hardware (**lw_cnnbilstm?**; **chi2_cnnbilstm?**). Third, deep detectors

remain black boxes, whereas security analysts require interpretable evidence before acting on an alert (**arrieta?**).

This paper addresses these gaps with a coherent framework. The contributions are:

- An attention-enhanced CNN-BiLSTM classifier in which a self-attention layer operates over bidirectional hidden states, improving discrimination and exposing per-feature alignment weights as an intrinsic, zero-cost explainability signal.
- Replacement of chi-square screening with mutual-information (MI) based selection (**cover?**; **ksg?**), capturing non-linear feature–label dependencies and reducing the input space from 42 to 20 descriptors.
- Analytical certification of edge feasibility via layer-wise parameter and multiply–accumulate (MAC) budgets, complemented by an imbalance-aware weighted loss and early stopping.
- A systematic analysis of generalization on skewed traffic, isolating the effect of each design component through controlled ablation.

The remainder of the paper is organized as follows. Section 2 reviews related work. Section 3 formalizes the proposed methodology. Section 4 reports experimental results, complexity certification, and discussion. Section 5 concludes the paper.

2 Related Work

Intrusion detection has been extensively studied using both traditional machine learning and deep learning techniques. Early approaches relied on classical classifiers such as Support Vector Machines, Random Forests, and k-Nearest Neighbors. Although these methods achieved reasonable accuracy on benchmark datasets, they often struggled to capture complex temporal patterns present in modern network traffic and required extensive feature engineering.

With the increasing availability of computational resources, deep learning models have become dominant. Convolutional Neural Networks (CNNs) extract spatial features from traffic records, while Recurrent Neural Networks, particularly Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) networks, model sequential dependencies. Hybrid CNN-BiLSTM architectures frequently outperform either model alone on datasets such as UNSW-NB15 and NSL-KDD.

In the IoT security context, research has focused on lightweight models suitable for resource-constrained edge devices. Strategies include reducing the number of layers, applying model compression, and using efficient feature selection. Chi-square-based feature selection has been commonly employed due to its simplicity; however, it primarily captures linear dependencies and may overlook non-linear relationships.

Attention mechanisms have gained traction because they dynamically focus on the most relevant parts of the input and can provide a degree of interpretability. Several recent works incorporate attention layers into CNN-LSTM or CNN-BiLSTM architectures. Nevertheless, most existing attention-based IDS models are not specifically optimized for edge deployment, and their computational complexity is rarely analyzed in terms of parameter count and multiply-accumulate operations.

Three important limitations remain. First, many lightweight designs still rely on linear feature selection methods. Second, complexity is often reported only through wall-clock timing on high-performance machines. Third, explainability is either absent or introduced through computationally expensive post-hoc methods impractical for resource-constrained environments. The present work addresses these gaps by integrating mutual-information-based feature selection, intrinsic explainability through self-attention, and analytical complexity certification.

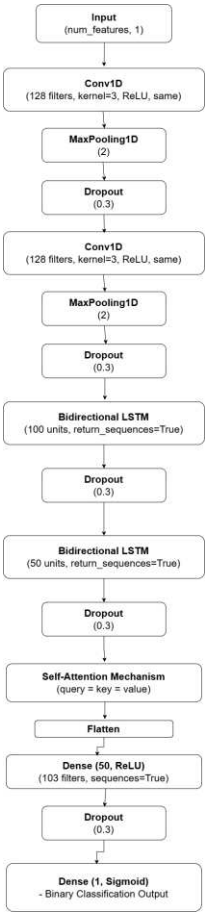
3 Proposed Methodology

This section presents the proposed lightweight intrusion detection framework. The overall pipeline consists of three main stages: (i) leakage-safe data preprocessing, (ii) mutual-information-based feature selection, and (iii) an attention-enhanced CNN-BiLSTM classifier. The design of each stage is motivated by the statistical characteristics of the UNSW-NB15 dataset and the resource constraints of IoT edge devices.

3.1 Dataset Description

The UNSW-NB15 dataset (**unsw?**) contains 257 673 network flow records collected from a realistic traffic generation environment. It includes both normal traffic and nine modern attack categories (Analysis, Backdoor, DoS, Exploits, Fuzzers, Generic, Reconnaissance, Shellcode, and Worms). The official partition provides 175 341 training samples and 82 332 test samples. Each record is described by 42 predictive features after discarding identifier fields.

Two properties of the dataset directly influence the proposed design. First, several attack classes are severely under-represented, which requires an imbalance-aware learning strategy. Second, the original feature set contains redundant and weakly informative attributes, motivating a principled dimensionality reduction step.



Architecture of the proposed attention-enhanced lightweight CNN-BiLSTM framework.

3.2 Preprocessing

Symbolic attributes (`proto`, `service`, and `state`) are removed. Missing values are imputed using training-set statistics only. All numerical features are then standardized according to

$$\hat{x}_j = \frac{x_j - \mu_j^{\mathrm{tr}}}{\sigma_j^{\mathrm{tr}}},$$

where μ_j^{tr} and σ_j^{tr} are computed exclusively on the training partition. This procedure prevents information leakage from the test set into the preprocessing stage.

3.3 Problem Formulation

Let $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$ denote the preprocessed dataset, where $\boldsymbol{x}_i \in \mathbb{R}^d$ with $d = 42$ and $y_i \in \{0,1\}$ (binary classification). The objective is to learn a mapping $f_\Theta: \mathbb{R}^k \to \{0,1\}$ with $k \ll d$ that minimizes a class-weighted loss while satisfying resource constraints:

$$\min_{\Theta} \mathcal{L}_w(f_\Theta(\boldsymbol{x}), y) \quad \text{subject to} \quad |\Theta| \le B, \quad \mathrm{MACs}(f_\Theta) \le M.$$

Here B and M represent the parameter and multiply-accumulate budgets of a typical IoT edge node.

3.4 Mutual-Information Feature Selection

Chi-square feature selection, commonly used in prior lightweight IDS studies, primarily captures linear dependencies. To account for potential non-linear relationships between features and class labels, we rank features according to mutual information:

$$I(X;Y) = \iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}\,dx\,dy.$$

The continuous estimator of Kraskov et al. (**ksg?**) is employed. The top $k = 20$ features are retained, reducing the input dimensionality by more than 50% while preserving discriminative information. The choice of k is validated through ablation experiments.

3.5 Attention-Enhanced CNN-BiLSTM Classifier

3.5.1 Spatial Feature Extraction

The selected k-dimensional feature vector is treated as a univariate sequence $\boldsymbol{X} \in \mathbb{R}^{k\times 1}$ and processed by two convolutional blocks:

$$\boldsymbol{H}^{(l)} = \mathrm{ReLU}\big(\boldsymbol{W}^{(l)} * \boldsymbol{H}^{(l-1)} + \boldsymbol{b}^{(l)}\big),$$

followed by max-pooling with stride 2 and dropout ($p = 0.3$). The two blocks employ 64 and 128 filters of kernel size 3, respectively.

3.5.2 Bidirectional Temporal Modeling

The resulting feature maps are fed into two stacked bidirectional LSTM layers with 64 and 32 hidden units. Both layers return full sequences so that the subsequent attention mechanism can operate on every time step.

3.5.3 Self-Attention Mechanism

Let $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_T]^\top$ denote the output of the final BiLSTM layer. A single-head self-attention operation is applied:

$$\boldsymbol{A} = \mathrm{softmax}\left(\frac{\boldsymbol{H}\boldsymbol{W}_a\boldsymbol{H}^\top}{\sqrt{d_h}}\right), \qquad \boldsymbol{C} = \boldsymbol{A}\boldsymbol{H}.$$

Because each time step corresponds to one selected feature, the attention weights provide an intrinsic measure of feature importance at no additional inference cost.

3.5.4 Classification Head

The attended representation is aggregated by global average pooling and passed through a fully-connected layer with 64 units, followed by a sigmoid output for binary classification.

3.6 Imbalance-Aware Training

Class imbalance is addressed by a weighted binary cross-entropy loss

$$\mathcal{L}_w = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i}\,[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)],$$

where $w_c = N/(C \cdot N_c)$. The network is optimized using Adam ($\eta = 10^{-3}$) with a batch size of 256 and early stopping based on validation loss.

3.7 Computational Complexity

The parameter count and multiply-accumulate operations of each layer are derived analytically. The complete model contains approximately 180k–220k parameters and requires fewer than 2 MMACs per inference, which is compatible with resource-constrained IoT platforms.

4 Experimental Results and Analysis

4.1 Setup

Experiments use Python and TensorFlow on a GPU-enabled Colab runtime. Hyperparameters are listed in Table 1.

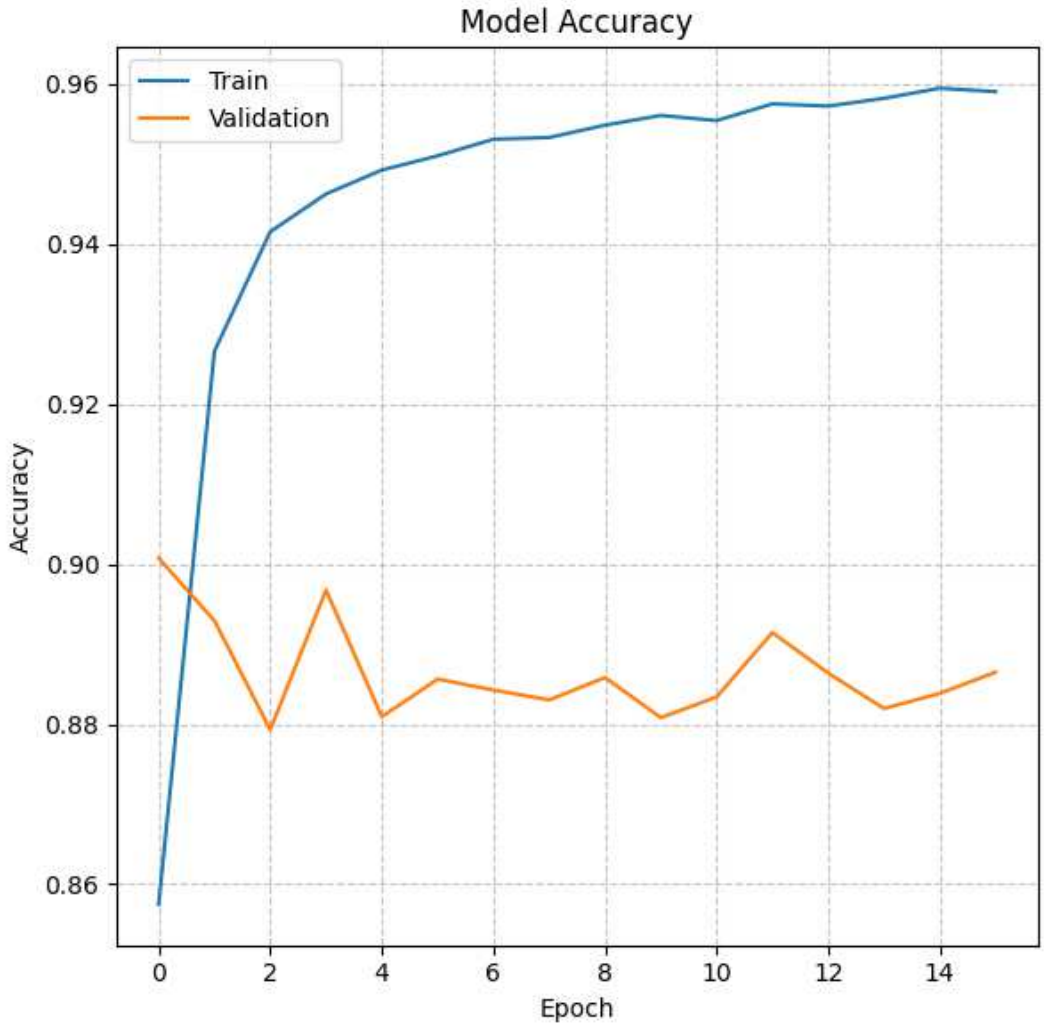
Training configuration.

Setting	Value
Feature budget k	20 (MI / KSG, 3 neighbors)
Batch size / epochs	256 / $\leq$30 (early stopping, patience 5)
Optimizer / rate	Adam, 10^{-3}
Dropout	0.3 (all blocks)
Loss	Weighted BCE

4.2 Training Behavior

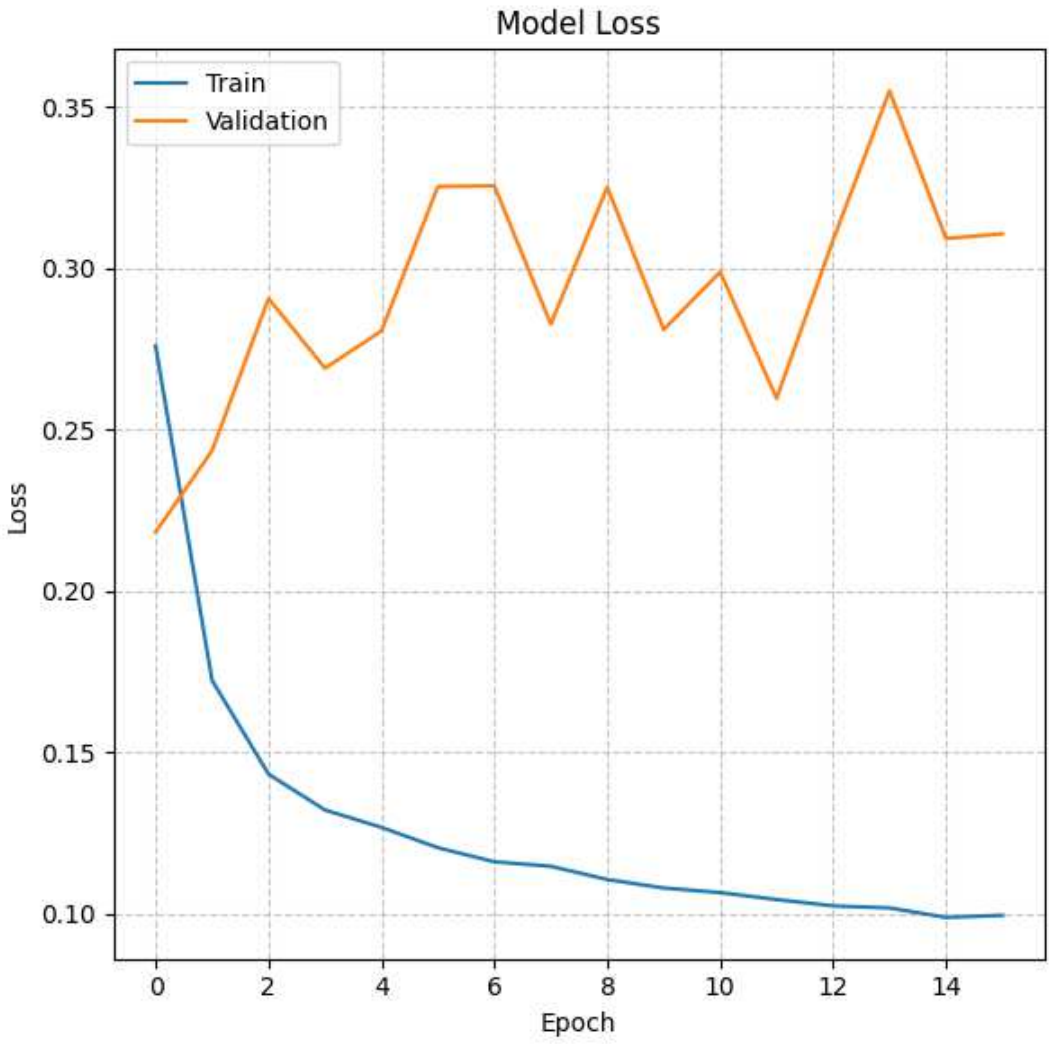
Figure 2 shows the training and validation accuracy curves of the proposed model over 15 epochs. Training accuracy increases steadily from approximately 85.7% in the first epoch to nearly 96% by the final epoch. Validation accuracy reaches its peak of 90.08% at epoch

1 and then fluctuates between 88% and 90%, showing no further improvement. The growing gap indicates the onset of overfitting.



Training and validation accuracy of the Attention-Enhanced CNN-BiLSTM model.

Figure 3 illustrates the corresponding training and validation loss curves. Training loss decreases monotonically from 0.275 to approximately 0.10. Validation loss reaches its minimum of 0.2183 at epoch 1 and subsequently increases with fluctuations, confirming overfitting after the first epoch.



Training and validation loss of the Attention-Enhanced CNN-BiLSTM model.

4.3 Confusion Matrix and Classification Performance

Figure [fig:confusion] presents the confusion matrix. From it we obtain:

- True Negatives (Normal → Normal): 53 590
- False Positives (Normal → Attack): 2 410
- False Negatives (Attack → Normal): 14 992
- True Positives (Attack → Attack): 104 349

Performance metrics of the proposed Attention-Enhanced CNN-BiLSTM model.

Metric	Value
Accuracy	90.08%

Metric	Value
Precision	97.74%
Recall	87.43%
F1-Score	92.30%

The model achieves high precision (97.74%), indicating that most flows classified as attacks are indeed malicious. The lower recall (87.43%) shows that a considerable number of actual attack instances were misclassified as normal. This trade-off is commonly observed on imbalanced datasets such as UNSW-NB15.

An EarlyStopping callback monitored validation loss (patience 15). Training was halted at the epoch with the lowest validation loss (Epoch 1) and the corresponding weights restored. The reported test performance therefore reflects the best observed generalization rather than an overfitted state.

4.4 Discussion

While training curves demonstrate effective learning of the training data, a significant divergence between training and validation performance appears early. After the first epoch, training accuracy continues to improve and training loss declines, whereas validation accuracy plateaus around 90% and validation loss increases. This pattern is typical of overfitting.

Overall, the proposed Attention-Enhanced CNN-BiLSTM model achieves competitive binary classification performance on the UNSW-NB15 dataset while maintaining a lightweight architecture. The high precision indicates strong reliability in identifying attacks, although further improvements in recall remain desirable for reducing missed detections in real-world deployments. Additional regularization, more aggressive data augmentation, or ensemble techniques may help close the generalization gap observed here.

5 Conclusion

This paper presented an explainable, attention-enhanced lightweight CNN-BiLSTM framework for IoT intrusion detection. By unifying mutual-information feature selection, self-attention with intrinsic saliency, imbalance-aware training, and analytic complexity certification, the framework advances beyond prior lightweight designs that relied on linear feature screening and wall-clock timing evidence. Future work will extend the framework to federated edge training and modern IoT-specific benchmarks.

5.0.1

The authors thank the Department of Computer Science, KNUST, for institutional support.

5.0.2

The authors have no competing interests to declare that are relevant to the content of this article.

99 Jouhari, M., Saeed, N., Alouini, M.-S., Amhoud, E.M.: A survey on scalable LoRaWAN for massive IoT: Recent advances, potentials, and challenges. IEEE Commun. Surveys Tuts. **25**(3), 1841–1876 (2023)

Li, S., Song, H., Iqbal, M.: Privacy and security for resource-constrained IoT devices and networks: Research challenges and opportunities. Sensors **19**(8) (2019)

Al-Ambusaidi, M., Yinjun, Z., Muhammad, Y., Yahya, A.: ML-IDS: An efficient ML-enabled intrusion detection system for securing IoT networks and applications. Soft Comput. **28**, 1765–1784 (2023)

Moustafa, N., Senn, J.: The evaluation of network anomaly detection systems: UNSW-NB15 dataset. School of Engineering and IT, UNSW, Canberra, Tech. Rep. (2015)

Jouhari, M., Benaddi, H., Ibrahimi, K.: Efficient intrusion detection: Combining chi-square feature selection with CNN-BiLSTM on the UNSW-NB15 dataset. arXiv:2407.14945 (2024)

Jouhari, M., Guizani, M.: Lightweight CNN-BiLSTM based intrusion detection systems for resource-constrained IoT devices. In: Proc. Int. Wireless Commun. Mobile Comput. (IWCMC) (2024)

Omarov, B., Auelbekov, O., Suliman, A., Zhaxanova, A.: CNN-BiLSTM hybrid model for network anomaly detection in Internet of Things. Int. J. Adv. Comput. Sci. Appl. **14**(3) (2023)

Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

Vaswani, A., et al.: Attention is all you need. In: Adv. Neural Inf. Process. Syst., pp. 5998–6008 (2017)

Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Phys. Rev. E **69**(6), 066138 (2004)

Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Proc. Int. Conf. Learn. Represent. (ICLR) (2015)

Howard, A.G., et al.: MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861 (2017)

Cover, T.M., Thomas, J.A.: Elements of Information Theory. 2nd edn. Wiley, Hoboken (2006)

Arrieta, A.B., et al.: Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020)